Do the Models *Hear* Us?

Artificial Intelligence and the (In)Visibility of Migrant Voices and Languages

Jasper Kyle Catapang

Doctoral Researcher, *Tokyo University of Foreign Studies*
AI Decision Quality Lead, *Money Forward Japan*

December 5, 2025

The Central Problem

AI systems increasingly mediate migrant mobility, communication, and access—yet we lack systematic understanding of how AI models interpret migrant voices.

- Language is "the indispensable agent of migration" (Borlongan, 2023).
- AI infrastructures mediate inclusion and exclusion (Benjamin, 2019; Eubanks, 2018).

Why This Is Linguistic, Not Technical

Migration Linguistics emphasizes:

- Migrants must be understood in languages they *use*, not those institutions prefer (Borlongan, 2023).
- AI systems assume standardized, idealized speakers.
- This mismatch produces structural vulnerability.
- It becomes a question of linguistic rights, not model performance.

Migration Linguistics as Framework

- Interdisciplinary and multidimensional (Borlongan, 2023).
- Human-centered study of language in mobility.
- Describes contact, shift, and hybridization.
- Insists on practical solutions for migrants.
- AI is “a cultural and political actor” in linguistic life (Borlongan & Catapang, forthcoming).

AI as a sociopolitical actor

"AI is a cultural and political actor shaping whose voices are heard."

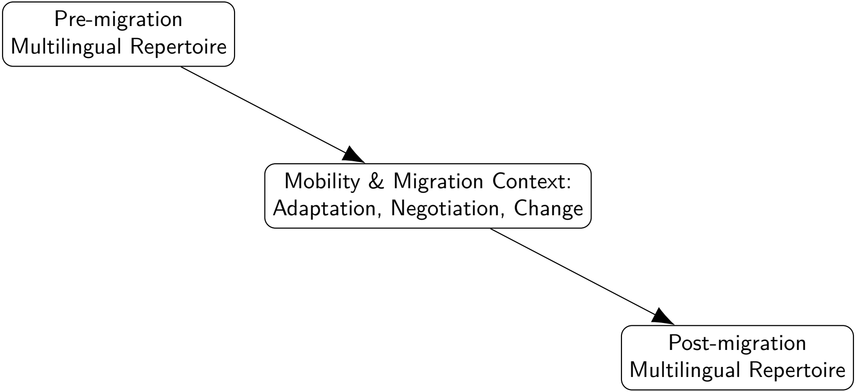
—Borlongan and Catapang (forthcoming)

Migrant Tongues and Hybrid Repertoires

Migration produces:

- Translingual practices
- Hybrid vocabularies
- Contact-induced forms
- Code-switching repertoires
- Accented L2 English varieties
- Dynamic repertoires changing across time (Borlongan, 2023)

Language in Motion (after Borlongan 2023)



The Five Aims of Migration Linguistics

1. Theorize migration-driven language change
2. Describe migrant linguistic practices
3. Use interdisciplinary empirical tools
4. Provide practical solutions for migrants
5. Engage migrants as participants (Borlongan, 2023)

AI Bias: What We Already Know

- ASR racial and accent disparities (Koenecke et al., 2020; Tatman, 2017)
- Embedding and representational bias (Bolukbasi et al., 2016; Caliskan et al., 2017)
- Toxicity misclassification of dialects (Sap et al., 2019; Xia et al., 2020)
- MT inequality for low-resource languages (Arivazhagan et al., 2019)
- Dataset gaps reflecting linguistic hierarchies (Blodgett et al., 2020)

Migration Linguistics vs. AI Fairness

Domain	Migration Linguistics	AI Research
Repertoires	Hybrid, dynamic	Penalized, misread
Rights	Legibility, access	Misclassification harms
Context	Mobility-grounded	Benchmark-grounded
Power	Visible gatekeeping	Hidden gatekeeping

Linguistic Rights and Digital Citizenship

- Migrants' rights include the right to be understood.
- AI mediates access to institutions (Borlongan & Catapang, forthcoming).
- Algorithmic systems reproduce linguistic hierarchies (Crawford, 2021; Pasquale, 2015).

