Novelty Assessment Report

Paper: Language Confusion Gate: Language-Aware Decoding Through Model Self-Distillation

PDF URL: https://openreview.net/pdf?id=JjJzbMDGsx

Venue: ICLR 2026 Conference Submission

Year: 2026

Report Generated: 2025-12-27

Abstract

Large language models (LLMs) often experience language confusion, which is the unintended mixing of languages during text generation. Current solutions to this problem either necessitate model retraining or cannot differentiate between harmful confusion and acceptable code-switching. This paper introduces the \textbf{Language Confusion Gate} (LCG), a lightweight, plug-in solution that filters tokens during decoding without altering the base LLM. The LCG is trained using norm-adjusted self-distillation to predict appropriate language families and apply masking only when needed. Our method is based on the findings that language confusion is infrequent, correct-language tokens are usually among the top predictions, and output token embedding norms are larger for high-resource languages, which biases sampling. When evaluated across various models, including Qwen3, GPT-OSS, Gemma3, Llama3.1, LCG decreases language confusion significantly—often by an order of magnitude—without negatively impacting task performance.

Core Task Landscape

This paper addresses: **Mitigating Unintended Language Mixing in Multilingual Language Models**

A total of **34 papers** were analyzed and organized into a taxonomy with **16 categories**.

Taxonomy Overview

The research landscape has been organized into the following main categories:

- **Decoding-Time Intervention Methods**
- **Training-Time and Architectural Approaches**
- **Cross-Lingual Knowledge Transfer and Interference**
- **Evaluation and Benchmarking**
- **Task-Specific Multilingual Applications**
- **Multilingual Model Development and Pretraining**

Complete Taxonomy Tree

- Mitigating Unintended Language Mixing in Multilingual Language Models Survey Taxonomy
- Decoding-Time Intervention Methods
  - Token-Level Filtering and Steering ★ (2 papers)
  - [0] Language Confusion Gate: Language-Aware Decoding Through Model Self-Distillation (Anon et al., 2026) View paper
  - [25] Language steering in latent space to mitigate unintended code-switching (Goncharov Andrey, 2025) View paper
  - Retrieval-Augmented Generation Adaptations (1 papers)
  - [24] Language Drift in Multilingual Retrieval-Augmented Generation: Characterization and Decoding-Time Mitigation (Bo Li, 2025) View paper
- Training-Time and Architectural Approaches
  - Preference Optimization and Alignment (3 papers)
  - [2] Controlling Language Confusion in Multilingual LLMs (Ko Hyunwoo, 2025) View paper
  - [21] Mitigating Language Confusion for Multimodal Foundation Models via Confusion-Aware Preference Optimization Pipeline (S Hwang, 2025) View paper
  - [23] CONGRAD: Conflicting Gradient Filtering for Multilingual Preference Alignment (Li Jiang-nan, 2025) View paper
  - Language-Specific Parameter Modulation (3 papers)
  - [14] LLaVA-NeuMT: Selective Layer-Neuron Modulation for Efficient Multilingual Multimodal Translation (Wei Jingxuan, 2025) View paper
  - [17] MPN: Leveraging Multilingual Patch Neuron for Cross-lingual Model Editing (Si, 2024) View paper
  - [28] MLM: Multi-linguistic LoRA Merging (J Lee, 2025) View paper
  - Continual Pretraining and Data Mixing Strategies (2 papers)
  - [7] Rethinking Multilingual Continual Pretraining: Data Mixing for Adapting LLMs Across Languages and Resources (Li Zihao, 2025) View paper
  - [29] Multilingual Adaptation in Large Language Models: Strategies and Insights from Continual Pretraining (Z Li, 2025) View paper
  - Tokenization and Vocabulary Design (2 papers)
  - [13] A Systematic Analysis of Subwords and Cross-Lingual Transfer in Multilingual Translation (Buys, 2024) View paper
  - [30] Parallel Tokenizers: Rethinking Vocabulary Design for Cross-Lingual Transfer (Koto, 2025) View paper
- Cross-Lingual Knowledge Transfer and Interference
  - Interference Analysis in Translation and Generation (4 papers)
  - [1] Causes and cures for interference in multilingual translation (Bhosale, 2023) View paper
  - [8] On negative interference in multilingual models: Findings and a meta-learning treatment (Zirui Wang, 2020) View paper
  - [9] Minimizing Language Interference for Multilingual Models (Xu, 2024) View paper

