From Oracle to Noisy Context: Mitigating Contextual Exposure Bias in Speech-LLMs

Anonymous ACL submission

Abstract

Contextual automatic speech recognition (ASR) with Speech-LLMs is typically trained with oracle conversation history, but relies on error-prone history at inference, causing a train–test mismatch in the context channel that we term contextual exposure bias. We propose a unified training framework to improve robustness under realistic histories: (i) Teacher Error Knowledge by using Whisper large-v3 hypotheses as training-time history, (ii) Context Dropout to regularize over-reliance on history, and (iii) Direct Preference Optimization (DPO) on curated failure cases. Experiments on TED-LIUM 3 (in-domain) and zero-shot LibriSpeech (out-of-domain) show consistent gains under predicted-history decoding. With a two-utterance history as context, SFT with Whisper hypotheses reduce WER from 5.59% (oracle-history training) to 5.47%, and DPO further improves to 5.17%. Under irrelevant-context attacks, DPO yields the smallest degradation (5.17% $\rightarrow$ 5.63%), indicating improved robustness to misleading context. Our code and models are published on `https://anonymous.4open.science/r/Contextual_Speech_LLMs-3210`.

1 Introduction

Automatic Speech Recognition (ASR) has evolved rapidly (Ma et al., 2025), progressing from traditional architectures like CTC (Graves et al., 2006), AED (Chan et al., 2016), and RNN-T (Graves et al., 2013) to robust pre-trained models such as HuBert (Hsu et al., 2021), Whisper (Radford et al., 2023), and Whistle (Yusuyin et al., 2025; Li et al., 2022). While the recent emergence of Speech-LLMs (Cui et al., 2024) has further expanded multimodal capabilities, effectively incorporating contextual cues—a problem known as Contextual ASR (Aleksic et al., 2015; Hall et al., 2015)—remains a critical challenge across these paradigms to compensate for ambiguous acoustic evidence.

In conventional ASR systems, the utilization of contextual information primarily follows two core paradigms: Shallow Fusion and Deep Fusion (Fang et al., 2025). Shallow Fusion functions essentially as an inference-stage ensemble strategy. Enhances the accuracy of recognition within specific contexts by utilizing an external language model to score the hypotheses generated by the acoustic model (McDermott et al., 2019; Ravi et al., 2020; Guo et al., 2023). In contrast, Deep Fusion involves a more fundamental architectural innovation. This approach encodes contextual information into vector embeddings that are directly integrated into the internal components of end-to-end models. Consequently, this enables the model to learn joint representations of acoustic features and contextual cues during the training phase (Toshniwal et al., 2018; Wang et al., 2024; Tang et al., 2024; Huang et al., 2024b,a; Kolokolov et al., 2024; Shi et al., 2024; Sudo et al., 2024). Such as Contextual RNN-T (Jain et al., 2020) demonstrated the effectiveness of attending to metadata embeddings for rare word recognition, while Hou et al. (Hou et al., 2022) extended this by integrating dialogue history into streaming RNN-T encoders. Hori et al. (Hori et al., 2020) explored Transformer-based architectures for long-context modeling, highlighting the importance of cross-utterance dependencies.

Benefiting from the robust contextual reasoning of LLMs, recent research has prioritized embedding contextual cues into Speech-LLMs prompts (Chen et al., 2024; Yang et al., 2024; Shen et al., 2025; Lakomkin et al., 2024; Cheng, 2024; Lei et al., 2025; Koshkin et al., 2024; Fang et al., 2025; Gong et al., 2024; Zhou et al., 2025). Approaches vary from using textual metadata like titles and descriptions (Lakomkin et al., 2024) to incorporating specific entity lists (Chen et al., 2024) and multi-modal auxiliary inputs (Yang et al., 2024). Notably, Lakomkin et al (Lakomkin et al., 2024) also examined the model's resilience to contextual perturbations. Building on this trend, Retrieval-Augmented Generation (RAG) has been proposed as a novel method to integrate Speech-LLMs (Shen et al., 2025; Li et al., 2024; Mu et al., 2025; Gourav et al., 2024). On the other hand, Step-Audio employs a text-based context manager to maintain conversation history and support multi-turn interactions (Huang et al., 2025). The paper does not provide a separate analysis of the impact of historical text on system performance.

We study utterance-by-utterance contextual ASR, where each utterance is decoded sequentially using transcripts from previous turns as textual history. In the training stage, the model can rely on oracle history, but in deployment, oracle history is unavailable and the model must condition on error-prone ASR hypotheses, yielding a distribution shift between training-time and inference-time context. We call this mismatch in the context channel contextual exposure bias. While the challenge of utilizing imperfect context has been approached—for instance, via context dropout in speech translation (Hussein et al., 2024) or noise representation learning in dialogue ASR (Lee et al., 2024) these methods often rely on implicit regularization or auxiliary modules. They lack a direct mechanism to align the model's generation preferences to explicitly reject contextual errors.

To mitigate exposure bias in continuous-utterance ASR, we propose a unified training framework that integrates three complementary strategies. Teacher Error Knowledge: instead of using ground-truth history, we feed Whisper large-v3 decoding hypotheses as training context, exposing the model to realistic "teacher errors" and better matching inference-time conditions. Context Dropout: we randomly mask the historical context with a fixed probability, reducing over-reliance on text history and encouraging acoustic-focused transcription, thereby alleviating history overfitting. Direct Preference Optimization (DPO): we further refine contextual generation by constructing preference pairs from selected hard negatives, explicitly training the model to avoid negative behaviors and improve output quality. The main contributions of this paper are summarized as follows:

- **Leveraging contextual exposure bias to improve Speech-LLMs ASR:** We identify contextual exposure bias as a key failure mode in context-conditioned Speech-LLMs based ASR and, explicitly exploit this train-test mismatch as a guiding principle to design training and alignment strategies that boost recognition performance under realistic, imperfect historical context.
- **Noise-aware training for imperfect context:** We propose a unified training framework that aligns training with inference by (i) using a strong teacher ASR system to generate realistic, error-prone historical context during training (instantiated with Whisper large-v3 in our experiments), (ii) applying context dropout to regularize reliance on history, and (iii) using DPO with preference pairs constructed from contextual failure cases to reduce error amplification.
- **Robustness under realistic and cross-domain evaluation:** Training with strong teacher-generated historical context substantially closes the oracle–inference gap, and DPO further improves robustness by reducing the model's sensitivity to irrelevant or erroneous context, consistently yielding the best in-domain and cross-domain performance under realistic decoding conditions.

