Article

Not peer-reviewed version

HCR: Hierarchical Collaborative Reasoning with Interactive Distillation and Swarm Reinforcement for Chinese Spelling Correction

Haoxiang Qi , Wenzhe Zhao , Guangyao Zhou , Yang Zhao [*]

Posted Date: 9 March 2026

doi: 10.20944/preprints202603.0580.v1

Keywords: Chinese spelling correction; interactive knowledge distillation; swarm reinforcement; debate-enhanced arbitration

doi:10.20944/preprints202603.0580.v1

Article

HCR: Hierarchical Collaborative Reasoning with Interactive Distillation and Swarm Reinforcement for Chinese Spelling Correction

Haoxiang Qi [1], Wenzhe Zhao [1], Guangyao Zhou [1] and Yang Zhao [2,*]

[1] The Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100190, China
[2] Peking University Health Science Center, Beijing 100191, China
* Correspondence: 2311210031@stu.pku.edu.cn

Abstract

Chinese spelling correction (CSC) remains challenging due to heterogeneous error types and domain-dependent variations. We propose HCR, a hierarchical collaborative reasoning framework that integrates interactive knowledge distillation, swarm reinforcement collaboration, and debate-enhanced arbitration into a unified multi-agent architecture. By specializing agents in orthographic, phonetic, and semantic reasoning and enabling adaptive collaboration, HCR effectively disentangles complex dependencies and refines predictions through iterative consensus. Extensive experiments on three public benchmarks and a real-world medical dataset demonstrate that HCR achieves state-of-the-art performance on both detection and correction tasks and exhibits strong robustness under domain shifts, establishing a solid foundation for advancing interpretable, adaptive, and generalizable collaborative reasoning in CSC.:

Keywords*:* Chinese spelling correction; interactive knowledge distillation; swarm reinforcement; debate-enhanced arbitration

1. Introduction

Chinese Spelling Correction (CSC) is a fundamental Natural Language Processing (NLP) task that ensures textual integrity and semantic consistency in downstream applications such as intelligent writing assistants (Yang, 2025), search engines (Wang et al., 2024b), and educational platforms (Liu et al., 2025b). Beyond its practical significance (Liu et al., 2025a; Li et al., 2025a), CSC serves as an essential foundation for language understanding, error-tolerant communication, and text standardization in large-scale knowledge systems. Nevertheless, the task remains intrinsically challenging due to the heterogeneous and overlapping nature of Chinese spelling errors (Liu et al., 2021). These errors arise from phonetic confusions between homophones, graphemic similarities among visually close characters, and semantic inconsistencies that require long-range contextual reasoning (Li et al., 2024; Zhou et al., 2024). In real-world scenarios, such as web corpora and domain-specific medical records, these error types often co-occur and interact, leading to compounded ambiguities. The coexistence of mixed linguistic patterns and noisy domain shifts further aggravates the difficulty of achieving robust, interpretable, and generalizable correction.

Recent advances in CSC have been propelled by the success of deep neural architectures and Transformer-based pre-trained language models, which provide strong representational capacity and contextual awareness (Xu et al., 2025). These models have significantly improved correction accuracy by leveraging bidirectional attention and large-scale pre-training. However, the majority of existing approaches adopt a single-agent paradigm that jointly handles orthographic, phonetic, and semantic errors (Liu et al., 2024a). Such unified modeling inevitably entangles heterogeneous error distributions, limiting the model's ability to specialize and generalize across diverse linguistic contexts. Detector–corrector frameworks (Zhu et al., 2022; Li et al., 2021) introduce modular

doi:10.20944/preprints202603.0580.v1

decomposition between error detection and correction, yet their collaboration is sequential rather than interactive, and the modules do not maintain explicit role specialization across orthographic, phonetic, and semantic dimensions. Ensemble-style methods and progressive refinement frameworks (Li et al., 2025b) improve robustness through multi-stage learning or prediction fusion, but their coordination is typically implicit, relying on averaging or cascaded optimization without structured inter-agent communication. Reinforcement-driven CSC approaches further introduce reward-based optimization; however, they generally operate with a flat scalar reward and a single optimization objective, lacking hierarchical reward decomposition and role-aware coordination. Consequently, existing CSC systems do not explicitly factorize heterogeneous reasoning roles nor establish a structured arbitration mechanism for conflict resolution. Moreover, while large language models (LLMs) demonstrate remarkable reasoning and adaptation abilities (Achiam et al., 2023; Yang et al., 2025; Guo et al., 2025), their internal mechanisms remain opaque, making it difficult to interpret or disentangle the complementary signals needed for fine-grained correction. Reinforcement learning techniques have also been introduced into CSC (Huang et al., 2023; Zhang et al., 2023), yet current frameworks typically focus on optimizing independent agents or single reward functions without explicit coordination. The absence of collaborative strategies restricts their capacity to leverage inter-agent synergy and exploit collective reasoning for ambiguous or domain-specific cases.

To address gaps, we propose HCR, a **H**ierarchical **C**ollaborative **R**easoning unified multi-agent architecture that integrates interactive knowledge distillation, a swarm-style cooperative reinforcement learning mechanism, and debate-enhanced arbitration. Within HCR, a large teacher model supervises three lightweight student agents specializing respectively in orthographic, phonetic, and semantic reasoning. Through interactive distillation, each agent acquires complementary expertise while maintaining alignment with global correction objectives. The swarm-style reinforcement component formulates agent interaction as coordinated multi-agent policy optimization under hierarchical rewards, where global rewards promote overall accuracy, fluency, and coherence, while local rewards encourage agent-specific specialization across distinct error categories. To ensure transparent and reliable decision-making, a debate-enhanced arbitration module enables agents to exchange, critique, and iteratively refine their predictions, ultimately achieving consensus through confidence-adaptive fusion.

By unifying specialized reasoning, adaptive cooperation, and interpretable arbitration, HCR advances the field of CSC beyond static single-agent correction. Extensive experiments on three public benchmarks and a real-world medical corpus demonstrate that HCR not only surpasses existing models in both detection and correction accuracy but also exhibits strong robustness under cross-domain shifts. The main contributions are as followed:

- We propose HCR, a hierarchical collaborative reasoning framework for Chinese spelling correction that explicitly factorizes orthographic, phonetic, and semantic reasoning into teacher-guided specialized agents. HCR establishes structured role-aware collaboration within a unified multi-agent architecture.
- We introduce a structured collaborative learning paradigm that jointly induces role specialization, reward-level coordination, and consensus-driven refinement within a unified CSC framework. By aligning reasoning trajectories, decomposing optimization signals hierarchically, and resolving inter-agent disagreement through confidence-aware interaction, the proposed approach establishes principled multi-agent collaboration beyond flat reinforcement optimization and heuristic fusion strategies.
- We conduct extensive experiments on three SIGHAN benchmarks and a real-world medical dataset, achieving pretty performance at both detection and correction levels.

2. Related Work

2.1. Neural Development for Chinese Spelling Correction

CSC has evolved from rule-based text normalization into a comprehensive neural reasoning task that requires phonological, visual, and semantic understanding. Early CSC systems primarily relied on phonetic mapping and dictionary matching techniques, using Pinyin-based edit distance and confusion sets constructed from linguistic statistics (Wu et al., 2013; Tseng et al., 2015). These methods were effective for limited vocabularies but failed to capture contextual dependencies, often producing ambiguous or grammatically inconsistent corrections. With the rise of neural language modeling, CSC research shifted from symbolic rules to data-driven learning. Recurrent and convolutional neural networks introduced sequence-level representations that improved robustness under noisy conditions (Li et al., 2022a; Zhang et al., 2020; Ji et al., 2021). The adoption of Transformer architectures and large-scale pre-trained models, such as BERT and RoBERTa (Liu et al., 2021; Zhang et al., 2021; Xu et al., 2021), further boosted accuracy by enabling global contextual encoding and self-attention over long text sequences. Later works refined these models by incorporating auxiliary tasks, including pronunciation prediction, error type classification, or language modeling, to better handle homophone and shape-similar errors (Guo et al., 2021; Yang & Yu, 2022).