- [12] Multilingual Neural Machine Translation with Integrated Language Adapters (Jin, 2024) View paper
- Cross-Lingual Representation Alignment (3 papers)
- [10] A WordNet view on crosslingual transformers (Wondimagegnhue Tufa, 2023) View paper
- [16] mOthello: When Do Cross-Lingual Representation Alignment and Cross-Lingual Transfer Emerge in Multilingual Models? (Yun Tian, 2024) View paper
- [32] Improving In-context Learning of Multilingual Generative Language Models with Cross-lingual Alignment (Li Chong, 2023) View paper
- Cross-Lingual Knowledge Consistency and Factuality (2 papers)
- [3] Crosslingual capabilities and knowledge barriers in multilingual large language models (Chua, 2024) View paper
- [18] Are Knowledge and Reference in Multilingual Language Models Cross-Lingually Consistent? (Xi Ai, 2025) View paper

- Evaluation and Benchmarking
  - Language Confusion Benchmarks (1 papers)
  - [6] Understanding and mitigating language confusion in llms (BÃ©rard, 2024) View paper
  - Multilingual Capability Assessment (3 papers)
  - [5] How do large language models handle multilingualism? (Lidong Bing, 2024) View paper
  - [19] Evaluating the Cross-Lingual Syntactic Capabilities of Language Models (S Salhan, 2024) View paper
  - [34] All Languages Matter: On the Multilingual Safety of Large Language Models (Wang, 2023) View paper
- Task-Specific Multilingual Applications
  - Multilingual Information Retrieval and RAG (2 papers)
  - [4] Retrieval-augmented generation in multilingual settings (Nadezhda Chirkova, 2024) View paper
  - [22] Multilingual Generative Retrieval via Cross-lingual Semantic Compression (Huang, 2025) View paper
  - Zero-Shot Cross-Lingual Generation (1 papers)
  - [26] Empirical study of pretrained multilingual language models for zero-shot cross-lingual knowledge transfer in generation (Chirkova, 2023) View paper
  - Cross-Lingual Reasoning and In-Context Learning (1 papers)
  - [27] Pushing on Multilingual Reasoning Models with Language-Mixed Chain-of-Thought (Son, 2025) View paper
  - Specialized Multilingual Tasks (3 papers)
  - [11] Debiasing Multilingual LLMs in Cross-lingual Latent Space (Peng Qiwei, 2025) View paper
  - [20] Leveraging Multilingual Training for Authorship Representation: Enhancing Generalization across Languages and Domains (Jung-Hwan Kim, 2025) View paper
  - [33] UZH_CLyp at SemEval-2023 Task 9: Head-First Fine-Tuning and ChatGPT Data Generation for Cross-Lingual Learning in Tweet Intimacy Prediction (Clematide, 2023) View paper
- Multilingual Model Development and Pretraining (2 papers)
  - [15] Multilinguality and Multiculturalism: Towards more Effective and Inclusive Neural Language Models (Choenni, 2025) View paper
  - [31] Responsibly Building Multilingual Language Models for Hundreds of Languages (Eghlidi, 2025) View paper

Narrative

Core task: Mitigating unintended language mixing during text generation in multilingual language models. The field addresses a fundamental challenge in multilingual NLP: ensuring that models generate text in the intended language without inadvertently switching or blending languages. The taxonomy reveals six major branches that capture complementary perspectives on this problem. Decoding-Time Intervention Methods focus on runtime strategies such as token-level filtering and steering to guide generation toward the target language, while Training-Time and Architectural Approaches modify model design or learning objectives to reduce confusion at its source. Cross-Lingual Knowledge Transfer and Interference examines how shared representations can both enable transfer and introduce unwanted mixing, a tension explored in works like Crosslingual Knowledge Barriers[3] and Interference Multilingual Translation[1]. Evaluation and Benchmarking provides diagnostic tools to measure language confusion, Task-Specific Multilingual Applications adapt these insights to domains like retrieval or translation, and Multilingual Model Development and Pretraining investigates foundational choices in tokenization and corpus balancing that shape a model's propensity for mixing.