2 Method

2.1 Utterance-level Contextual ASR

We consider an utterance-level contextual ASR setting, in which each utterance is recognized sequentially and the transcripts of preceding utterances are provided as textual context for the current decoding.

Formally, an input stream is segmented into a sequence of utterances $\{X_t\}_{t=1}^{T}$, where X_t denotes the acoustic signal of the t-th utterance, and Y_t denotes its reference transcript. For each utterance t, the recognizer produces a hypothesis $\hat{Y}_t$ conditioned on the current acoustics X_t and an available context C_t:

$$\hat{Y}_t \sim p_\theta(Y|X_t, C_t) \quad (1)$$

We define the textual context C_t as a function of preceding utterances' transcripts. In the simplest case, C_t is the concatenation of the most recent N utterance-level transcripts:

$$C_t = concat(S_{t-N}, \cdots, S_{t-1}) \quad (2)$$

where S_i is the transcript used as history for utterance i, and N controls the context window size

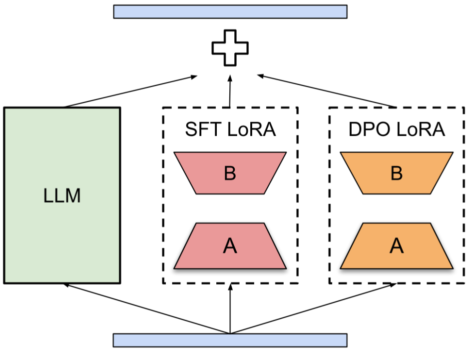

Figure 1: Model architecture

(with $N = 0$ corresponding to the no-context baseline).

A central aspect of this setting is that the quality of historical transcripts available at inference may differ from that used during training. In many experimental setups, S_i is taken to be the oracle transcript Y_i, yielding an oracle-context condition. In realistic deployment, however, the system does not have access to Y_i and must instead rely on automatically generated hypotheses from upstream recognition of previous turns. We denote this predicted-context condition by setting $S_i = \hat{Y}_i$, where $\hat{Y}_i$ may contain recognition errors, relative to reference Y_i.

This discrepancy induces a distribution mismatch in the conditioning context between training and inference: training commonly conditions on oracle history, while inference conditions on imperfect, model-generated history. In this work, we refer to this train–test mismatch in the contextual channel as contextual exposure bias. Importantly, this is a context-level mismatch: even if the model is trained with standard teacher-forcing for the current utterance, the historical context provided to the model at inference can be error-prone and may systematically differ from the oracle context used in training.

In the remainder of the paper, we study how this contextual exposure bias affects contextual ASR in Speech-LLMs, and we develop training strategies that better align training-time context with inference-time conditions.

2.2 Model Architecture

We adopt a training framework consisting of an off-the-shelf speech encoder, a frozen LLM backbone, and a trainable MLP projector for modality adaptation. We fine-tune the LLM with Low-Rank Adaptation (LoRA) (Hu et al.) in two stages: (i) a dedicated SFT LoRA module for SFT, and (ii) a separate DPO LoRA module for preference alignment. Following (Jiang et al., 2025), the DPO stage introduces an additional, distinct LoRA block rather than reusing the SFT adapter, enabling independent control of preference optimization. The overall architecture is illustrated in Figure 1.

2.3 Training Methodology

We study contextual ASR under a realistic train–test discrepancy: during training, the contextual history is typically provided as oracle transcripts, whereas at inference time the history must be obtained from upstream ASR outputs and therefore contains recognition errors. To bridge this gap, we propose a training framework that (i) replaces oracle history with teacher-generated hypotheses to simulate imperfect context during training, where the teacher is a strong off-the-shelf ASR system (instantiated as Whisper large-v3 in our experiments), and (ii) regularizes the model's reliance on history to prevent error amplification when the context is noisy or misleading.

2.3.1 Teacher Error Knowledge

To approximate inference-time conditions, we introduce Teacher Error Knowledge, where the history transcripts $\hat{Y}$ are taken from Whisper large-v3 decoded hypotheses instead of always using oracle history [1]. Concretely, we pre-decode the training set with Whisper and store hypotheses $\hat{Y}_i^{Whisper}$. During training, we construct C_t using Whisper large-v3 decoded hypotheses as the history source. This exposes the model to realistic contextual noise and encourages robustness to history errors.

We optimize the standard sequence negative log-likelihood conditioned on the constructed context:

$$\mathcal{L}_{SFT} = -\sum_{t} \log p_\theta(Y_t|X_t, C_t, P) \qquad (3)$$

where P is the task prompt and θ denotes trainable parameters.

2.3.2 Context Dropout

Even with noisy histories, the model may still over-trust contextual text and become brittle when the history is misleading.To further regularize context usage, we apply Context Dropout during training: with probability p_{drop}, we mask the textual history C_t, while keeping the current utterance speech X_t

[1]Our method does not rely on a specific teacher model.

unchanged:

$$\tilde{C}_t = \begin{cases} \varnothing, & \text{with } p_{\text{drop}}, \\ C_t, & \text{otherwise} \end{cases} \quad (4)$$

The SFT objective is then computed conditioned on $\tilde{C}_t$. This mechanism prevents the model from depending on history as a shortcut and encourages it to retain strong speech information, while still allowing it to benefit from context when available.

2.3.3 Preference Optimization on Hard Negatives with DPO

While SFT with noisy history improves overall robustness, we observe that a small subset of failure cases can still be triggered. To explicitly suppress such undesirable behaviors, we further apply DPO on curated hard negatives.