Despite these advancements, current neural CSC systems remain constrained by the single-agent paradigm, which jointly models heterogeneous error sources in a unified embedding space. Such entanglement hinders specialization and interpretability: phonetic confusions require acoustic priors, while graphemic misuses depend on visual similarity; semantic inconsistencies, in contrast, rely on high-level discourse reasoning. Existing PLM-based methods often overfit to data biases and exhibit performance degradation under domain shifts, especially in out-of-distribution contexts such as medical or legal text. Reinforcement learning (RL) techniques were later introduced to introduce adaptive error exploration and reward-guided correction (Liu et al., 2024b; Wang et al., 2024a; Foerster et al., 2016; Lowe et al., 2017), but most frameworks employ a single-agent optimization strategy with a global scalar reward, lacking any notion of collective learning or cooperative negotiation. Consequently, the model struggles to capture multi-perspective reasoning or to establish transparent decision paths, motivating the exploration of collaborative, interpretable, and multi-agent frameworks for CSC.

2.2. Collaborative and Multi-Agent Reasoning Paradigms

The concept of multi-agent collaboration has gained increasing attention across artificial intelligence, offering a structured way to decompose complex reasoning tasks into specialized subproblems (Sunehag et al., 2017; Rashid et al., 2020). In natural language processing, collaborative reasoning frameworks have been applied to question answering, dialogue generation, code synthesis, and reasoning-intensive evaluation, where multiple agents communicate, argue, and refine shared hypotheses (Irving et al., 2018; Wei et al., 2022). These approaches are inspired by cognitive and social theories, positing that collective intelligence emerges through interaction, debate, and self-correction. With the integration of LLMs, multi-agent frameworks have become increasingly capable of managing contextual complexity through role-based specialization and explicit communication protocols (Li et al., 2023; Shinn et al., 2023; Li et al., 2022b). Such designs improve not only task performance but also interpretability, as the reasoning process can be traced through structured debates and consensus formation.

3. Method

3.1. Overview

The HCR framework integrates three core components into a unified architecture for Chinese spelling correction, including multi-agent knowledge distillation, swarm reinforcement collaboration, and debate-enhanced arbitration. A large teacher model guides three specialized student agents to acquire expertise in orthographic similarity, phonetic consistency, and contextual semantics through interactive knowledge transfer. All three student agents are architecturally identical, sharing the same backbone and prediction head, and their specialization is induced purely by different supervision signals and optimization objectives. The distilled agents collaborate to exchange reasoning signals and dynamically adapt their contributions via a swarm reinforcement mechanism, enabling robust and flexible decision-making under diverse error patterns. Finally, a debate-enhanced arbitration module refines predictions by integrating outputs and confidence estimates from all agents, providing transparent and interpretable correction results. Figure 1 shows the overall pipeline.

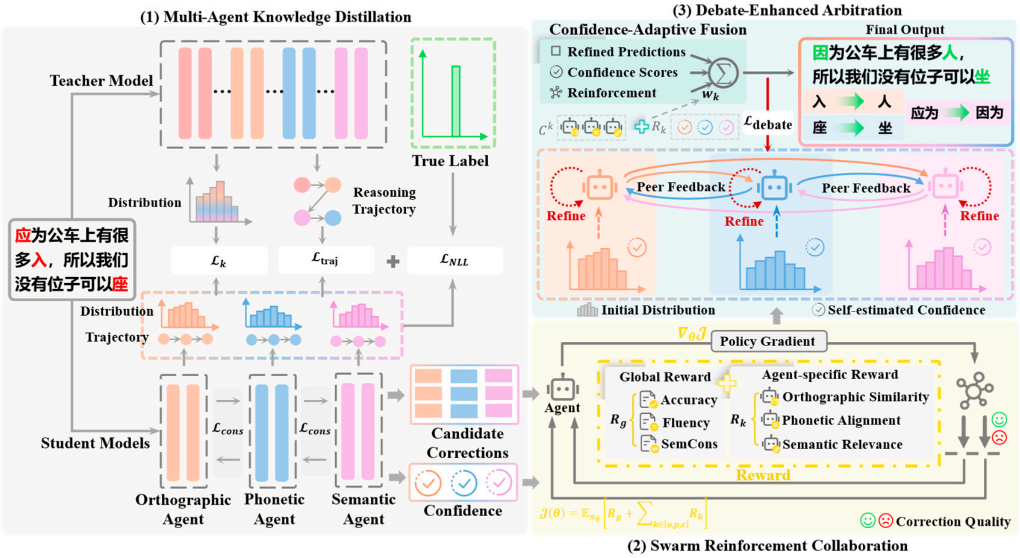

Figure 1. Overview of HCR.

3.2. Multi-Agent Knowledge Distillation

HCR employs a multi-agent knowledge distillation strategy to enable three specialized student agents to acquire complementary reasoning capabilities from a large teacher model while preserving distinct expertise in orthographic similarity, phonetic consistency, and contextual semantics. The three student agents are architecturally identical and share the same backbone and prediction head; their functional specialization does not come from structural differences, but is induced by different supervision signals, distillation targets, and optimization biases. Unlike conventional approaches that directly mimic the teacher's outputs, HCR leverages both predictive distributions and intermediate reasoning trajectories to guide agent-specific learning. Given an input sequence $\mathbf{X} = [x_1, x_2, \ldots, x_n]$, the teacher model $\mathcal{T}$ produces a probability distribution $\boldsymbol{P}^T$ over the candidate vocabulary $\mathcal{V}$. Each student agent $\mathcal{S}_k$, parameterized by θ_k, generates its own distribution $\boldsymbol{P}^k$ under the teacher's supervision.

In addition to the output distribution, the teacher model T is further prompted to generate a structured reasoning trajectory $\mathbf{R}^T = [r_1, r_2, \ldots, r_m]$, which is defined as an ordered token sequence

doi:10.20944/preprints202603.0580.v1

that explicitly encodes the intermediate decision process from the input $\mathbf{X}$ to the final correction. Concretely, $\mathbf{R}^T$ is constructed under a predefined format that represents a sequence of reasoning states, each corresponding to a well-defined diagnostic or decision step, such that the entire sequence forms a complete and self-consistent decision path. This trajectory is generated by the teacher via autoregressive decoding and stored as an intermediate supervision signal.

To ensure effective learning, we incorporate a supervised negative log-likelihood loss that directly optimizes the likelihood of predicting the correct sequence:

$$\mathcal{L}_{\text{NLL}} = -\sum_{i=1}^{n} \log\, P(y_i \mid \mathbf{X}),$$

where $P(y_i \mid \mathbf{X})$ is the final predicted probability for the ground-truth character y_i at position i.

In parallel, the teacher's knowledge is transferred to each student via an agent-specific distillation loss defined by the Kullback-Leibler divergence between softened teacher and student distributions:

$$\mathcal{L}_k = \tau^2 \cdot \mathrm{KL}\left(\sigma\left(\frac{\mathbf{P}^T}{\tau}\right) \parallel \sigma\left(\frac{\mathbf{P}^k}{\tau}\right)\right), \tag{2}$$

where $\sigma(\cdot)$ denotes the softmax function, τ is the temperature parameter controlling distribution smoothness, and $\mathbf{P}^k$ represents the predictive distribution generated by $\mathcal{S}_k$. This formulation allows the student agents to approximate the teacher's knowledge while retaining stable gradients for efficient training.

Beyond output-level distillation, we further introduce a trajectory-level supervision that explicitly transfers the teacher's reasoning process to each student. Given the teacher-generated trajectory $\mathbf{R}^T = [r_1, \dots, r_m]$, each student agent $\mathcal{S}_k$ is trained to model the conditional distribution $P^k(\mathbf{R} \mid \mathbf{X})$. The corresponding trajectory distillation loss is defined as:

$$\mathcal{L}_{\text{traj}} = -\sum_{t=1}^{m} \log P_k(r_t \mid \mathbf{X}, \mathbf{r}_{<t}), \tag{3}$$

which enforces the student to reproduce the same ordered sequence of intermediate decisions as the teacher, thereby aligning not only the final output but also the underlying reasoning process.

