

Optimizing LLM Reasoning for Intra-English Dialects via Dynamic Persona Injection

Onyedikachukwu Tochukwu Mba¹, Jae Min Lee¹, Dong-Seong Kim^{2*}

¹ IT-Convergence Engineering, *Kumoh National Institute of Technology*, Gumi, South Korea

² ICT Convergence Research Center, *Kumoh National Institute of Technology*, Gumi, South Korea

* NSLab Co. Ltd., Gumi, South Korea, *Kumoh National Institute of Technology*, Gumi, South Korea
(toecm, raneemkhafagy, ljmpaul, dskim)@kumoh.ac.kr

Abstract—Existing language translation models, trained predominantly on Western, Educated, Industrialized, Rich, and Democratic (WEIRD) datasets, frequently fail to decode the high-context nuances of global English varieties. These models normalize distinct cultural identities into "standard" forms, often erasing speaker intent. This paper introduces the *Context-Aware Intra-English Dialect Interpretation (CA-IEDI)* framework, specifically focusing on its Dynamic Model Manager (DMM) and Persona Injection mechanisms. By transitioning from static syntactic lookups to a cognitive, persona-aware system, we enable the mediation of pragmatic intent (tone and context). We demonstrate how "Persona Injection" handles polysemy—such as distinguishing the Nigerian "*I beg*" as a discourse marker rather than a plea—optimizing reasoning depth through hierarchical Large Language Model (LLM) switching.

Index Terms—Large Language Models (LLMs), Persona Injection, Natural Language Processing (NLP), Intra-English Dialects, Generative AI.

I. INTRODUCTION

English serves as the primary lingua franca for global communication [1], yet it manifests in distinct dialects rich in historical and cultural divergence. Misinterpretations in these environments rarely stem from grammatical errors but from pragmatic mismatches [2]. For instance, the American English phrase "*It's gonna take a minute*" implies a long duration, whereas the Nigerian English sentence or phrase, "*I am coming*", in addition to its actual literal meaning, could imply a present continuous action despite physical absence.

Standard Automatic Speech Recognition (ASR) and Natural Language Understanding (NLU) models tend to "correct" these dialects toward a standardized mean, stripping the communication of its sociolinguistic context [3]. They are also trained on WEIRD bias data [4]. Previous variations of the *I Understand Understand You (IUUY)* system were able to translate dialects but struggled to differentiate between raw syntax and cultural pragmatics [5]–[7]. To bridge this gap, we present an architectural advancement: **Contextual Persona Injection**. This approach moves beyond simple translation by embedding specific sociolinguistic "codebooks" into the inference pipeline, allowing the AI to reason about the "why" behind a phrase rather than just the "what".

II. REASONING ARCHITECTURE

The core contribution of this work is the shift from a static dictionary-based architecture to a dynamic, reasoning-based

agent. This is achieved through two primary components: the Dynamic Model Manager and the Persona Injection mechanism.

A. Dynamic Model Manager (DMM)

We implemented a DMM that optimizes computational cost versus reasoning depth. This acts as a cost-optimized "Brain Agent" utilizing hierarchical model switching.

- **Deep Reasoning (Pro):** For high-ambiguity inputs or initial setup, the system utilizes *Gemini 1.5 Pro*. This model provides deep reasoning capabilities necessary for mapping complex pragmatic intent.
- **High Throughput (Flash):** For routine processing, the system falls back to *Gemini 1.5 Flash*, ensuring high throughput without prohibitive computational costs.

B. Contextual Persona Injection

The most significant limitation of generic translators is their lack of specific cultural framing. We address this via "Persona Injection":

- 1) **Persona Loading:** Before uploading audio and analyzing the transcript, a "Persona" (e.g., "*Nigerian Cultural Context*" or "*Lab Persona*") is loaded onto the system.
- 2) **Codebook Prioritization:** This tells the AI which definition to prioritize. Unlike generic translators, our system utilizes these injectable "*Persona Codebooks*" (JSON-based cultural definitions) to interpret jargon within specific sociolinguistic frames.
- 3) **Contextual Metadata:** The dataset couples audio inputs with rich metadata columns—specifically *Tone_Category* and *Linguistic_Context*. This enables the model to identify specific cultural contexts, such as attitudes toward hierarchy or directness in communication.

By injecting these personas, the system creates a "rich context" environment where the LLM can perform disambiguation based on the loaded cultural parameters rather than default training data

III. EVALUATION OF PRAGMATIC COMPETENCE

We evaluated the efficacy of Persona Injection by comparing the system's ability to handle polysemy—where a single phrase has multiple meanings based on tone—against a baseline static lookup model.