For each selected utterance, we construct a preference pair (Y^+, Y^-) under the same inputs $\mathbf{X} = (X_t, C_t, P)$, where Y^+ is the ground-truth and Y^- is the inference transcript of the model. DPO optimizes the policy to assign higher likelihood to Y^+ than Y^- without requiring an explicit reward model. Following the standard DPO formulation, we minimize:

$$\Delta_\theta = \log \pi_\theta(Y^+ \mid \mathbf{X}) - \log \pi_\theta(Y^- \mid \mathbf{X}) \quad (5)$$

$$\Delta_r = \log \pi_r(Y^+ \mid \mathbf{X}) - \log \pi_r(Y^- \mid \mathbf{X}) \quad (6)$$

$$m = \beta\,(\Delta_\theta - \Delta_{\text{r}}) \quad (7)$$

$$\mathcal{L}_{\text{DPO}} = -\log \sigma(m) \quad (8)$$

In our experiments, we set the reference policy π_r to the SFT checkpoint (frozen), β is a temperature coefficient that controls the strength of preference sharpening, in this work, we use $\beta = 0.1$, and $\sigma(\cdot)$ is the sigmoid function. Here, Δ_θ and Δ_r measure the log-likelihood gaps between the preferred and dispreferred responses under the current policy π_θ and the reference policy π_r, respectively, and m scales their difference to form the DPO objective (Ouyang et al., 2022; Rafailov et al., 2023).

In our implementation, we use a separate LoRA block for DPO to decouple preference optimization from the main SFT adaptation (Jiang et al., 2025) as in Figure 1.

3 Experiment Setup

3.1 Models and Modules

We build our system on an in-house Speech-LLMs training framework. We adopt the Whisper large-v3 (Radford et al., 2023), discard the decoder and only use the encoder as a feature extractor. Based on previous work on Speech-LLMs, which has shown that LLMs with SFT significantly outperform vanilla pre-trained models in speech recognition tasks (Ma et al., 2025), we choose vicuna-7B-v1.5 (Zheng et al., 2023) with LoRA adapter as the LLM for our system. Unless otherwise specified, we keep the Whisper large-v3 encoder frozen and train a lightweight MLP projector together with LoRA parameters. We report WER as the primary metric.

3.2 Datasets

We conduct both training and in-domain evaluation on the TED-LIUM 3 (Hernandez et al., 2018) dataset following an utterance-level segmentation. Each utterance is paired with its reference transcript, and contextual history is constructed from preceding utterances within the same session. We use the official splits for train, dev and test.

To assess cross-domain generalization, we evaluate on LibriSpeech (Panayotov et al., 2015) without any additional training. We report results on test-clean and test-other.

3.3 Training Protocol

We train our system in three stages. All experiments are conducted on 4 NVIDIA A100 GPUs. In the first Stage, we train a context-free base model by freezing the Whisper large-v3 encoder and Vicuna v1.5-7B, and optimizing only the MLP projector with a standard ASR objective. This stage establishes a reliable alignment between speech and text, which serves as a stable initialization for subsequent contextual training. For this stage, the batch size is set to 6, the learning rate to 1e-4, and the warmup step to 1000.

In the second stage, we perform SFT of the Speech-LLMs using paired speech–text data under the Utterance-level Contextual ASR setting. During SFT, we replace oracle history with these Whisper hypotheses to expose the model to history noise. Importantly, Whisper is only used offline to construct Teacher Error Knowledge for training; our inference doesn't require Whisper as an auxiliary component. To further reduce over-reliance on the history channel, we apply Context Dropout, which randomly masks the provided textual history with probability $p = 0.5$ (and keeps the rest of the training pipeline unchanged). This regularizer forces the model to remain grounded in acoustic information when the history is missing or unreliable.

In this stage, the batch size is set to 6, the learning rate is 1e-4, and the warmup step is 1000.

In the third stage, we refine the model on a subset of challenging examples, termed 'hard negatives', which are identified by decoding the training set using the optimal model from Stage 2 and selecting instances with a WER exceeding 20%. On this specific subset, we conduct a comparative analysis of two refinement strategies: an additional SFT pass and DPO, where the latter utilizes preference pairs to explicitly mitigate error propagation under noisy-history conditions. To implement these strategies, we introduce a dedicated LoRA module distinct from the previous SFT phase while keeping the model backbone frozen. In this stage, for the additional SFT, the batch size is 6; for DPO, the batch size is 2, the accumulate grad is 16. Both strategies use 0 warmup steps and learning rate is 1e-5.

3.4 Inference Details

During the inference phase, we employ beam search decoding to balance generation quality and diversity. Specifically, we set the beam width to 4 and the maximum generation length to 200 tokens. To ensure deterministic and reproducible evaluation, we disable sampling strategies (setting do_sample=False, top_p=1.0, and temperature=1.0). We also maintain neutral penalty settings with both repetition_penalty and length_penalty set to 1.0.

Inference-time LoRA strength adjustment. We introduce an explicit inference-time strength factor γ to control the contribution of the DPO LoRA adapter, while keeping the LoRA rank and scaling hyperparameters fixed across training and inference.

LoRA formulation. Let W_{LLM} denote the frozen base LLM weights. We attach two LoRA adapters: an SFT adapter and a DPO adapter. The resulting effective weight matrix used in the forward pass is:

$$W = W_{\text{LLM}} + \frac{\alpha}{r}\,\Delta W_{\text{SFT}} + \gamma\,\frac{\alpha'}{r'}\,\Delta W_{\text{DPO}} \quad (9)$$

where ΔW_{SFT} and ΔW_{DPO} are low-rank updates (e.g., $\Delta W = BA$) learned in the SFT and DPO stages, respectively, and $\frac{\alpha}{r}$ (resp., $\frac{\alpha'}{r'}$) controls the magnitude of the injected LoRA features (Hu et al.). In our implementation, we use the same rank and scaling for both adapters, i.e.,

$$r = r', \qquad \alpha = \alpha' \quad (10)$$

specifically, $r = 8$ and $\alpha = 32$, so the only additional degree of freedom for the DPO adapter at inference is γ.

Training vs. inference. During DPO training, we set $\gamma = 1$, so the DPO adapter is applied with its full strength:

$$W_{\text{train}} = W_{\text{LLM}} + \frac{\alpha}{r}\,\Delta W_{\text{SFT}} + \frac{\alpha}{r}\,\Delta W_{\text{DPO}} \quad (11)$$

At inference, we keep r and α unchanged and instead tune $\gamma < 1$ to soften the effect of the DPO adapter:

$$W_{\text{infer}} = W_{\text{LLM}} + \frac{\alpha}{r}\,\Delta W_{\text{SFT}} + \gamma\,\frac{\alpha}{r}\,\Delta W_{\text{DPO}} \quad (12)$$

In our experiments, we set $\gamma = 0.25$, which effectively reduces the DPO adapter strength by $4\times$ compared to training, while leaving the learned parameters and all other LoRA hyperparameters intact.