To further encourage the agents to specialize in distinct reasoning spaces, we introduce a cross-agent consistency regularization that penalizes redundant representation learning. For hidden states $\mathbf{H}^k$ and $\mathbf{H}^j$ from agents $\mathcal{S}_k$ and $\mathcal{S}_j$, the regularization is defined as:

$$\mathcal{L}_{\text{cons}} = \sum_{k \neq j} \parallel \mathbf{H}^k - \mathbf{H}^j \parallel_2^2, \tag{4}$$

which drives the intermediate representations of different agents toward diverse and complementary subspaces. Finally, we combine the teacher alignment loss, cross-agent regularization, and supervised ground-truth learning into the unified objective for multi-agent knowledge distillation:

$$\mathcal{L}_{\text{MKD}} = \alpha \sum_{k} \mathcal{L}_{k} + \beta \mathcal{L}_{\text{cons}} + \gamma \mathcal{L}_{\text{NLL}} + \delta \sum_{k} \mathcal{L}^{k}{}_{traj}\,, \tag{5}$$

where α, β, γ, and δ are hyperparameters balancing the contributions of each component. This formulation enables each agent to achieve specialization in its designated reasoning domain while collectively contributing to the overall correction performance.

3.3. Swarm Reinforcement Collaboration

While interactive multi-agent distillation enables the student agents to acquire complementary reasoning capabilities, effective collaboration among them remains critical for achieving robust and adaptive correction. To address this challenge, we introduce Swarm Reinforcement Collaboration (SRC), which models the three agents as a cooperative multi-agent reinforcement learning system, metaphorically described as a coordinated swarm, that jointly optimizes correction policies through structured policy updates and shared reward signals to maximize overall correction quality. Instead of treating the agents as independent learners, SRC enables them to dynamically adapt their contributions based on input characteristics and error patterns, resulting in more flexible and reliable decision-making.

Formally, given an input sequence $\mathbf{X}$, three agents $\{\mathcal{S}_o,\ \mathcal{S}_p,\ \mathcal{S}_c\}$ generate candidate corrections and confidence estimates. A global reward R_g evaluates the overall performance of the swarm, encouraging the agents to collaborate toward shared objectives of accuracy, fluency, and semantic consistency:

$$R_g = \lambda_1 \cdot \text{Acc} + \lambda_2 \cdot \text{Fluency} + \lambda_3 \cdot \text{SemCons}, \tag{6}$$

where Acc denotes sentence-level correction accuracy, defined as a binary indicator of whether $\widehat{\mathbf{Y}}$ exactly matches the ground-truth correction $\mathbf{Y}^{*}$; Fluency measures linguistic well-formedness and is computed as the negative length-normalized log-likelihood under a pretrained language model P_{LM}:

$$\text{Fluency}(\hat{Y}) = -\frac{1}{|\hat{Y}|} \sum_{i=1}^{|\hat{Y}|} \log P_{\text{LM}}(\hat{y}_i \mid \hat{y}_{<i})\,, \tag{7}$$

and SemCons evaluates semantic consistency between the input and the correction by the cosine similarity between their sentence embeddings:

$$\text{SemCons}(X, \hat{Y}) = \cos\left(f_{\text{enc}}(X), f_{\text{enc}}(\hat{Y})\right), \tag{8}$$

and λ_1, λ_2 , λ_3 are weighting coefficients.

To encourage specialization and complementary decision-making, each agent $\mathcal{S}_k$ also receives an agent-specific reward R_k measuring its performance on the designated reasoning dimension:

$$R_k = \mu_1 \cdot \text{Orth}_k + \mu_2 \cdot \text{Phon}_k + \mu_3 \cdot \text{Ctx}_k, \tag{7}$$

where Orth_k, Phon_k, and Ctx_k quantify orthographic similarity, phonetic alignment, and contextual relevance, respectively, and μ_1, μ_2, μ_3 balance their contributions. Specifically, Orth_k is defined as the normalized character-level edit similarity:

$$\mathrm{Orth}_k\left(X,\hat{Y}\right)=1-\frac{\mathrm{EditDistance}\left(X,\hat{Y}\right)}{\max\left(|X|,\left|\hat{Y}\right|\right)}, \tag{8}$$

Phon_k is computed analogously on the corresponding pinyin sequences:

$$\mathrm{Phon}_k\left(X,\hat{Y}\right)=1-\frac{\mathrm{EditDistance}\left(\phi(X),\phi\left(\hat{Y}\right)\right)}{\max\left(|\phi(X)|,\left|\phi\left(\hat{Y}\right)\right|\right)}, \tag{9}$$

where $\phi(\,)$ maps a character sequence to its phonetic representation; and Ctx_k is defined as the sentence-level semantic similarity:

$$\mathrm{Ctx}_k\left(X,\hat{Y}\right)=\cos\left(f_{\mathrm{enc}}(X),f_{\mathrm{enc}}\left(\hat{Y}\right)\right). \tag{10}$$

The overall objective of SRC is to maximize the expected cumulative reward of the swarm, combining both global and agent-specific signals:

$$\mathcal{J}(\theta)=\mathbb{E}_{\pi_\theta}\left[R_g+\sum\nolimits_{k\in\{o,p,c\}}R_k\right], \tag{11}$$

where π_θ denotes the joint policy parameterized by θ. The optimization is performed via policy gradient methods, updating the parameters according to:

$$\nabla_\theta\mathcal{J}=\mathbb{E}_{\pi_\theta}\left[\nabla_\theta\log\pi_\theta(\,a\mid\mathbf{X}\,)\cdot\left(R_g+\sum\nolimits_k R_k\right)\right]. \tag{12}$$

By integrating global coordination and agent-level specialization, SRC enables HCR to achieve adaptive multi-agent collaboration, balancing collective objectives with individual expertise and improving correction performance under diverse and noisy scenarios.

3.4. Debate-Enhanced Arbitration

While multi-agent knowledge distillation enables the student agents to acquire complementary expertise and swarm reinforcement collaboration optimizes their cooperative strategies, conflicts may still arise when the agents produce divergent correction hypotheses due to their specialized reasoning perspectives. To resolve these conflicts and achieve reliable final predictions, we introduce a Debate-Enhanced Arbitration (DEA) mechanism, which refines agent outputs through iterative debates and confidence-adaptive fusion. The debate process is performed at inference time and does not update model parameters; instead, it operates on the prediction distributions to reach a consensus.

Given an input sequence $\mathbf{X}$, each agent $\mathcal{S}_k$ produces an initial prediction distribution $\boldsymbol{P}^k$ and a self-estimated confidence score C^k. During the debate phase, the agents iteratively exchange hypotheses and counterarguments, incorporating peer feedback to refine their predictions. Let $\mathbf{P}_t^k$ denote the prediction of agent $\mathcal{S}_k$ at debate round t, and Δ_t^k represent the adjustment derived from peer feedback. Specifically, Δ_t^k is defined as the confidence-weighted disagreement between agent $\mathcal{S}_k$ and the other agents:

$$\Delta_t^k = \sum_{j \neq k} \alpha_{kj} \cdot \left(\boldsymbol{P}_t^j - \boldsymbol{P}_t^k \right), \tag{13}$$

where the peer weight α_{kj} is computed from the confidence scores via a softmax normalization, $\alpha_{kj} = \frac{\exp(C^j)}{\sum_{l \neq k} \exp(C^l)}$. The refined prediction at round $t + 1$ is updated as:

$$\mathbf{P}_{t+1}^k = \mathbf{P}_t^k + \eta \cdot \Delta_t^k, \tag{13}$$

followed by a normalization step to ensure $\mathbf{P}_{t+1}^k$ remains a valid probability distribution. η is the debate learning rate controlling the extent to which external feedback influences the agent's predictions. This update is not gradient-based and does not involve backpropagation; it is a deterministic, confidence-weighted refinement at the distribution level.

After T rounds of debate, the arbitration module computes the final prediction probability $P(y_i \mid \mathbf{X})$P defined in Section 3.2 by aggregating the refined outputs $\{\mathrm{P}_T^k\}$ from all agents using a confidence-adaptive fusion strategy:

$$P(y_i \mid \mathbf{X}) = \frac{\sum_k w_k \cdot P_T^k(y_i \mid \mathbf{X})}{\sum_k w_k}, \tag{14}$$

where $P(y_i \mid \mathbf{X})$ is the final prediction from agent $\mathcal{S}_k$, and w_k denotes the arbitration weight assigned to agent $\mathcal{S}_k$.