Recent work highlights contrasting strategies for controlling language confusion. Some approaches intervene during decoding by steering latent representations or filtering undesired tokens, as seen in Language Steering Latent[25] and Controlling Language Confusion[2], while others address the issue earlier through training modifications or architectural constraints like language adapters. The original paper, Language Confusion Gate[0], sits squarely within the Decoding-Time Intervention branch, specifically under Token-Level Filtering and Steering. It shares this focus with Language Steering Latent[25], yet differs in mechanism: where Language Steering Latent[25] manipulates hidden states to enforce language consistency, Language Confusion Gate[0] introduces a gating mechanism to selectively suppress cross-lingual tokens at generation time. This positions it as a lightweight, inference-time solution that complements training-based methods like Mitigating Language Confusion[6] and offers an alternative to heavier architectural interventions, addressing the practical need for post-hoc control in deployed multilingual systems.

Related Works in Same Category

The following **1 sibling papers** share the same taxonomy leaf node with the original paper:

1. Language steering in latent space to mitigate unintended code-switching

Authors: Goncharov Andrey, Zaytsev, Alexey | **Year/Venue**: 2025 | **URL**: View paper

Abstract

Multilingual Large Language Models (LLMs) often exhibit unintended code-switching, reducing reliability in downstream tasks. We propose latent-space language steering, a lightweight inference-time method that identifies language directions via PCA on parallel translations and steers token embeddings along these axes to control language identity. Our approach mitigates code-switching while preserving semantics with negligible computational overhead and requires only minimal parallel data for cali...

Relationship Analysis

Both papers belong to the Token-Level Filtering and Steering category, employing decoding-time interventions to control language identity during generation without modifying base model weights. They overlap in addressing unintended language mixing through token-level manipulation: the original paper (LCG) uses a trained MLP gate to predict permissible language families and mask inappropriate tokens, while the candidate paper applies latent-space steering via PCA-derived language directions to shift token embeddings. The key difference is that LCG operates on logits with learned language family predictions and dynamic masking, whereas the candidate performs geometric steering in hidden representation space using linear projections along principal components.

Contributions Analysis

Overall novelty summary. The paper proposes a Language Confusion Gate (LCG), a plug-in decoding-time filter that masks inappropriate language tokens during generation without retraining the base model. It resides in the Token-Level Filtering and Steering leaf, which contains only two papers including this one. This leaf sits within the broader Decoding-Time Intervention Methods branch, indicating a relatively sparse research direction compared to training-based approaches. The taxonomy shows 34 papers across 16 leaf nodes, suggesting the field is moderately populated but this specific decoding-time filtering niche remains underexplored.

The taxonomy reveals that neighboring work clusters around training-time solutions (Preference Optimization, Language-Specific Parameter Modulation) and cross-lingual interference analysis. The sibling paper in the same leaf, Language Steering Latent, manipulates hidden states rather than filtering tokens, highlighting a methodological divergence within the same problem space. The exclude_note clarifies that methods requiring model retraining belong elsewhere, positioning LCG as a lightweight alternative to heavier architectural interventions like language adapters or continual pretraining strategies found in adjacent branches.

Among 20 candidates examined, the LCG mechanism itself shows no clear refutation across 10 candidates reviewed. The norm-adjusted self-distillation training method was not examined against prior work. The specialized training and evaluation datasets contribution examined 10 candidates and found 1 refutable match, suggesting some overlap in dataset construction approaches. The limited search scope means these findings reflect top-K semantic matches rather than exhaustive coverage, with the core gating mechanism appearing more distinctive than the dataset contribution within the examined literature.