Why adjust γ at inference. Empirically, we observe that the DPO-tuned adapter can be overly aggressive during decoding, occasionally exhibiting reward over-optimization (i.e., the model exploits preference signals at the expense of coherent and faithful generation) (Gao et al., 2023). Introducing γ provides a simple and stable knob to control the influence of DPO at test time: lowering γ mitigates over-optimization and helps preserve the generalization and fluency of the underlying base model, while retaining most of the robustness gains from preference alignment.

4 Result and Analysis

4.1 Main Result

We verify the effectiveness of our proposed framework on the TED-LIUM 3 dataset. Table 1 presents the WER comparisons between baselines and our proposed methods. To ensure a fair comparison under realistic deployment conditions, we primarily focus on the settings where Con_{inf} is hypotheses, meaning the model must rely on predicted history rather than oracle transcripts.

Impact of Context and Training Strategies: The comparison with the baseline without context (7.89%) reveals that incorporating context does not inherently guarantee performance improvements; it depends heavily on the training strategy. As shown in Table 1, the configuration with $\text{Con}_{\text{train}}$ as Whisper large-v3 generated context without context dropout (0 Dropout) yields a WER of 8.15%

N	**Con$_{inf}$/Con$_{train}$**	**0 Dropout WER (%)↓**				**0.5 Dropout WER (%)↓**			
		TED	**Test-clean**	**Test-other**	**LS-Ave.**	**TED**	**Test-clean**	**Test-other**	**LS-Ave.**
0	- / -	7.89	4.79	9.83	7.310	-	-	-	-
	GT / GT	5.6	4.49	10.36	7.425	7.89	4.31	9.68	6.995
	hyp / GT	5.85	**4.54**	10.63	7.585	7.47	4.74	9.94	7.340
1	hyp / Whisper	**5.62**	4.67	**9.46**	**7.065**	7.21	5.37	9.96	7.665
	+ DPO	5.69	4.71	9.57	7.140	**5.32**	**4.56**	9.38	**6.970**
	+ SFT2	5.76	4.67	9.49	7.080	7.26	5.14	**9.30**	7.220
	GT / GT	6.73	4.10	8.36	6.230	5.66	4.10	8.37	6.235
	hyp / GT	6.89	4.85	9.88	7.365	5.59	5.15	**9.10**	7.130
2	hyp / Whisper	8.15	5.57	12.00	8.785	5.47	5.14	9.50	7.320
	+ DPO	**5.07**	4.87	**9.51**	**7.190**	**5.17**	**4.84**	9.19	**7.015**
	+ SFT2	6.90	**4.55**	11.17	7.860	6.10	5.43	9.66	7.545
	GT / GT	7.35	4.24	8.29	6.265	10.42	4.89	10.36	7.625
	hyp / GT	7.05	**5.03**	10.68	7.855	12.62	5.28	10.93	8.105
3	hyp / Whisper	10.06	5.36	10.69	8.025	7.87	5.93	10.39	8.160
	+ DPO	**5.98**	5.60	**9.96**	**7.780**	**5.18**	**4.73**	**9.36**	**7.045**
	+ SFT2	9.30	5.22	10.49	7.855	8.01	6.11	10.20	8.155
	GT / GT	8.54	4.26	9.01	6.635	9.22	**4.75**	10.23	7.490
	hyp / GT	7.74	4.87	11.07	7.970	10.87	**4.75**	10.23	7.490
4	hyp / Whisper	87.37	**4.66**	10.81	7.735	7.81	**4.75**	9.82	7.285
	+ DPO	**4.93**	4.79	**9.97**	**7.380**	**5.69**	4.79	**9.25**	**7.020**
	+ SFT2	113.95	4.90	11.34	8.120	9.16	**4.75**	9.83	7.290
	GT / GT	8.72	5.46	9.49	7.475	9.57	4.90	10.04	7.470
	hyp / GT	10.34	5.08	10.76	7.920	8.19	5.36	11.29	8.325
5	hyp / Whisper	135.57	**4.59**	10.87	7.730	8.5	4.95	9.33	7.140
	+ DPO	**5.34**	4.67	**9.85**	**7.260**	**4.96**	**4.55**	**9.24**	**6.895**
	+ SFT2	72.55	4.90	10.57	7.735	8.51	5.23	9.33	7.280

Table 1: WER comparison on TED-LIUM 3 and out-of-domain Librispeech dataset across different context window sizes (N). The column **Con$_{inf}$ / Con$_{train}$** specifies the source of history used during inference and training, respectively. **hyp** denotes using the model's own predictions as history during inference. Regarding training configuration, **GT** uses ground-truth history, while **Whisper** indicates the model was trained using context decoded by Whisper to simulate historical errors. **+ DPO** and **+ SFT2** are additional fine-tuning stages applied to the SFT model.

γ	**TED-LIUM 3 (WER %) ↓**			**LibriSpeech (WER %) ↓**		
	Attacks/o	**Attacks/w**	**Gap ↓**	**Test-clean**	**Test-other**	**Ave.**
0	5.47	7.93	2.46	5.14	9.50	7.320
0.0625	5.37	7.13	1.76	5.12	9.31	7.215
0.125	5.11	5.76	0.65	5.02	9.53	7.275
0.1875	**5.06**	5.69	0.63	**4.70**	**9.08**	**6.890**
0.25	5.17	**5.63**	0.46	4.84	9.19	7.015
0.375	5.55	5.73	**0.18**	4.85	9.63	7.240
0.5	8.39	8.67	0.28	6.44	12.14	9.290
0.625	53.26	57.15	3.89	27.11	28.96	28.035

Table 2: Impact of DPO LoRA scaling factor (γ) during inference. TED-LIUM 3 Gap denotes the WER degradation caused by irrelevant context attacks. "Attacks/o" refers to relevant context inference, while "Attacks/w" refers to irrelevent context randomly selected from the test set.

at $N = 2$, which underperforms the context-free baseline. This degradation indicates that without regularization, the model may overfit to noisy history or become overly reliant on it. In contrast, the model trained with 0.5 Dropout effectively leverages the context. Under the same configuration ($N = 2$, Whisper history), it achieves a WER of 5.47%, significantly surpassing the No Context baseline. Consequently, we identify the model trained with Whisper large-v3 generated context and 0.5 dropout model as our best-performing SFT model, which serves as the robust baseline for subsequent optimization.