To incorporate both the agent's confidence and its reinforcement performance introduced in Section 3.3, the weight is computed as:

$$w_k = \frac{\exp(\delta_1 C^k + \delta_2 R_k)}{\sum_j \exp\left(\delta_1 C^j + \delta_2 R_j\right)}, \tag{15}$$

where δ_1 and δ_2 are balancing coefficients for confidence and reinforcement contributions, respectively.

To encourage consensus among agents, we introduce a debate consistency loss that penalizes significant divergence between their final predictions:

$$\mathcal{L}_{\text{debate}} = \frac{1}{K^2} \sum_{k=1}^{K} \sum_{j=1}^{K} \| \mathbf{P}_T^k - \mathbf{P}_T^j \|_2^2, \tag{16}$$

where K is the number of student agents.

Finally, we integrate the objectives of multi-agent knowledge distillation, swarm reinforcement collaboration, and debate-enhanced arbitration into a unified loss function:

$$\mathcal{L}_{\text{HCR}} = \gamma_1 \cdot \mathcal{L}_{\text{MKD}} - \gamma_2 \cdot \mathcal{J}(\theta) + \gamma_3 \cdot \mathcal{L}_{\text{debate}}, \tag{17}$$

where γ_1, γ_2, γ_3 are non-negative hyperparameters that balance the relative contributions of knowledge distillation, reinforcement optimization, and debate consistency within the overall objective.

Training alternates between knowledge distillation, swarm reinforcement updates, and multi-round debates until convergence, while inference starts with agents generating initial predictions, refining them through debates, and reaching consensus via confidence-adaptive arbitration. For clarity, we summarize the overall training procedure and stage-wise parameter updates in the Algorithm.

Algorithm 1 End-to-End Training of HCR

```
Require: Training set $\mathcal{D}$; teacher $T$ (frozen); students $\{A_o, A_p, A_s\}$ with parameters $\theta$
Ensure: Trained parameters $\theta$
for epoch $= 1$ to $E$ do
    for minibatch $(x, y) \sim \mathcal{D}$ do
        Stage 1: Multi-Agent Knowledge Distillation
        Obtain teacher outputs $(P_T, \tau_T) \leftarrow T(x)$
        Each agent computes $P_i = \pi_i(\cdot|x;\theta)$
        Compute $\mathcal{L}_{MKD}$
        $\theta \leftarrow \theta - \eta \nabla_\theta \mathcal{L}_{MKD}$

        Stage 2: Swarm Reinforcement Collaboration
        Sample actions $a_i \sim \pi_i(\cdot|x;\theta)$
        Compute $R = R_g + \sum_i R_i$
        $\theta \leftarrow \theta + \alpha \nabla_\theta E_{\pi_\theta}[R]$

        Stage 3: Debate-Enhanced Arbitration
        Initialize $P_i^{(0)} = P_i$
        for $t = 1$ to $T_{deb}$ do
            Update $P_i^{(t)}$ via confidence-weighted disagreement
        end for
        Compute $\mathcal{L}_{deb}$
        $\theta \leftarrow \theta - \eta \nabla_\theta \mathcal{L}_{deb}$
    end for
end for
return $\theta$
```

4. Experimental Results

4.1. Experiment Setup

4.1.1. Datasets

The evaluation of the proposed HCR framework was conducted on four public benchmarks and one real-world administrative dataset to comprehensively assess its performance and generalization capability. The public datasets SIGHAN13 (Wu et al., 2013), SIGHAN14 (Yu et al., 2014), and SIGHAN15 (Tseng et al., 2015) are among the most widely used corpora for Chinese Spelling Correction and contain a diverse range of linguistic phenomena across formal and informal text. Each dataset provides sentence-level annotations with character-level error labels that allow precise evaluation of both detection and correction capabilities. They include phonetic confusions, visually similar character substitutions, and context-driven semantic inconsistencies, offering a balanced distribution of error types representative of real usage.

To further evaluate robustness under practical deployment, a medical administrative dataset containing 1000 manually verified sentences from de-identified hospital records was constructed. The sentences were sampled from routine clinical administrative texts, including discharge summaries and examination descriptions, after strict de-identification. All personal identifiable information, including patient names, IDs, contact information, and exact dates, was removed or anonymized before annotation.

Annotation was performed following a three-stage protocol. First, two annotators with medical background independently corrected each sentence and marked erroneous spans. Second, a senior annotator with clinical experience adjudicated all disagreements and produced the final gold-standard correction. Third, a random subset of the dataset was manually reviewed for quality control to ensure consistency with the annotation guidelines. The annotation guidelines require minimal corrections that preserve the original clinical meaning while fixing orthographic, phonetic, or contextual errors.

The resulting dataset covers a mixture of error types, including orthographic confusions, phonetic substitutions, context-driven semantic errors, and medical term normalization errors. The detailed distribution of these error types is reported in Table 1. This dataset introduces genuine challenges including specialized medical terminology, abbreviation ambiguity, and mixed-script expressions that frequently occur in clinical documents. To assess annotation reliability, we measured inter-annotator agreement before adjudication using span-level F1 for error detection and Cohen's κ for error-type classification, and the results indicate a high level of consistency between annotators, as shown in Table 2.

Table 1. Error-type distribution in the medical administrative dataset (N=1000).

Error Type	#Sentences	Percentage (%)	Description
Orthographic	210	21	Visually similar character substitutions
Phonetic	170	17	Phonetically similar character confusions
Semantic / Contextual	330	33	Context-driven inappropriate word usage
Medical Term Normalization	190	19	Non-standard or inconsistent medical term expressions
Numeric / Unit / Format	100	10	Errors in numbers, units, or formatting
Total	1000	100	—

Table 2. Inter-annotator agreement on the medical administrative dataset before adjudication.

Task	Metric	Score (%)
Error Detection (span-level)	Precision	91.4
Error Detection (span-level)	Recall	88.0
Error Detection (span-level)	F1-score	89.7
Error Type Classification	Cohen's κ	83.9
Sentence-level Correction	Exact Match Agreement	85.4

4.1.2. Implementation Details

HCR was implemented using PyTorch 2.2 and the Transformers 4.40 framework. The architecture followed a teacher–student paradigm in which a fine-tuned Qwen3-8B (Yang et al., 2025) served as the teacher model, while three student agents based on Qwen3-1.7B (Yang et al., 2025) were designed for orthographic, phonetic, and semantic reasoning. All experiments were conducted on NVIDIA A100 GPUs under mixed-precision settings to ensure computational efficiency and numerical stability.

Model optimization employed the AdamW algorithm with a learning rate of 2×10^{-5}, a batch size of 32 for each GPU, and gradient accumulation across four steps. The learning rate followed a linear warmup schedule with a warmup ratio of 0.1. The input sequence length was limited to 256 tokens, which covered nearly all samples across datasets. The teacher temperature was set to 2.0 to facilitate smooth knowledge transfer during interactive distillation.

Training involved three rounds of inter-agent debate where agents exchanged intermediate predictions and refined their reasoning through iterative interaction. The arbitration module combined confidence-based and reinforcement-based weighting with δ_1 set to 0.7 and δ_2 set to 0.3. The joint loss integrated supervised correction loss, distillation alignment loss, and reinforcement reward loss, weighted by $\gamma_1 = 1.0$, $\gamma_2 = 0.3$, and $\gamma_3 = 0.5$ respectively. For the global swarm reward, we set $(\lambda_1, \lambda_2, \lambda_3) = (0.50, 0.15, 0.35)$. For the agent-specific reward, we set $(\mu_1, \mu_2, \mu_3) = (0.40, 0.25, 0.35)$. Training proceeded for twenty epochs with early stopping after five stagnant validation steps. Random seeds 42 were fixed for all runs, and model checkpoints and configurations were preserved for replication.