Based on the limited search of 20 candidates, the work appears to occupy a relatively novel position within decoding-time token filtering, though the dataset contribution shows some prior overlap. The sparse population of its taxonomy leaf and the methodological contrast with its sole sibling paper suggest a distinct approach, but the analysis does not cover the full landscape of multilingual generation control methods beyond top semantic matches.

This paper presents **3 main contributions**, each analyzed against relevant prior work:

Contribution 1: Language Confusion Gate (LCG)

Description: The authors propose a lightweight two-layer MLP intervention mechanism that dynamically filters inappropriate tokens at decoding time by predicting permissible language families and applying masking only when necessary, without modifying the base LLM weights.

This contribution was assessed against **10 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

---

1. MrT5: Dynamic Token Merging for Efficient Byte-level Language Models

URL: View paper

Brief Assessment

Dynamic Token Merging[45] focuses on token deletion mechanisms for byte-level models to improve efficiency, not on filtering inappropriate language tokens during multilingual decoding.

---

2. DeFTX: Denoised Sparse Fine-Tuning for Zero-Shot Cross-Lingual Transfer

URL: View paper

Brief Assessment

Denoised Sparse Finetuning[43] focuses on cross-lingual transfer using sparse fine-tuning with SVD-based denoising for parameter selection, not on token filtering mechanisms during decoding for language confusion mitigation.

---

3. Embedding Structure Matters: Comparing Methods to Adapt Multilingual Vocabularies to New Languages

URL: View paper

Brief Assessment

Embedding Structure Adaptation[48] focuses on vocabulary adaptation and embedding re-initialization for multilingual models during language-adaptive pre-training, not on token filtering mechanisms during decoding to prevent language confusion.

---

4. CLOVER: Cross-Layer Orthogonal Vectors Pruning and Fine-Tuning

URL: View paper

Brief Assessment

Cross Layer Orthogonal[47] focuses on model compression through SVD-based pruning and fine-tuning of attention layers, not on token filtering mechanisms for multilingual language confusion during decoding.

---

5. Adaptive Originality Filtering: Rejection Based Prompting and RiddleScore for Culturally Grounded Multilingual Riddle Generation

URL: View paper

Brief Assessment

Adaptive Originality Filtering[49] addresses semantic filtering for creative text generation (riddles) to enforce novelty and cultural fidelity, not token-level language confusion mitigation during multilingual decoding.

---

6. Reducing Computation Costs in Transformers with Token Pruning

URL: View paper

Brief Assessment

Token Pruning Transformers[52] focuses on computational efficiency through token pruning mechanisms, not on language-aware decoding or multilingual token filtering during generation.

---

7. Efficient continual pre-training of llms for low-resource languages

URL: View paper

Brief Assessment

Continual Pretraining Low Resource[44] focuses on vocabulary augmentation and corpus selection for continual pre-training of LLMs in low-resource languages, not on token filtering mechanisms during decoding for language confusion mitigation.

---

8. Dynamic token pruning for LLMs: leveraging task-specific attention and adaptive thresholds

URL: View paper

Brief Assessment

Dynamic Token Pruning[46] focuses on input token pruning for efficient decoding in pre-trained language models, not on filtering inappropriate language tokens during multilingual generation.

9. Retraining-Free Pruning Text-to-Speech Synthesis Model for Speaker Cloning

URL: View paper

Brief Assessment

Retraining Free Pruning[50] focuses on pruning text-to-speech synthesis models for speaker cloning, not on token filtering mechanisms for multilingual language models during decoding.

10. Control Extreme Multi-label Generation via Level-Guided Token Filtering

URL: View paper

Brief Assessment

Level Guided Filtering[51] focuses on extreme multi-label generation tasks with token filtering for label control, not multilingual language confusion mitigation during decoding.

Contribution 2: Norm-adjusted self-distillation training method

Description: The authors introduce a training approach that leverages the model's own debiased top-k/p predictions by adjusting logits with token embedding norms to remove systemic bias toward high-resource languages, enabling the gate to learn from the model's corrected language predictions.