Sensitivity to Regularization and Exposure Bias: We analyze how different training data sources affect the model's sensitivity to context dropout. As shown in Table 1, the baseline trained with ground-truth context ($\text{Con}_{\text{train}} = \text{GT}$) exhibits an inconsistent preference for dropout rates depending on the context length. Specifically, for $N = 1, 3, 4$, the model performs better without dropout (0 Dropout), whereas for $N = 2, 5$, it requires 0.5 Dropout to achieve optimal results. This fluctuation suggests that models with $\text{Con}_{\text{train}} =$ GT are structurally unstable, requiring context-length-specific hyperparameter tuning to avoid performance degradation (e.g., using the wrong 0.5 Dropout at $N = 3$ leads to a high WER of 12.62%). In contrast, our model using Whisper training context ($\text{Con}_{\text{train}} = \text{Whisper}$) shows a clear and consistent pattern for sequential tasks. While 0 Dropout is preferred for the shortest context ($N = 1$), for all multi-turn scenarios ($N \geq 2$), 0.5 Dropout consistently yields superior performance and is essential for stability. Comparing the optimal configurations, our method ($\text{Con}_{\text{train}} =$ Whisper with 0.5 Dropout) achieves the best overall performance at $N = 2$ (5.47% vs. GT's best 5.59%). Furthermore, unlike the ground-truth baseline which oscillates between needing and rejecting regularization, our method provides a reliable strategy ($p = 0.5$) that ensures robustness across varying context lengths, avoiding the catastrophic failures seen in unregularized Whisper models.

Improvement with DPO: To further suppress recognition errors, we refine each SFT checkpoint using the "hard negatives" mined by the same strategy, comparing additional SFT (SFT2) against DPO. Table 1 shows that SFT2 is not a reliable refinement strategy: it often provides negligible gains and frequently degrades performance (e.g., at $N = 2$ with 0.5 dropout, TED WER increases from 5.47% to 6.10%). In contrast, DPO consistently improves WER across almost all configurations (9 out of 10), with particularly large gains under longer context windows where error accumulation is more severe (e.g., 7.87%→5.18% at $N = 3$, 0.5 dropout; 8.50%→4.96% at $N = 5$, 0.5 dropout). We attribute this to the fact that SFT only forces the model to mimic the ground truth, whereas DPO explicitly optimizes the model to prefer the ground truth over its erroneous hypotheses, yielding more stable improvements.

4.2 Cross-Domain Generalization

To assess out-of-domain generalization, we evaluate all models on LibriSpeech (test-clean/test-other) in a zero-shot setting using utterance-level decoding, where the inference history is formed by the model's own hypotheses. As shown in Table 1, SFT with noisy history does not reliably improve cross-domain transfer: the *hyp/Whisper* SFT model at N=2 with 0.5 dropout achieves 7.32% LS-Ave., which is essentially on par with the no-context baseline (7.31%), and increasing the context window can even degrade performance (e.g., 8.16% at N=3, 0.5 dropout).

In contrast, DPO consistently improves robustness across multiple context window sizes. For example, DPO reduces LS-Ave. from 7.32% to 7.015% at N=2 (0.5 dropout), and from 8.16% to 7.045% at N=3 (0.5 dropout). The best out-of-domain result is obtained with N=5 and 0.5 dropout, where DPO achieves 6.895% LS-Ave., outperforming both the no-context baseline and the corresponding SFT baselines. This finding suggests that DPO effectively mitigates domain-specific overfitting. By penalizing the repetition of history errors and hallucinations through preference learning, the model learns to better balance contextual cues with acoustic evidence, leading to superior generalization on unseen out-of-domain data.

4.3 Impact of DPO Inference Scaling

To determine the optimal inference strategy for the DPO module, we analyze the sensitivity of the LoRA scaling factor γ. As shown in Table 2, there is a critical trade-off derived from the DPO signal strength. While $\gamma = 0.1875$ yields the best clean accuracy, increasing γ to 0.25 significantly enhances robustness, reducing the performance drop under attack (TED Gap) to 0.46%. This confirms that a stronger DPO weight helps the model

N	Con_{inf} / $\text{Con}_{\text{train}}$	0 Dropout WER (%)↓			0.5 Dropout WER (%)↓		
		Attacks/o	Attacks/w	Gap ↓	Attacks/o	Attacks/w	Gap ↓
1	hyp / GT	5.85	8.32	2.47	7.47	8.82	1.35
	hyp / Whisper	**5.62**	6.27	0.65	7.21	7.09	**-0.12**
	+ DPO	5.69	**5.5**	**-0.19**	**5.32**	**5.23**	-0.09
2	hyp / GT	6.89	8.43	1.54	5.59	9.23	3.46
	hyp / Whisper	8.15	10.37	2.22	5.47	7.93	2.46
	+ DPO	**5.07**	**6.59**	**1.52**	**5.17**	**5.63**	**0.46**
3	hyp / GT	7.05	7.64	0.59	12.62	10.19	**-2.43**
	hyp / Whisper	10.06	11.18	1.12	7.87	9.53	1.66
	+ DPO	**5.98**	**6.24**	**0.26**	**5.18**	**5.31**	0.13
4	hyp / GT	7.74	9.75	2.01	10.87	13.15	2.28
	hyp / Whisper	87.37	8.8	**-78.57**	7.81	8.82	**1.01**
	+ DPO	**4.93**	**5.44**	0.51	**5.69**	**7.58**	1.89
5	hyp / GT	10.34	11.83	1.49	8.19	10.41	2.22
	hyp / Whisper	135.57	10.74	**-124.83**	8.5	11.34	2.84
	+ DPO	**5.34**	**6.20**	0.86	**4.96**	**5.51**	**0.55**

Table 3: Robustness analysis against irrelevant context attacks on TED-LIUM 3. We evaluate model resilience by replacing the historical context with randomly sampled irrelevant context.

reject misleading context. However, excessive scaling ($\gamma > 0.375$) causes reward over-optimization. We therefore select $\gamma = 0.25$ to prioritize stability and error suppression.

4.4 Robustness to Irrelevant Context

To verify whether the model genuinely comprehends the contextual information, we conduct a robustness attack test. In this experiment, we replace the ground-truth history with semantically irrelevant context randomly sampled from the TED dataset during inference. A robust contextual ASR model should be able to identify the irrelevance of the context and back off to the acoustic signal, thereby minimizing performance degradation.