4.1.3. Metrics

Performance evaluation was conducted at both the detection and correction levels. At the detection level, a prediction is regarded as correct when all erroneous tokens in a sentence are accurately identified. At the correction level, both identification and substitution must correspond exactly to the reference annotation. Precision, recall, and F1-score were used to measure performance at both levels, providing a consistent view of detection sensitivity and correction accuracy. All results were averaged over five independent runs.

To provide inferential statistical support for the main performance comparisons, we conduct paired bootstrap resampling on the fixed test sets. For each dataset, we compare HCR with the strongest baseline and estimate the significance of the F1 difference using 10,000 bootstrap samples. The strongest baseline on each dataset is defined as the non-HCR model achieving the highest correction-level F1 score in Table 3. Improvements are considered statistically significant when the two-sided p-value is below 0.05.

4.2. Experimental Results and Comparison

Table 3 presents a comprehensive comparison between HCR and a range of competitive baselines across four benchmark datasets. HCR consistently achieves superior performance in both detection and correction tasks. On SIGHAN13, HCR attains an F1-score of 89.1 percent at the detection level and 89.2 percent at the correction level, outperforming ProTEC (Li et al., 2025b) by 3.1 and 3.9 percentage points respectively. The gain demonstrates that multi-agent specialization effectively captures the interplay between phonetic, orthographic, and semantic reasoning, enabling more accurate error identification and correction. A similar trend is observed on SIGHAN15, where HCR improves the F1-score to 85.5 percent in detection and 84.3 percent in correction, surpassing the best baseline by over 1.5 percent. These results confirm that interactive distillation and debate-driven arbitration jointly enhance both stability and convergence across datasets with balanced linguistic complexity.

Performance on SIGHAN14, which is known for its higher lexical ambiguity and domain noise, further validates the robustness of HCR. The model obtains relative F1 improvements of 2.3 percent in detection and 1.9 percent in correction compared with ProTEC (Li et al., 2025b). The observed advantage can be attributed to the swarm reinforcement mechanism that dynamically calibrates cooperation strength among agents, ensuring a better equilibrium between specialization and coordination.

We note that we do not include large prompted LLM baselines on the SIGHAN benchmarks. This is because these datasets were constructed under specific annotation conventions and language usage styles that differ from the distribution of modern general-domain corpora used to pretrain current LLMs. Without task-specific fine-tuning, such models exhibit strong systematic bias and unstable behavior on character-level CSC benchmarks, making the comparison unfair and difficult to interpret; while fine-tuning them would turn them into task-specific systems and blur the distinction between general LLM baselines and trained CSC models. Therefore, we follow the standard and reproducible evaluation protocol on SIGHAN and compare HCR with strong, trainable, and fully reproducible baselines.

When evaluated on the Medical Records dataset, we additionally include strong open-source LLM baselines, including Qwen3-32B (Yang et al., 2025) and DeepSeek-Distill-Qwen3-32B (Guo et al., 2025), to reflect the current LLM-level performance under reproducible settings. HCR achieves the highest scores of 92.0 percent and 92.6 percent for detection and correction, respectively, exceeding the strongest baseline by about 3 percent. The consistent superiority across both standard and real-world data indicates that HCR generalizes effectively to complex, domain-specific corpora where conventional single-agent models and even large prompted LLMs exhibit performance degradation.

Table 3. Performance comparison of HCR with representative baseline models on SIGHAN13, SIGHAN14, SIGHAN15, and the Medical Records dataset. Precision (P), Recall (R), and F1-score (F1) are reported at both detection and correction levels. The best results are highlighted in bold. * denotes statistically significant improvement over the strongest baseline on the same dataset (paired bootstrap resampling on sentence-level test sets, two-sided $p < 0.05$).

Dataset	Models	Detection Level			Correction Level		
		P (%)	R (%)	F1 (%)	P (%)	R (%)	F1 (%)
SIGHAN13	SpellGCN (Cheng et al., 2020)	80.1	74.4	77.2	78.3	72.7	75.4
	UMRSpell (He et al., 2023)	83.0	73.6	78.0	80.0	71.0	75.2
	MSC (Wang et al., 2024c)	86.6	80.0	83.2	85.2	78.7	81.8
	SPMSpell (He et al., 2024)	87.7	83.7	85.6	86.9	82.8	84.6
	ProTEC (Li et al., 2025b)	88.5	83.7	86.0	87.7	83.0	85.3
	IPCK-IME (Zhao et al., 2025)	85.0	80.5	82.7	84.0	78.4	81.1
	HCR (Ours)	**89.4**	**88.7**	**89.1***	**90.2**	**88.3**	**89.2***
SIGHAN14	SpellGCN (Cheng et al., 2020)	65.1	69.5	67.2	63.1	67.2	65.3
	UMRSpell (He et al., 2023)	69.0	56.6	62.2	63.9	57.2	60.4
	MSC (Wang et al., 2024c)	65.7	68.3	67.0	65.8	67.3	67.1
	SPMSpell (He et al., 2024)	68.6	73.5	70.5	67.0	71.2	69.0
	ProTEC (Li et al., 2025b)	70.2	73.3	71.7	69.3	**72.3**	70.7
	IPCK-IME (Zhao et al., 2025)	67.0	70.1	68.5	66.0	68.9	67.4
	HCR (Ours)	**73.1**	**74.9**	**74.0***	**73.4**	71.8	**72.6***
SIGHAN15	SpellGCN (Cheng et al., 2020)	74.8	80.7	77.7	72.1	77.7	75.9
	UMRSpell (He et al., 2023)	77.2	72.2	75.0	69.3	64.8	67.0
	MSC (Wang et al., 2024c)	77.0	80.3	78.6	75.9	79.9	76.9

	SPMSpell (He et al., 2024)	81.7	**85.6**	83.6	79.4	83.4	81.3
	ProTEC (Li et al., 2025b)	82.9	84.8	83.8	80.3	82.3	81.3
	IPCK-IME (Zhao et al., 2025)	78.0	81.7	79.8	76.0	78.9	77.4
	HCR (Ours)	**85.6**	85.3	**85.5***	**83.8**	**84.7**	**84.3***
Medical Records	SpellGCN (Cheng et al., 2020)	82.3	81.1	81.7	82.1	80.5	81.3
	UMRSpell (He et al., 2023)	84.2	80.8	82.5	83.5	81.0	82.2
	MSC (Wang et al., 2024c)	86.7	84.9	85.8	87.0	84.5	85.7
	SPMSpell (He et al., 2024)	88.4	86.1	87.2	89.1	86.3	87.7
	ProTEC (Li et al., 2025b)	89.5	87.9	88.7	90.4	87.7	89.0
	IPCK-IME (Zhao et al., 2025)	88.5	85.4	86.9	87.0	83.3	85.1
	Qwen3-32B (Yang et al., 2025)	90.9	87.2	89.0	89.6	87.3	88.4
	DeepSeek-32B (Guo et al., 2025)	88.4	85.0	86.7	90.1	83.7	86.8
	HCR (Ours)	**92.7**	**91.3**	**92.0***	**93.5**	**91.7**	**92.6***

4.3. Ablation Study

The ablation analysis on the SIGHAN15 dataset reveals the individual contributions of each component within HCR, as shown in Table 4. When multi-agent knowledge distillation is removed, the F1-score drops from 85.5 to 81.7 in detection and from 84.3 to 79.7 in correction. This reduction indicates that teacher-guided specialization is essential for aligning orthographic, phonetic, and semantic reasoning, allowing the agents to maintain complementary expertise. Without this supervision, the agents tend to converge toward redundant representations, leading to limited generalization and less stable convergence during training.

Table 4. Ablation analysis of HCR on the SIGHAN15 dataset. Each variant removes one core component from the full model to evaluate its contribution. P, R, and F1 denote Precision, Recall, and F1-score at detection and correction levels.