This contribution was assessed against **0 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

Contribution 3: Specialized training and evaluation datasets

Description: The authors collect and release datasets specifically designed for training the language confusion gate and evaluating language confusion across diverse multilingual contexts, covering over 200 languages and approximately 78,000 samples.

This contribution was assessed against **10 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

1. ILT-Iterative LoRA Training through Focus-Feedback-Fix for Multilingual Speech Recognition

URL: View paper

Brief Assessment

Iterative LoRA Training[40] focuses on multilingual speech recognition datasets for ASR tasks, not language confusion evaluation. The datasets serve different purposes: the original paper addresses language confusion in text generation across 200+ languages, while the candidate addresses speech-to-text recognition challenges.

2. Reducing language confusion for code-switching speech recognition with token-level language diarization

URL: View paper

Brief Assessment

Token Level Diarization[38] focuses on code-switching speech recognition datasets (SEAME corpus) for training language diarization modules, not on multilingual datasets for evaluating language confusion in text generation across 200+ languages as in the original paper.

3. Language Surgery in Multilingual Large Language Models

URL: View paper

Brief Assessment

Language Surgery[42] focuses on multilingual datasets for training language control mechanisms and evaluating cross-lingual capabilities, not specifically on language confusion detection and evaluation as described in the original paper's contribution.

4. Optimism, Expectation, or Sarcasm? Multi-Class Hope Speech Detection in Spanish and English

URL: View paper

Brief Assessment

Hope Speech Detection[39] focuses on emotion classification datasets for hope speech detection in English and Spanish, not on language confusion or multilingual language model evaluation across 200+ languages.

5. Factual Consistency of Multilingual Pretrained Language Models

URL: View paper

Brief Assessment

Factual Consistency Multilingual[36] focuses on multilingual paraphrase datasets for measuring consistency in factual knowledge retrieval, not on language confusion detection or training confusion gates across 200+ languages.

6. Against all odds: Overcoming typology, script, and language confusion in multilingual embedding inversion attacks

URL: View paper

Brief Assessment

Embedding Inversion Attacks[35] focuses on multilingual embedding inversion attacks across 20 languages with datasets from CulturaX for security vulnerability assessment, not on language confusion detection in LLMs. The datasets serve fundamentally different purposes: security testing versus language confusion evaluation.

7. SentiXRL: An advanced large language Model Framework for Multilingual Fine-Grained Emotion Classification in Complex Text Environment

URL: View paper

Brief Assessment

SentiXRL Emotion Classification[37] focuses on sentiment classification datasets with unified labels across fine-grained emotion annotation datasets, not on multilingual language confusion datasets covering 200+ languages for training language confusion gates.

8. Understanding and mitigating language confusion in llms

URL: View paper

Prior Art Analysis

Mitigating Language Confusion[6] demonstrates that specialized datasets for training and evaluating language confusion in multilingual contexts were created and released prior to the original paper. The candidate paper explicitly describes creating the Language Confusion Benchmark (LCB) covering 15 typologically diverse languages with approximately 7,600 prompts (2,600 monolingual + 4,500 cross-lingual), sourced from multiple datasets including human-generated and machine-translated content. The candidate also details specific data collection efforts including 'native prompts' and 'complex prompts' created by human annotators. This directly refutes the novelty claim of being the first to collect and release datasets specifically designed for training language confusion gates and evaluating language confusion across diverse multilingual contexts.

Evidence

Evidence 1 - **Rationale**: Both papers describe creating specialized datasets for evaluating language confusion. The candidate paper's LCB was created to evaluate language confusion across 15 languages, demonstrating prior work in this area. - **Original**: we collect and open-source specialized training and evaluation datasets, and evaluate lcg on open-source models covering diverse architectures and both thinking and no-think modes. - **Candidate**: we create the language confusion benchmark (lcb) to evaluate such failures, covering 15 typologically diverse languages with existing and newly-created english and multilingual prompts.