We compare the performance degradation of the proposed method against the standard oracle-trained baseline. As reported in Table 3, the DPO optimized model exhibits remarkable resilience. Across all context window sizes ($N = [1, 5]$) and both dropout settings (0 and 0.5), although the degradation gap (Attacks/w − Attacks/o) of DPO is not always the smallest, the DPO refined model consistently achieves the lowest attacked WER. Overall, these results suggest that incorporating noisy teacher history and context dropout, together with DPO refinement, helps suppress performance degradation when the provided history is unreliable.

4.5 Data Selection and Inference Scaling for DPO

We conducted an ablation study on the WER selection threshold (based on the $N = 2$, 0.5 dropout SFT model). As shown in Table 4 in Appendix A, our method is remarkably robust to data strictness, yielding consistent gains across all thresholds without requiring precise tuning. Furthermore, the optimal inference scaling γ remains stable in all of these configurations, consistent with Sec. 4.3. This shows that the optimal inference strategy is effectively decoupled from the data curation process, significantly simplifying deployment.

5 Conclusion

In this work, we address contextual exposure bias in Speech-LLMs by proposing a unified training framework that integrates Teacher Error Knowledge, context dropout, and DPO. We demonstrate that training with Teacher Error Knowledge effectively bridges the train-test gap, significantly outperforming oracle-trained baselines. Crucially, context dropout proves to be a decisive factor for stability; it prevents the model from over-relying on textual history. Furthermore, DPO explicitly suppresses error propagation, yielding superior in-domain performance and robust cross-domain generalization. Collectively, our approach establishes a reliable paradigm for long-form ASR under realistic, imperfect conditions.

Limitations

While our proposed framework effectively mitigates contextual exposure bias and enhances robustness in Speech-LLMs, several limitations remain to be addressed in future work.

First, Multi-Speaker Overlap. Our current experimental setup assumes sequential, turn-based speech (as found in TED-LIUM 3 and LibriSpeech). We have not evaluated the model's performance in scenarios involving overlapping speech or "cocktail party" environments. Since Speech-LLMs process audio as a single sequence, handling simultaneous speakers with heavy overlap likely requires specialized architectural modifications or data curation strategies that are beyond the scope of this work.

Second, Limited Diversity of Teacher Error Sources. Although Teacher Error Knowledge is a model-agnostic concept that can, in principle, be instantiated with hypotheses from arbitrary ASR systems, our current implementation relies on a single teacher model (Whisper large-v3) to generate erroneous contextual knowledge. Consequently, the simulated errors exposed during training are biased toward the error characteristics of this teacher, and may not exhaustively represent the diverse failure modes of Speech-LLMs or the wide range of acoustic distortions encountered in unconstrained environments.

References

Petar Aleksic, Mohammadreza Ghodsi, Assaf Michaely, Cyril Allauzen, Keith Hall, Brian Roark, David Rybach, and Pedro Moreno. 2015. Bringing contextual information to google speech recognition. In *Proc. Interspeech 2015*, pages 468–472.

William Chan, Navdeep Jaitly, Quoc Le, and Oriol Vinyals. 2016. Listen, attend and spell: A neural network for large vocabulary conversational speech recognition. In *2016 IEEE international conference on acoustics, speech and signal processing (ICASSP)*, pages 4960–4964. IEEE.

Zhehuai Chen, He Huang, Andrei Andrusenko, Oleksii Hrinchuk, Krishna C Puvvada, Jason Li, Subhankar Ghosh, Jagadeesh Balam, and Boris Ginsburg. 2024. Salm: Speech-augmented language model with in-context learning for speech recognition and translation. In *ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 13521–13525. IEEE.

Jian Cheng. 2024. Context-aware speech recognition using prompts for language learners. *Interspeech 2024*, pages 4009–4013.

Wenqian Cui, Dianzhi Yu, Xiaoqi Jiao, Ziqiao Meng, Guangyan Zhang, Qichao Wang, Yiwen Guo, and Irwin King. 2024. Recent advances in speech language models: A survey. *arXiv preprint arXiv:2410.03751*.

Yangui Fang, Jing Peng, Yu Xi, Xu Li, Haoyu Li, Chengwei Zhang, Guohui Zhong, and Kai Yu. 2025. Joint decoding method for controllable contextual speech recognition based on speech llm. *arXiv preprint arXiv:2508.08585*.

Leo Gao, John Schulman, and Jacob Hilton. 2023. Scaling laws for reward model overoptimization. In *International Conference on Machine Learning*, pages 10835–10866. PMLR.

Xun Gong, Anqi Lv, Zhiming Wang, and Yanmin Qian. 2024. Contextual biasing speech recognition in speech-enhanced large language model. *Proc. Interspeech. ISCA*, pages 257–261.

Aditya Gourav, Jari Kolehmainen, Prashanth Shivakumar, Yile Gu, Grant Strimel, Ankur Gandhe, Ariya Rastrow, and Ivan Bulyko. 2024. Multi-modal retrieval for large language model based speech recognition. In *Findings of the Association for Computational Linguistics ACL 2024*, pages 4435–4446.

Alex Graves, Santiago Fernández, Faustino Gomez, and Jürgen Schmidhuber. 2006. Connectionist temporal classification: labelling unsegmented sequence data with recurrent neural networks. In *Proceedings of the 23rd international conference on Machine learning*, pages 369–376.

Alex Graves, Abdel-rahman Mohamed, and Geoffrey Hinton. 2013. Speech recognition with deep recurrent neural networks. In *2013 IEEE international conference on acoustics, speech and signal processing*, pages 6645–6649. Ieee.

Yachao Guo, Zhibin Qiu, Hao Huang, and Chng Eng Siong. 2023. Improved keyword recognition based on aho-corasick automaton. In *2023 International Joint Conference on Neural Networks (IJCNN)*, pages 1–7. IEEE.

Keith Hall, Eunjoon Cho, Cyril Allauzen, Françoise Beaufays, Noah Coccaro, Kaisuke Nakajima, Michael Riley, Brian Roark, David Rybach, and Linda Zhang. 2015. Composition-based on-the-fly rescoring for salient n-gram biasing. In *Proc. Interspeech 2015*, pages 1418–1422.

François Hernandez, Vincent Nguyen, Sahar Ghannay, Natalia Tomashenko, and Yannick Estève. 2018. Tedlium 3: Twice as much data and corpus repartition for experiments on speaker adaptation. In *Speech and Computer*, pages 198–208. Springer International Publishing.