Model	**Detection Level**			**Correction Level**		
	P (%)	**R (%)**	**F1 (%)**	**P (%)**	**R (%)**	**F1 (%)**
HCR (Full Model)	**85.6**	**85.3**	**85.5**	**83.8**	**84.7**	**84.3**
w/o Multi-Agent Knowledge Distillation	81.3	82.0	81.7	79.1	80.3	79.7
w/o Swarm Reinforcement Collaboration	82.8	81.4	82.1	80.5	79.8	79.6
w/o Debate-Enhanced Arbitration	83.5	80.9	82.2	81.0	78.6	78.8
w/o Distillation and Arbitration	79.7	80.4	80.1	78.2	78.0	78.1

Eliminating the swarm reinforcement collaboration results in a notable decrease in both detection and correction accuracy, highlighting its importance in balancing cooperation and individual specialization. The decline from 85.5 to 82.1 in detection F1 confirms that adaptive reward sharing enhances inter-agent synergy and prevents overfitting to specific error types. Similarly, removing the debate-enhanced arbitration reduces interpretability and final decision reliability. The absence of structured interaction during arbitration leads to inconsistent predictions among agents and a 5.5 percent F1 degradation in correction accuracy. When both distillation and arbitration are removed simultaneously, the model experiences the most severe performance collapse, with overall F1 dropping to 80.0 and 78.1 for detection and correction respectively. This observation underscores that hierarchical integration among the three modules is not only complementary but also indispensable for stable reasoning. The results collectively demonstrate that HCR's strength arises from the synergy of specialized learning, adaptive collaboration, and iterative debate, which together establish a robust and interpretable correction mechanism.

5. Discussion

5.1. Error-Type Analysis on the Medical Dataset

A fine-grained evaluation was conducted on the medical dataset to examine model performance across four error categories, including pronunciation errors (Pro.), graphemic errors (Gly.), combined pronunciation and graphemic errors (Gly.&Pro.), and other errors (Oth.). As reported in Table 5, HCR consistently achieves the highest F1-scores across all categories, confirming the effectiveness of its hierarchical collaborative design in addressing heterogeneous linguistic phenomena.

Table 5. F1-scores (%) of different models on the medical dataset across error types. Pron. for Pronunciation errors, Graph. for Graphemic errors, Graph. & Pron. for Combined pronunciation and graphemic errors, Oth. for Other errors.

Models	Pro.	Graph.	Graph. & Pro.	Oth.
SpellGCN (Cheng et al., 2020)	77.3	74.1	71.8	67.5
UMRSpell (He et al., 2023)	76.4	70.6	73.2	65.9
MSC (Wang et al., 2024c)	84.2	75.5	76.8	72.4
SPMSpell (He et al., 2024)	82.7	80.2	78.6	74.3
ProTEC (Li et al., 2025b)	86.5	82.4	80.9	77.8
IPCK-IME (Zhao et al., 2025)	85.1	74.8	79.5	70.8
HCR (Ours)	**91.3**	**88.2**	**89.1**	**90.4**

For pronunciation-related errors, which mainly involve homophonic substitutions within domain-specific terminology, HCR reaches an F1-score of 91.3 percent. This advantage stems from the specialized pronunciation agent guided by multi-agent knowledge distillation, which enables accurate modeling of phonological similarity and effectively mitigates confusions between homophones. In graphemic errors, where visually similar characters often arise in handwritten or OCR-based medical records, HCR achieves 88.2 percent F1, outperforming all baselines. The improvement demonstrates the contribution of the orthographic agent that explicitly encodes character-level structural patterns and shape resemblance.

When pronunciation and graphemic variations occur simultaneously, the swarm reinforcement collaboration facilitates interaction between the two specialized agents, allowing complementary reasoning that reduces ambiguity and strengthens error disambiguation. HCR obtains 89.1 percent F1 in this category, highlighting the importance of adaptive cooperation for compound error resolution. In the remaining category of other errors, which encompasses abbreviation inconsistencies, mixed-script expressions, and context-dependent lexical substitutions, HCR achieves a remarkable 90.4 percent F1, representing the largest relative improvement among all types. This result illustrates the capability of the debate-enhanced arbitration module to integrate contextual semantics and agent confidence dynamically, refining predictions through iterative consensus formation. Overall, the results verify that HCR's layered specialization and collaborative reasoning yield robust correction performance across both surface-level and context-driven errors.

5.2. Hyperparameter Sensitivity Analysis

We investigate the sensitivity of HCR to the key hyperparameters on the SIGHAN13 benchmark, and the results are summarized in Figure 2. Figure 2(a) shows the effect of the temperature τ used in knowledge distillation. The performance exhibits a clear unimodal trend and reaches the optimum around τ = 2.0, indicating that too small τ leads to over-confident and less informative soft targets, while too large τ over-smooths the distribution and weakens the distillation signal.

Figure 2(b) analyzes the influence of the balancing coefficients in the overall objective $L_{\mathrm{HCR}} = \gamma_1 L_{\mathrm{MKD}} - \gamma_2 J(\theta) + \gamma_3 L_{\mathrm{debate}}$. We vary one coefficient at a time while fixing the others. The results show that the performance is stable in a relatively wide neighborhood and peaks around $\gamma_1 = 1.0$, $\gamma_2 = 0.3$, and $\gamma_3 = 0.5$, confirming that a proper balance between knowledge distillation, reinforcement optimization, and debate consistency is crucial, and that the chosen defaults provide a robust trade-off.

Figure 2(c) reports the sensitivity to the global reward weights $(\lambda_1, \lambda_2, \lambda_3)$ in R_g under the constraint $\lambda_1 + \lambda_2 + \lambda_3 = 1$. The heatmap shows a smooth performance surface with a broad high-performance region, and the optimum lies near a balanced combination of accuracy-oriented and semantic-consistency-oriented rewards, while purely emphasizing fluency or accuracy leads to suboptimal results. This indicates that the model benefits from jointly considering task correctness and linguistic quality.

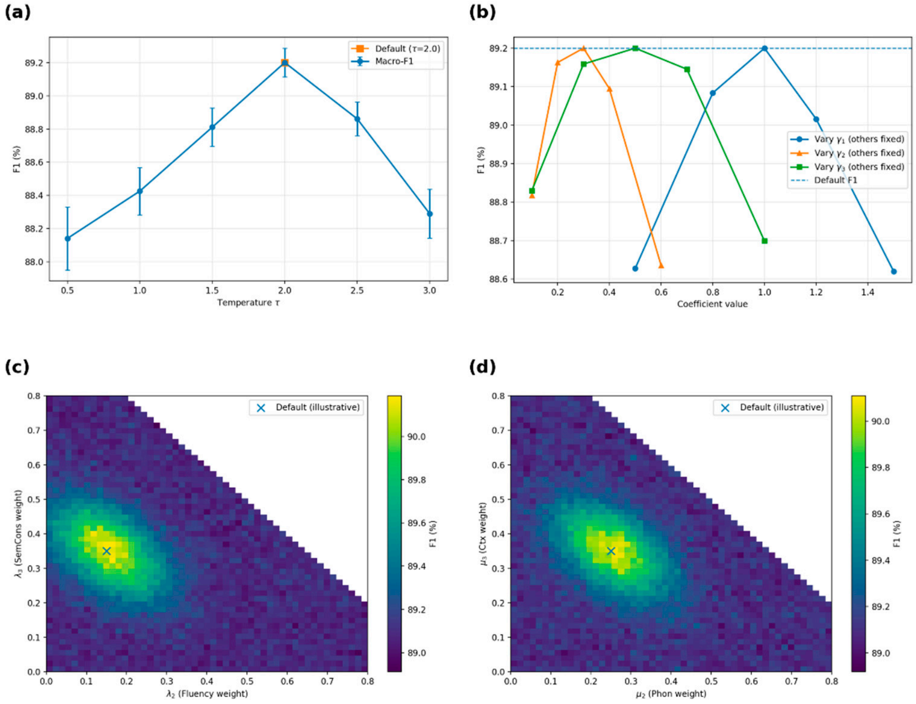

Figure 2. Sensitivity analysis of key hyperparameters in HCR on the SIGHAN13 benchmark. (a) Effect of temperature τ in knowledge distillation. (b) Sensitivity to the loss balancing coefficients γ in L_{HCR}. (c) Sensitivity to the global reward weights λ in R_g. (d) Sensitivity to the agent-specific reward weights μ in R_k.

Similarly, Figure 2(d) shows the sensitivity to the agent-specific reward weights (μ_1, μ_2, μ_3) in R_k. The performance again forms a smooth landscape with a clear high-performance region around the default setting, suggesting that moderately emphasizing contextual signals while keeping orthographic and phonetic cues well balanced leads to the most stable collaboration among agents.