Evidence 2 - **Rationale**: The candidate paper explicitly states they created and released a benchmark for language confusion evaluation, covering diverse languages with both existing and newly-created data. - **Original**: third, we collect and open-source specialized training and evaluation datasets, and evaluate lcg on open-source models covering diverse architectures and both thinking and no-think modes. - **Candidate**: we create and release a language confusion benchmark covering 15 typologically diverse languages, sourcing prompts from public english and multilingual instruction datasets, and additionally creating new data with more complex prompts.

Evidence 3 - **Rationale**: Both papers describe collecting specialized data with human annotators. The candidate paper explicitly commissioned native annotators to create prompts for underrepresented languages and collected complex prompts, demonstrating prior work in creating specialized training/evaluation data for language confusion. - **Original**: training data for the gate.we trained the lcg on a composite dataset of approximately78,000 samplescovering over200 languagesto ensure it learns to handle a wide variety of linguistic contexts. - **Candidate**: native prompts (ours) for japanese and korean, under-represented in the above datasets, as well as spanish and french, we commission native annotators to collect our own prompts (see §a.2). complex prompts (ours) as prompts from the above sources are relatively simple, we collect complex english pro...

---

9. Cosda-ml: Multi-lingual code-switching data augmentation for zero-shot cross-lingual nlp

URL: View paper

Brief Assessment

Code Switching Augmentation[41] focuses on generating code-switched data for cross-lingual transfer using bilingual dictionaries, not on creating datasets for training and evaluating language confusion detection mechanisms across 200+ languages as in the original paper.

---

10. Minimizing Language Interference for Multilingual Models

URL: View paper

Brief Assessment

Minimizing Language Interference[9] focuses on multilingual NMT and cross-lingual transfer tasks, not on language confusion detection in LLM text generation. The datasets described cover translation benchmarks (FLORES+, OPUS-100, IWSLT) and NLU tasks, which serve different purposes than the language confusion gate training/evaluation datasets.

Appendix: Text Similarity Detection

Textual similarity detection checked 21 papers and found 2 similarity segment(s) across 2 paper(s).

The following **2 paper(s)** were detected to have high textual similarity with the original paper. These may represent different versions of the same work, duplicate submissions, or papers with substantial textual overlap. Readers are advised to verify these relationships independently.

1. Against all odds: Overcoming typology, script, and language confusion in multilingual embedding inversion attacks

Detected in: Contribution: contribution_3

⚠ **Note**: This paper shows substantial textual similarity with the original paper. It may be a different version, a duplicate submission, or contain significant overlapping content. Please review carefully to determine the nature of the relationship.

---

2. Language Surgery in Multilingual Large Language Models

Detected in: Contribution: contribution_3

⚠ **Note**: This paper shows substantial textual similarity with the original paper. It may be a different version, a duplicate submission, or contain significant overlapping content. Please review carefully to determine the nature of the relationship.

References

- [0] Language Confusion Gate: Language-Aware Decoding Through Model Self-Distillation View paper
- [1] Causes and cures for interference in multilingual translation View paper
- [2] Controlling Language Confusion in Multilingual LLMs View paper
- [3] Crosslingual capabilities and knowledge barriers in multilingual large language models View paper
- [4] Retrieval-augmented generation in multilingual settings View paper
- [5] How do large language models handle multilingualism? View paper
- [6] Understanding and mitigating language confusion in llms View paper
- [7] Rethinking Multilingual Continual Pretraining: Data Mixing for Adapting LLMs Across Languages and Resources View paper
- [8] On negative interference in multilingual models: Findings and a meta-learning treatment View paper
- [9] Minimizing Language Interference for Multilingual Models View paper
- [10] A WordNet view on crosslingual transformers View paper
- [11] Debiasing Multilingual LLMs in Cross-lingual Latent Space View paper
- [12] Multilingual Neural Machine Translation with Integrated Language Adapters View paper
- [13] A Systematic Analysis of Subwords and Cross-Lingual Transfer in Multilingual Translation View paper
- [14] LLaVA-NeuMT: Selective Layer-Neuron Modulation for Efficient Multilingual Multimodal Translation View paper
- [15] Multilinguality and Multiculturalism: Towards more Effective and Inclusive Neural Language Models View paper
- [16] mOthello: When Do Cross-Lingual Representation Alignment and Cross-Lingual Transfer Emerge in Multilingual Models? View paper