Takaaki Hori, Niko Moritz, Chiori Hori, and Jonathan Le Roux. 2020. Transformer-based long-context end-to-end speech recognition. In *Interspeech*, pages 5011–5015.

Junfeng Hou, Jinkun Chen, Wanyu Li, Yufeng Tang, Jun Zhang, and Zejun Ma. 2022. Bring dialogue-context into rnn-t for streaming asr. In *INTERSPEECH*, pages 2048–2052.

Wei-Ning Hsu, Benjamin Bolte, Yao-Hung Hubert Tsai, Kushal Lakhotia, Ruslan Salakhutdinov, and Abdelrahman Mohamed. 2021. Hubert: Self-supervised speech representation learning by masked prediction of hidden units. *IEEE/ACM transactions on audio, speech, and language processing*, 29:3451–3460.

Edward J Hu, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, Weizhu Chen, and 1 others. Lora: Low-rank adaptation of large language models. In *International Conference on Learning Representations*.

Ailin Huang, Boyong Wu, Bruce Wang, Chao Yan, Chen Hu, Chengli Feng, Fei Tian, Feiyu Shen, Jingbei Li, Mingrui Chen, and 1 others. 2025. Step-audio: Unified understanding and generation in intelligent speech interaction. *CoRR*.

Ruizhe Huang, Mahsa Yarmohammadi, Sanjeev Khudanpur, and Daniel Povey. 2024a. Improving neural biasing for contextual speech recognition by early context injection and text perturbation. In *Proc. Interspeech 2024*, pages 752–756.

Zhiqi Huang, Diamantino Caseiro, Kandarp Joshi, Christopher Li, Pat Rondon, Zelin Wu, Petr Zadrazil, and Lillian Zhou. 2024b. Optimizing large-scale context retrieval for end-to-end asr. *Proc. Interspeech. ISCA*, pages 4573–4577.

Amir Hussein, Brian Yan, Antonios Anastasopoulos, Shinji Watanabe, and Sanjeev Khudanpur. 2024. Enhancing end-to-end conversational speech translation through target language context utilization. In *ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 11971–11975. IEEE.

Mahaveer Jain, Gil Keren, Jay Mahadeokar, Geoffrey Zweig, Florian Metze, and Yatharth Saraf. 2020. Contextual rnn-t for open domain asr. In *Proc. Interspeech 2020*, pages 11–15.

Lifan Jiang, Boxi Wu, Jiahui Zhang, Xiaotong Guan, and Shuang Chen. 2025. Huvidpo: Enhancing video generation through direct preference optimization for human-centric alignment. *arXiv preprint arXiv:2502.01690*.

Konstantin Kolokolov, Pavel Pekichev, and Karthik Raghunathan. 2024. Self-consistent context aware conformer transducer for speech recognition. *arXiv preprint arXiv:2402.06592*.

Roman Koshkin, Katsuhito Sudoh, and Satoshi Nakamura. 2024. Llms are zero-shot context-aware simultaneous translators. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pages 1192–1207.

Egor Lakomkin, Chunyang Wu, Yassir Fathullah, Ozlem Kalinli, Michael L Seltzer, and Christian Fuegen. 2024. End-to-end speech recognition contextualization with large language models. In *ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 12406–12410. IEEE.

Wonjun Lee, San Kim, and Gary Geunbae Lee. 2024. Enhancing dialogue speech recognition with robust contextual awareness via noise representation learning. In *Proceedings of the 25th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, pages 333–343.

Zhihong Lei, Xingyu Na, Mingbin Xu, Ernest Pusateri, Christophe Van Gysel, Yuanyuan Zhang, Shiyi Han, and Zhen Huang. 2025. Contextualization of asr with llm using phonetic retrieval-based augmentation. In *ICASSP 2025-2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1–5. IEEE.

Jinyu Li and 1 others. 2022. Recent advances in end-to-end automatic speech recognition. *APSIPA Transactions on Signal and Information Processing*, 11.

Shaojun Li, Hengchao Shang, Daimeng Wei, Jiaxin Guo, Zongyao Li, Xianghui He, Min Zhang, and Hao Yang. 2024. La-rag: Enhancing llm-based asr accuracy with retrieval-augmented generation. *CoRR*.

Ziyang Ma, Guanrou Yang, Yifan Yang, Zhifu Gao, Jiaming Wang, Zhihao Du, Fan Yu, Qian Chen, Siqi Zheng, Shiliang Zhang, and 1 others. 2025. Speech recognition meets large language model: Benchmarking, models, and exploration. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pages 24840–24848.

Erik McDermott, Hasim Sak, and Ehsan Variani. 2019. A density ratio approach to language model fusion in end-to-end automatic speech recognition. In *2019 IEEE Automatic Speech Recognition and Understanding Workshop (ASRU)*, pages 434–441. IEEE.

Bingshen Mu, Hexin Liu, Hongfei Xue, Kun Wei, and Lei Xie. 2025. Hearing more with less: Multi-modal retrieval-and-selection augmented conversational llm-based asr. *arXiv preprint arXiv:2508.01166*.

Long Ouyang, Jeffrey Wu, Xu Jiang, Diogo Almeida, Carroll Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, and 1 others. 2022. Training language models to follow instructions with human feedback. *Advances in neural information processing systems*, 35:27730–27744.

Vassil Panayotov, Guoguo Chen, Daniel Povey, and Sanjeev Khudanpur. 2015. Librispeech: an asr corpus based on public domain audio books. In *2015 IEEE international conference on acoustics, speech and signal processing (ICASSP)*, pages 5206–5210. IEEE.

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

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

Vijay Ravi, Yile Gu, Ankur Gandhe, Ariya Rastrow, Linda Liu, Denis Filimonov, Scott Novotney, and Ivan Bulyko. 2020. Improving accuracy of rare words for rnn-transducer through unigram shallow fusion. *arXiv preprint arXiv:2012.00133*.

Peng Shen, Xugang Lu, and Hisashi Kawai. 2025. Retrieval-augmented speech recognition approach for domain challenges. *arXiv preprint arXiv:2502.15264*.

Xian Shi, Yexin Yang, Zerui Li, Yanni Chen, Zhifu Gao, and Shiliang Zhang. 2024. Seaco-paraformer: A non-autoregressive asr system with flexible and effective hotword customization ability. In *ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 10346–10350. IEEE.