5.3. Computational Cost and Efficiency Analysis

We analyze the computational cost of HCR in terms of training time, inference latency, and model size. The detailed comparison with a single-agent Qwen3-1.7B baseline is reported in Table 6.

Since HCR adopts a teacher–student framework, the teacher model is only used during training for knowledge distillation and is completely discarded at inference time. Therefore, the inference cost of HCR is solely determined by the three student agents and the lightweight debate–arbitration module, resulting in a total of approximately 5.1B parameters during inference.

Table 6. Computational cost comparison on medical test set.

Method	#Params (Inference)	Training Time / Epoch	Total Training Time	Inference Time / Sentence
Qwen3	1.7B	38 min	12.5 h	42 ms
HCR (Ours)	≈5.1B	96 min	32 h	93 ms

As shown in Table 6, compared with the single-agent baseline, the training time per epoch of HCR increases from 38 minutes to 96 minutes, which is about 2.5× slower due to the three parallel student agents, reinforcement optimization, and multi-round debate interactions. Nevertheless, the total training process remains practical, and the full training of 20 epochs finishes within about 32 hours.

During inference, HCR requires three parallel forward passes and three rounds of lightweight debate before arbitration. This increases the average inference latency from 42 ms to 93 ms per sentence, corresponding to approximately 2.2× overhead compared with the single-agent model. This overhead is moderate and acceptable for offline or batch-processing scenarios such as document-level text normalization and medical record cleaning.

5.4. Impact of Teacher and Student Model Capacity

To study the adaptability of HCR to different student model capacities and backbone architectures, we conduct an ablation study on SIGHAN13 by varying the student agents under the constraint that the student capacity is strictly smaller than the teacher model size. The results are reported in Table 7.

When using Qwen3-8B as the teacher model, replacing the default Qwen3-1.7B students with Llama-7B leads to a decrease in F1 from 89.2% to 88.5%, indicating that cross-architecture student models are less effective for Chinese character-level correction due to differences in tokenization and pretraining data distribution. This suggests that, although HCR is architecture-agnostic, backbone–data alignment still plays an important role in this task.

Table 7. Correction Level Effect of teacher model scale on SIGHAN13.

Teacher Model	Student Agents	P (%)	R (%)	F1 (%)
Qwen3-8B	Qwen3-1.7B	90.2	88.3	89.2
Qwen3-8B	Llama-7B	89.6	87.4	88.5
Qwen3-32B	Qwen3-1.7B	90.7	88.9	89.8
Qwen3-32B	Llama-7B	90.2	88.2	89.2
Qwen3-32B	Qwen3-8B	91.4	89.6	90.5
Qwen3-32B	Qwen3-14B	91.7	90	90.8

When a stronger Qwen3-32B teacher is used, the performance improves consistently across all student configurations. Specifically, the F1-score increases from 89.8% with Qwen3-1.7B students to 90.5% with Qwen3-8B students, and further to 90.8% with Qwen3-14B students. This shows that HCR can effectively exploit both stronger teachers and higher-capacity students, as better reasoning trajectories and softer targets provide higher-quality supervision during multi-agent distillation.

It is also observed that the performance gain exhibits a diminishing-return trend as the student capacity increases, suggesting that once the student models reach a certain scale, the collaboration mechanism and training strategy become the main performance bottleneck rather than raw model size.

5.5. Limitations and Future Work

While HCR demonstrates consistent improvements across datasets, several aspects still warrant further refinement. Although the proposed framework shows strong robustness when transferring from general-domain benchmarks to the medical administrative domain, this setting does not cover extreme domain shifts such as legal text, low-resource dialectal content, or highly informal user-generated text, where linguistic conventions, terminology distribution, and error patterns may differ substantially. Under such scenarios, the current collaboration and arbitration mechanisms may require domain-specific calibration or partial retuning to maintain stability and performance. The current framework may also be sensitive to the teacher model's quality and the balance among agents during reinforcement updates, which could influence stability under extreme domain shifts. In addition, the debate mechanism introduces modest computational overhead that may be optimized for large-scale deployment. Future work will systematically investigate cross-domain adaptation of HCR under more radical distribution shifts and explore whether lightweight domain-adaptive reweighting or partial reconfiguration of agent roles is sufficient to preserve its effectiveness, while future extensions toward lighter collaboration strategies and more adaptive knowledge transfer could further enhance HCR's efficiency and generalization without altering its core design.

6. Conclusions

In this work, we proposed HCR, a hierarchical collaborative reasoning framework that unifies interactive knowledge distillation, swarm reinforcement collaboration, and debate-enhanced arbitration to address heterogeneous errors in Chinese spelling correction. By leveraging multi-agent specialization and adaptive collaboration, HCR effectively disentangles orthographic, phonetic, and semantic dependencies while refining predictions through iterative consensus. Extensive experiments on three public benchmarks and a real-world medical dataset demonstrate that HCR achieves state-of-the-art performance in both detection and correction tasks and exhibits strong robustness under domain shifts, establishing a solid foundation for future research on interpretable, adaptive, and generalizable collaborative reasoning systems.

Ethics Statement: This study used retrospectively collected administrative medical text that was fully de-identified prior to analysis. All direct identifiers, including personal names, identification numbers, contact information, and exact dates, were removed or anonymized before annotation. According to institutional policy, the use of fully de-identified textual data does not constitute human subject research and therefore does not require formal Institutional Review Board (IRB) approval.

Data Governance Statement: Data access was restricted to authorized researchers within the hosting institution. All annotation and model development procedures were conducted under institutional data governance and privacy protection regulations.

Conflicts of Interest: The authors declare no competing interests.

Reference

Liu, L., Wu, H., & Zhao, H. (2025, January). Driving chinese spelling correction from a fine-grained perspective. In *Proceedings of the 31st International Conference on Computational Linguistics* (pp. 10727-10737).

Li, Y., Huang, H., Wang, B., & Gao, Y. (2025). DRMSpell: dynamically reweighting multimodality for Chinese spelling correction. *Frontiers of Information Technology & Electronic Engineering*, *26*(3), 354-366.

Yang, A. (2025). Generative AI chatbots: the future of grammar and spelling correction in English learning. *International Journal of Information and Communication Technology*, *26*(22), 72-87.

Wang, Y., Zheng, Z., Tang, Z., Li, J., Liu, Z., Chen, K., ... & Zhang, M. (2024, March). Towards better chinese spelling check for search engines: A new dataset and strong baseline. In *Proceedings of the 17th ACM International Conference on Web Search and Data Mining* (pp. 769-778).

Liu, C., Zhang, K., Jiang, J., Kong, Z., Liu, Q., & Chen, E. (2025). Chinese spelling correction: a comprehensive survey of progress, challenges, and opportunities. *arXiv preprint arXiv:2502.11508*.

Liu, S., Yang, T., Yue, T., Zhang, F., & Wang, D. (2021, August). PLOME: Pre-training with misspelled knowledge for Chinese spelling correction. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)* (pp. 2991-3000).

Li, C., Zhang, M., Zhang, X., & Yan, Y. (2024). MCRSpell: A metric learning of correct representation for Chinese spelling correction. *Expert Systems with Applications*, *237*, 121513.

Zhou, H., Li, Z., Zhang, B., Li, C., Lai, S., Zhang, J., ... & Zhang, M. (2024, November). A simple yet effective training-free prompt-free approach to Chinese spelling correction based on large language models. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing* (pp. 17446-17467).

Xu, M., Liu, J., Peng, K., & Li, Z. (2025). Chinese spelling correction based on Long Short-Term Memory Network-enhanced Transformer and dynamic adaptive weighted multi-task learning. *Natural Language Processing*, *31*(5), 1265-1284.

Zhu, C., Ying, Z., Zhang, B., & Mao, F. (2022, May). MDCSpell: A multi-task detector-corrector framework for Chinese spelling correction. In *Findings of the association for computational linguistics: ACL 2022* (pp. 1244-1253).