- [17] MPN: Leveraging Multilingual Patch Neuron for Cross-lingual Model Editing View paper
- [18] Are Knowledge and Reference in Multilingual Language Models Cross-Lingually Consistent? View paper
- [19] Evaluating the Cross-Lingual Syntactic Capabilities of Language Models View paper
- [20] Leveraging Multilingual Training for Authorship Representation: Enhancing Generalization across Languages and Domains View paper
- [21] Mitigating Language Confusion for Multimodal Foundation Models via Confusion-Aware Preference Optimization Pipeline View paper
- [22] Multilingual Generative Retrieval via Cross-lingual Semantic Compression View paper
- [23] CONGRAD: Conflicting Gradient Filtering for Multilingual Preference Alignment View paper
- [24] Language Drift in Multilingual Retrieval-Augmented Generation: Characterization and Decoding-Time Mitigation View paper
- [25] Language steering in latent space to mitigate unintended code-switching View paper
- [26] Empirical study of pretrained multilingual language models for zero-shot cross-lingual knowledge transfer in generation View paper
- [27] Pushing on Multilingual Reasoning Models with Language-Mixed Chain-of-Thought View paper
- [28] MLM: Multi-linguistic LoRA Merging View paper
- [29] Multilingual Adaptation in Large Language Models: Strategies and Insights from Continual Pretraining View paper
- [30] Parallel Tokenizers: Rethinking Vocabulary Design for Cross-Lingual Transfer View paper
- [31] Responsibly Building Multilingual Language Models for Hundreds of Languages View paper
- [32] Improving In-context Learning of Multilingual Generative Language Models with Cross-lingual Alignment View paper
- [33] UZH_CLyp at SemEval-2023 Task 9: Head-First Fine-Tuning and ChatGPT Data Generation for Cross-Lingual Learning in Tweet Intimacy Prediction View paper
- [34] All Languages Matter: On the Multilingual Safety of Large Language Models View paper
- [35] Against all odds: Overcoming typology, script, and language confusion in multilingual embedding inversion attacks View paper
- [36] Factual Consistency of Multilingual Pretrained Language Models View paper
- [37] SentiXRL: An advanced large language Model Framework for Multilingual Fine-Grained Emotion Classification in Complex Text Environment View paper
- [38] Reducing language confusion for code-switching speech recognition with token-level language diarization View paper
- [39] Optimism, Expectation, or Sarcasm? Multi-Class Hope Speech Detection in Spanish and English View paper
- [40] ILT-Iterative LoRA Training through Focus-Feedback-Fix for Multilingual Speech Recognition View paper
- [41] Cosda-ml: Multi-lingual code-switching data augmentation for zero-shot cross-lingual nlp View paper
- [42] Language Surgery in Multilingual Large Language Models View paper
- [43] DeFTX: Denoised Sparse Fine-Tuning for Zero-Shot Cross-Lingual Transfer View paper
- [44] Efficient continual pre-training of llms for low-resource languages View paper
- [45] MrT5: Dynamic Token Merging for Efficient Byte-level Language Models View paper
- [46] Dynamic token pruning for LLMs: leveraging task-specific attention and adaptive thresholds View paper
- [47] CLOVER: Cross-Layer Orthogonal Vectors Pruning and Fine-Tuning View paper
- [48] Embedding Structure Matters: Comparing Methods to Adapt Multilingual Vocabularies to New Languages View paper
- [49] Adaptive Originality Filtering: Rejection Based Prompting and RiddleScore for Culturally Grounded Multilingual Riddle Generation View paper
- [50] Retraining-Free Pruning Text-to-Speech Synthesis Model for Speaker Cloning View paper
- [51] Control Extreme Multi-label Generation via Level-Guided Token Filtering View paper
- [52] Reducing Computation Costs in Transformers with Token Pruning View paper