Yui Sudo, Yosuke Fukumoto, Muhammad Shakeel, Yifan Peng, and Shinji Watanabe. 2024. Contextualized automatic speech recognition with dynamic vocabulary. In *2024 IEEE Spoken Language Technology Workshop (SLT)*, pages 78–85. IEEE.

Jiyang Tang, Kwangyoun Kim, Suwon Shon, Felix Wu, and Prashant Sridhar. 2024. Improving asr contextual biasing with guided attention. In *ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 12096–12100. IEEE.

Shubham Toshniwal, Anjuli Kannan, Chung-Cheng Chiu, Yonghui Wu, Tara N Sainath, and Karen Livescu. 2018. A comparison of techniques for language model integration in encoder-decoder speech recognition. In *2018 IEEE spoken language technology workshop (SLT)*, pages 369–375. IEEE.

Mengzhi Wang, Shifu Xiong, Genshun Wan, Hang Chen, Jianqing Gao, and Lirong Dai. 2024. Deep clas: Deep contextual listen, attend and spell. *arXiv preprint arXiv:2409.17603*.

Guanrou Yang, Ziyang Ma, Fan Yu, Zhifu Gao, Shiliang Zhang, and Xie Chen. 2024. Mala-asr: Multimedia-assisted llm-based asr. *CoRR*.

Saierdaer Yusuyin, Te Ma, Hao Huang, Wenbo Zhao, and Zhijian Ou. 2025. Whistle: Data-efficient multilingual and crosslingual speech recognition via weakly phonetic supervision. *IEEE Transactions on Audio, Speech and Language Processing*.

Lianmin Zheng, Wei-Lin Chiang, Ying Sheng, Siyuan Zhuang, Zhanghao Wu, Yonghao Zhuang, Zi Lin, Zhuohan Li, Dacheng Li, Eric Xing, and 1 others. 2023. Judging llm-as-a-judge with mt-bench and chatbot arena. *Advances in neural information processing systems*, 36:46595–46623.

Yue Zhou, Yuxuan Yuan, Chengwei Zhang, and Xiaodong Shi. 2025. Boosting context-aware speech translation with large language models. *IEEE Signal Processing Letters*.

A Data Selection and Inference Scaling for DPO

WER (%) Threshold	γ	TED-LIUM 3 (WER %) ↓			LibriSpeech (WER %) ↓		
		Attacks/o	Attacks/w	Gap ↓	Test-clean	Test-other	Ave.
5	0	5.47	7.93	2.46	5.14	9.50	7.320
	0.0625	5.40	7.02	1.62	5.11	**9.32**	**7.215**
	0.125	5.43	7.34	1.91	5.18	9.41	7.295
	0.1875	**5.04**	6.76	1.72	**5.09**	9.70	7.395
	0.25	5.23	**5.99**	0.76	5.18	9.51	7.345
	0.375	6.21	6.42	**0.21**	5.49	10.19	7.840
	0.5	9.18	9.63	0.45	8.58	14.65	11.615
	0.625	50.40	50.97	0.57	69.98	70.66	70.320
10	0	5.47	7.93	2.46	5.14	9.50	7.320
	0.0625	5.40	7.10	1.70	5.11	**9.30**	7.205
	0.125	5.32	6.91	1.59	5.19	9.61	7.400
	0.1875	**5.04**	6.63	1.59	5.04	9.67	7.355
	0.25	5.27	**6.00**	0.73	**4.62**	9.50	**7.060**
	0.375	6.29	6.38	**0.09**	5.06	10.22	7.640
	0.5	10.22	10.87	0.65	8.65	14.96	11.805
	0.625	85.55	89.97	4.42	88.93	85.49	87.210
15	0	5.47	7.93	2.46	5.14	9.50	7.320
	0.0625	5.40	7.43	2.03	5.12	9.30	7.210
	0.125	5.15	6.63	1.48	5.21	9.58	7.395
	0.1875	5.12	5.91	0.79	4.98	9.46	7.220
	0.25	**5.07**	**5.67**	**0.60**	**4.77**	**9.27**	**7.020**
	0.375	5.79	6.14	0.35	5.04	9.84	7.440
	0.5	8.27	8.84	0.57	6.99	12.89	9.940
	0.625	33.58	35.59	2.01	22.72	27.96	25.340
20	0	5.47	7.93	2.46	5.14	9.50	7.320
	0.0625	5.37	7.13	1.76	5.12	9.31	7.215
	0.125	5.11	5.76	0.65	5.02	9.53	7.275
	0.1875	**5.06**	5.69	0.63	**4.70**	**9.08**	**6.890**
	0.25	5.17	**5.63**	0.46	4.84	9.19	7.015
	0.375	5.55	5.73	**0.18**	4.85	9.63	7.240
	0.5	8.39	8.67	0.28	6.44	12.14	9.290
	0.625	53.26	57.15	3.89	27.11	28.96	28.035
25	0	5.47	7.93	2.46	5.14	9.50	7.320
	0.0625	5.39	7.40	2.01	5.13	9.33	7.230
	0.125	5.28	6.00	0.72	5.21	9.57	7.390
	0.1875	**5.08**	5.68	0.60	**4.70**	**9.18**	**6.940**
	0.25	5.28	**5.34**	**0.06**	4.84	9.33	7.085
	0.375	5.74	6.04	0.30	5.12	10.01	7.565
	0.5	10.26	10.50	0.24	8.25	14.64	11.445
	0.625	31.01	30.48	-0.53	28.03	35.84	31.935
30	0	5.47	7.93	2.46	5.14	9.50	7.320
	0.0625	5.32	7.38	2.06	5.14	**9.28**	7.210
	0.125	5.12	6.06	0.94	5.14	9.48	7.310
	0.1875	5.00	5.51	0.51	**4.70**	9.54	7.120
	0.25	**4.97**	5.39	0.42	4.74	**9.28**	**7.010**
	0.375	**4.97**	**5.20**	**0.23**	4.83	9.33	7.080
	0.5	5.49	5.78	0.29	5.11	10.00	7.555
	0.625	8.77	9.11	0.34	8.16	14.73	11.445

Table 4: Impact of Hard Negatives threshold and DPO LoRA scaling factor (γ) during inference. TED-LIUM 3 Gap denotes the WER degradation caused by irrelevant context attacks.