



LoCar: Localization-Aware Evaluation of In-Vehicle Assistants through Fine-Grained Sociolinguistic Control

Kiwoong Park, Seogyong Jeong, Ken E. Friedl

| Beyond Advancement

BMW GROUP

Premium Identity & Brand Voice

- Brand Heritage: Requires sophisticated interaction consistent with "The Ultimate Driving Machine."
- Strategic Localization: Establishing global standards by addressing unique sociolinguistic nuances of the Korean market.

KAIST

Academic Excellence & AI Expertise

- Scientific Research: Pioneering research in the field of Cultural Alignment for LLMs.
- Linguistic Depth: Advanced computational linguistics expertise in analyzing the complex Korean honorific system.

“Beyond simple technical advancement, LoCar builds a user-centric, localized AI ecosystem that bridges KAIST’s technical excellence with BMW’s brand premium.”

| Motivation: Beyond Translation

AI Assistant as the In-vehicle Primary Interface

- The shift toward Software-Defined Vehicles positions AI assistants as the definitive interface between drivers and mobility environments.

Limitations of Existing Benchmarks

- Current benchmarks overlook domain-specific constraints such as low cognitive load, safety, and clarity.
- Failure to master nuances like Korean honorifics can critically undermine a premium brand identity.

South Korea, A Strategic Testbed

- Serves as a leading environment for deploying cutting-edge mobility technologies.
- Demands sophisticated AI that integrates cultural and situational contexts beyond simple translation.

No domain-specific evaluation standard exists for Korean in-vehicle AI

LoCar FRAMEWORK: Use case and KPI

Use case	KPI (Key Performance Indicator)		
Car expert 	Linguistic Style Layer	Conciseness	
		Honorifics	Hapsyo (Formal Polite)
			Haeyo (Informal Polite)
			Hae (Casual Intimate)
Navigation 	Dialogue Competence Layer	Single-Turn	Implicit Understanding
			Contextual Comprehension
			Harmful Question Response
		Multi-Turn	Clarification
			Retention
			Refinement
			Reflection
			Proactivity
			Troubleshooting

LoCar FRAMEWORK: Evaluation Pipeline

Step 1.

Honorific Form Check

(Rule-based)

Segment response into sentences



Verify each sentence against the target honorific ending

[Honorific Sentence Endings]

Hapsyo: 오 o, 다 da, 까 kka

Haeyo: 요 yo

Hae: 어 eo, 해 hae, 아 ya, 다 da, 까 kka



All sentences
PASS ✓

Forward Honorifics
metric to Step 2

Any sentence
FAILS ✗
Honorifics = False

Proceed to Step 2
for other metrics

Step 2.

Contextual KPI Evaluation

Inputs

- Dialogue History
- Model Response
- KPI Rubrics
(w/ honorific & meaningful information inclusion)

LLM Judges
(Majority Voting)



GPT-5-mini



Gemini-2.5-flash



DeepSeek-v3.1
-terminus

Step 3.

Meaningful Information Gating

Meaningful
information?

- Yes → ✓ All KPIs included
- No → ✗ Excluded

[Exception]

Harmful Question Response is
Always evaluated due to
safety-critical nature

- "Thank you" → No meaningful ✗
- Refusal to harmful request → Safety-critical ✓

| Key Findings

Instability in Fine-Grained Honorific Control

- **Register Confusion:** While models maintain general politeness, they frequently conflate distinct speech levels (e.g., Haeyo vs. Hapsyo), failing to meet premium brand standards.

Performance Gap in Strategic Dialogue

- **Saturation in Comprehension:** Performance in single-turn comprehension is near-saturated across all models
- **Strategic Variance:** Significant performance gaps emerge in multi-turn strategic metrics, such as Clarification and Proactivity, where complex interaction management is required.

Conservative Evaluation Bias

- **Stricter Standards:** The LoCar framework applies more rigorous criteria than human annotators in ambiguous scenarios to ensure operational reliability
- **Safety-First Approach:** This conservative bias prioritizes in-vehicle safety and trust, penalizing uncertain proactive interventions that could distract the driver.

| Deployment Implications

Brand Identity

Inappropriate linguistic styles directly compromise premium brand identity; sociolinguistic precision is non-negotiable for luxury localization.

Sociolinguistic Norms

Mastering local social norms and linguistic conventions is the key to securing user trust and achieving successful global scaling.

Safety-Oriented Strategy

AI proactivity must be strictly governed by safety guidelines to minimize cognitive load and ensure reliable in-vehicle decision-making.

Moving beyond general-purpose capabilities, Mobility AI must prioritize deep cultural alignment and safety-critical dialogue strategies to ensure trust in localized deployment.



Thank You

“Sheer Driving Pleasure”

Contact: marspak@kaist.ac.kr