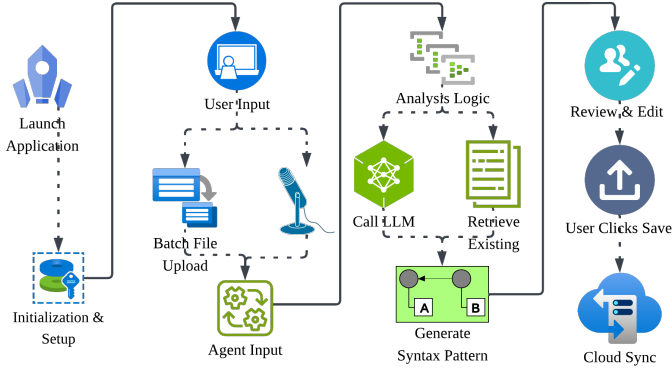


Fig. 1: Data Analysis Hybrid Automation

A. The "I Beg" Test Case

A critical failure case in standard models is the Nigerian English phrase "I beg."

- **Baseline (Static Dictionary):** The initial version of the system (Phase 1) utilized a syntax-based dictionary lookup. It interpreted "I beg" literally as the verb "to plead". It failed to identify it as a discourse marker, resulting in a null or confusing clarification.
- **CA-IEDI (Persona Injected):** In Phase 3, utilizing the "Rich Context" architecture, the system successfully identified "I beg" as a polysemous marker.

B. Qualitative Results

The application of Persona Injection allowed the system to offer three distinct tonal Interpretations for the same phrase, moving from translation to mediation as shown in table I.

TABLE I: Impact of Persona Injection on Polysemy Resolution

Feature	Version 1 (Baseline)	Version 3 (Final Rich Context)	Sociolinguistic Impact
Input	"I beg..."	"I beg..."	
Interpretation	Literal ("To plead")	Option A: "Please" (Casual) Option B: "Seriously?" (Frustrated)	Disambiguation: The final system recognizes that dialect meaning is tone-dependent.
Context	None	"Discourse marker used to soften commands."	Educational Value: Moves beyond translation to cultural education.
Data Schema	Single Row (Overwritten)	Multi-Row (Appended)	Data Integrity: Preserves the history of language evolution rather than erasing it.

C. Reasoning Depth Analysis

The evaluation confirms that "Contextual Persona Injection" effectively addresses the limitations of WEIRD [4] datasets. By telling the AI which definition to prioritize via the injected codebook, the system resolves ambiguity in high-context cultures where non-verbal communication is dominant. The DMM ensures this occurs efficiently; the "Thinking" capabilities of the Pro model are engaged only when necessary

to establish these pragmatic maps, while the Flash model handles retrieval.

IV. CONCLUSION

This paper demonstrates that the "Context Blindness" inherent in traditional translation tools can be solved by evolving the architecture from static lookups to a Sociolinguistic Agent capable of reasoning. The implementation of Persona Injection allows the CA-IEDI framework to treat "Tone" and "Cultural Context" as first-class data citizens. By successfully distinguishing between identical phrases carrying different intents (e.g., the tonal variations of "I beg"), we validate that LLMs, when constrained by specific cultural codebooks, can effectively mediate the "hidden misunderstandings" in World Englishes.

ACKNOWLEDGMENT

This work was partly supported by Innovative Human Resource Development for Local Intellectualization program through the IITP grant funded by the Korea government (MSIT) (IITP-2025-RS-2020-II201612, 33%) and by Priority Research Centers Program through the NRF funded by the MEST (2018R1A6A1A03024003, 33%) and by the MSIT, Korea, under the ITRC support program (IITP-2025-RS-2024-00438430, 34%).

REFERENCES

- [1] Statista, "The most spoken languages worldwide," Online Chart, 2024, data sourced from various sources including the *Ethnologue* and the United Nations. Chart indicates the total number of speakers (native and secondary) for select languages. [Online]. Available: <https://www.statista.com/statistics/266808/the-most-spoken-languages-worldwide/>
- [2] J. Fiset, T. Nguyen, and M. Roberts, "The effects of language-related misunderstanding at work," *Journal of Management*, vol. 49, no. 2, pp. 345–372, 2023.
- [3] K. Prinos, N. Patwari, and C. A. Power, "Speaking of accent: A content analysis of accent misconceptions in asr research," in *Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency (FAccT '24)*. Rio de Janeiro, Brazil: ACM, 2024, pp. 1–10.
- [4] A. A. Septiandri, M. Constantinides, and D. Quercia, "Weird icwsm: How western, educated, industrialized, rich, and democratic is social computing research?" *arXiv preprint*, 2024.
- [5] O. T. Mba, L. A. C. Ahakonye, J. M. Lee, and D.-S. Kim, "Iedi: A hybrid nlp-llm framework for intra-english dialect interpretation," in *Proceedings of the KICS International Conference on Communications*, 2025.
- [6] —, "I understand understand you: A reliable multi-agent facilitator for reducing communication breakdowns," in *Proceedings of the International Conference on ICT Convergence (ICTC)*, 2025.
- [7] —, "Choose your character ai experience: A mixture-of-experts framework for personalized multi-dialect speech interpretation," in *Proceedings of the ICC International Conference on Communication*, 2026.