Liu, C., Zhang, K., Jiang, J., Liu, Z., Tao, H., Gao, M., & Chen, E. (2024, November). ARM: An alignment-and-replacement module for Chinese spelling check based on LLMs. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing* (pp. 10156-10168).

Li, J., Wu, G., Yin, D., Wang, H., & Wang, Y. (2021, July). Dcspell: A detector-corrector framework for chinese spelling error correction. In *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval* (pp. 1870-1874).

Achiam, J., Adler, S., Agarwal, S., Ahmad, L., Akkaya, I., Aleman, F. L., ... & McGrew, B. (2023). Gpt-4 technical report. *arXiv preprint arXiv:2303.08774*.

Yang, A., Li, A., Yang, B., Zhang, B., Hui, B., Zheng, B., ... & Qiu, Z. (2025). Qwen3 technical report. *arXiv preprint arXiv:2505.09388*.

Guo, D., Yang, D., Zhang, H., Song, J., Wang, P., Zhu, Q., ... & He, Y. (2025). Deepseek-r1: Incentivizing reasoning capability in llms via reinforcement learning. *arXiv preprint arXiv:2501.12948*.

Huang, H., Ye, J., Zhou, Q., Li, Y., Li, Y., Zhou, F., & Zheng, H. T. (2023, December). A frustratingly easy plug-and-play detection-and-reasoning module for Chinese spelling check. In *Findings of the Association for Computational Linguistics: EMNLP 2023* (pp. 11514-11525).

Wu, S. H., Liu, C. L., & Lee, L. H. (2013, October). Chinese spelling check evaluation at SIGHAN bake-off 2013. In *Proceedings of the seventh SIGHAN workshop on Chinese language processing* (pp. 35-42).

Tseng, Y. H., Lee, L. H., Chang, L. P., & Chen, H. H. (2015, July). Introduction to SIGHAN 2015 bake-off for Chinese spelling check. In *Proceedings of the Eighth SIGHAN Workshop on Chinese Language Processing* (pp. 32-37).

Li, Y., Zhou, Q., Li, Y., Li, Z., Liu, R., Sun, R., ... & Zheng, H. T. (2022, May). The past mistake is the future wisdom: Error-driven contrastive probability optimization for chinese spell checking. In *Findings of the Association for Computational Linguistics: ACL 2022* (pp. 3202-3213).

Zhang, S., Huang, H., Liu, J., & Li, H. (2020, July). Spelling error correction with soft-masked BERT. In *Proceedings of the 58th annual meeting of the association for computational linguistics* (pp. 882-890).

Ji, T., Yan, H., & Qiu, X. (2021, November). SpellBERT: A lightweight pretrained model for Chinese spelling check. In *Proceedings of the 2021 conference on empirical methods in natural language processing* (pp. 3544-3551).

Zhang, R., Pang, C., Zhang, C., Wang, S., He, Z., Sun, Y., ... & Wang, H. (2021, August). Correcting Chinese spelling errors with phonetic pre-training. In *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021* (pp. 2250-2261).

Xu, H. D., Li, Z., Zhou, Q., Li, C., Wang, Z., Cao, Y., ... & Mao, X. L. (2021, August). Read, listen, and see: Leveraging multimodal information helps Chinese spell checking. In *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021* (pp. 716-728).

Guo, Z., Ni, Y., Wang, K., Zhu, W., & Xie, G. (2021, August). Global attention decoder for Chinese spelling error correction. In *Findings of the association for computational linguistics: ACL-IJCNLP 2021* (pp. 1419-1428).

Yang, S., & Yu, L. (2022, August). CoSPA: an improved masked language model with copy mechanism for Chinese spelling correction. In *Uncertainty in Artificial Intelligence* (pp. 2225-2234). PMLR.

Liu, L., Wu, H., & Zhao, H. (2024, March). Chinese spelling correction as rephrasing language model. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 38, No. 17, pp. 18662-18670).

Wang, Y., Zheng, Z., Li, J., Liu, Z., Chang, J., Zhang, Q., ... & Zhang, M. (2024, May). Towards more realistic Chinese spell checking with new benchmark and specialized expert model. In Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024) (pp. 16570-16580).

Foerster, J., Assael, I. A., De Freitas, N., & Whiteson, S. (2016). Learning to communicate with deep multi-agent reinforcement learning. *Advances in neural information processing systems*, *29*.).

Lowe, R., Wu, Y. I., Tamar, A., Harb, J., Pieter Abbeel, O., & Mordatch, I. (2017). Multi-agent actor-critic for mixed cooperative-competitive environments. *Advances in neural information processing systems*, *30*.

Sunehag, P., Lever, G., Gruslys, A., Czarnecki, W. M., Zambaldi, V., Jaderberg, M., ... & Graepel, T. (2017). Value-decomposition networks for cooperative multi-agent learning. arXiv preprint arXiv:1706.05296.

Rashid, T., Samvelyan, M., De Witt, C. S., Farquhar, G., Foerster, J., & Whiteson, S. (2020). Monotonic value function factorisation for deep multi-agent reinforcement learning. *Journal of Machine Learning Research*, *21*(178), 1-51.

Irving, G., Christiano, P., & Amodei, D. (2018). AI safety via debate. *arXiv preprint arXiv:1805.00899*.

Wei, J., Wang, X., Schuurmans, D., Bosma, M., Xia, F., Chi, E., ... & Zhou, D. (2022). Chain-of-thought prompting elicits reasoning in large language models. *Advances in neural information processing systems*, *35*, 24824-24837.

Li, Y., Huang, H., Ma, S., Jiang, Y., Li, Y., Zhou, F., ... & Zhou, Q. (2023). On the (in) effectiveness of large language models for chinese text correction. *arXiv preprint arXiv:2307.09007*.

Shinn, N., Cassano, F., Gopinath, A., Narasimhan, K., & Yao, S. (2023). Reflexion: Language agents with verbal reinforcement learning. *Advances in neural information processing systems*, *36*, 8634-8652.

Li, Y., Ma, S., Zhou, Q., Li, Z., Yangning, L., Huang, S., ... & Zheng, H. (2022, December). Learning from the dictionary: Heterogeneous knowledge guided fine-tuning for chinese spell checking. In *Findings of the Association for Computational Linguistics: EMNLP 2022* (pp. 238-249).

Yu, L. C., Lee, L. H., Tseng, Y. H., & Chen, H. H. (2014, October). Overview of SIGHAN 2014 bake-off for Chinese spelling check. In *Proceedings of The Third CIPS-SIGHAN Joint Conference on Chinese Language Processing* (pp. 126-132).

Li, Y., Ma, S., Chen, S., Huang, H., Huang, S., Li, Y., ... & Shen, Y. (2025). Correct like humans: Progressive learning framework for chinese text error correction. *Expert Systems with Applications*, *265*, 126039.

Cheng, X., Xu, W., Chen, K., Jiang, S., Wang, F., Wang, T., ... & Qi, Y. (2020, July). Spellgcn: Incorporating phonological and visual similarities into language models for chinese spelling check. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics* (pp. 871-881).

He, Z., Zhu, Y., Wang, L., & Xu, L. (2023, July). UMRSpell: Unifying the detection and correction parts of pre-trained models towards Chinese missing, redundant, and spelling correction. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* (pp. 10238-10250).

Wang, Y., Wang, Y., & Liu, Y. (2024, January). Chinese Spelling Correction Method Based on Multi-feature Fusion and Attention Mechanism. In *Proceedings of the 3rd International Conference on Computer, Artificial Intelligence and Control Engineering* (pp. 481-487).

He, L., Liu, F., Liu, J., Duan, J., & Wang, H. (2024). Self-distillation and Pinyin character prediction for Chinese spelling correction based on multimodality. *Applied Sciences*, *14*(4), 1375.

Zhao, W., Wang, X., & An, X. (2025). Incorporating Confused Phraseological Knowledge Based on Pinyin Input Method for Chinese Spelling Correction. *IEEE Transactions on Big Data*.

Zhang, D., Li, Y., Zhou, Q., Ma, S., Li, Y., Cao, Y., & Zheng, H. T. (2023, June). Contextual similarity is more valuable than character similarity: An empirical study for chinese spell checking. In *ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 1-5). IEEE.