Linguistic Invisibility

- Misrecognition becomes procedural exclusion.
- Toxicity classifiers penalize non-standard English (Sap et al., 2019).
- ASR fails in high-stakes contexts (Koenecke et al., 2020).
- MT errors reshape bureaucratic interpretation.

What AI Misses Entirely

- Code-switching dynamics
- Migrant hybrid repertoires
- Shifting proficiency over time
- Culturally marked expressions
- Interaction under power asymmetry

How do AI systems respond to migrant voices?

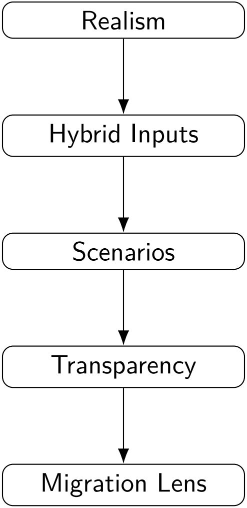
Why Existing Methods Are Insufficient

- Benchmarks are static and unrealistic.
- No scenario-specific evaluation.
- No hybrid or multilingual repertoires.
- No pragmatic or cultural context.
- No attention to institutional power.
- No migration lens whatsoever.

Conceptual Requirements

1. Sociolinguistic realism
2. Hybrid and multilingual inputs
3. Scenario-based evaluation
4. Judgment transparency
5. Migration-centered interpretation

Conceptual Stack



Reframing AI Bias

- Variation is not a model error.
- A migration-linguistic violation.
- A linguistic rights failure.

Toward an Evaluation Environment

Grounded in:

- Migration Linguistics (theory + rights)
- AI Fairness Research (methods)

Designed to examine how AI systems interpret migrant voices.

From Theory to Practice

Migration Linguistics tells us:

- migrant repertoires are hybrid, dynamic, and context-shaped;
- AI systems misrecognize these repertoires in systematic ways;
- evaluation must be grounded in real scenarios of migrant life.

But how do we test this in practice?

Demonstration

Evaluating AI systems through Migration Linguistics

For this segment, I will be showing the **Bridgr.AI** web app demo.

Academic Impact

- Establishes an empirical program for Migration Linguistics grounded in real-world migrant discourse.
- Reveals systematic patterns in how AI systems misrecognize multilingual and hybrid repertoires.
- Positions migrant language practices as central to evaluating fairness, interpretability, and model behavior.
- Opens a conceptual and methodological bridge between sociolinguistics, computational linguistics, and AI fairness.

Societal Impact

- Enhances visibility and legitimacy for migrant linguistic repertoires in automated systems.
- Provides actionable evidence for rights-based advocacy and policy intervention.
- Helps institutions understand how AI shapes access, recognition, and everyday bureaucratic navigation for migrants.

References I

Arivazhagan, N., et al. (2019).Massively multilingual neural machine translation. *NAACL.*

Benjamin, R. (2019). *Race after technology*. Polity.

Blodgett, S. L., Barocas, S., Daumé III, H., & Wallach, H. (2020).Language (technology) is power. *ACL.*

Bolukbasi, T., et al. (2016).Man is to computer programmer as woman is to homemaker? *NeurIPS.*

Borlongan, A. M. (2023).Migration linguistics: A synopsis. *AILA Review, 36*(1), 38–63. https://doi.org/10.1075/aila.22014.bor

Borlongan, A. M., & Catapang, J. K. (forthcoming). Migration, language, and artificial intelligence [Forthcoming book, Bloomsbury Academic].

Caliskan, A., Bryson, J., & Narayanan, A. (2017).Semantics derived automatically from language corpora contain human-like biases. *Science, 356*(6334), 183–186.

Crawford, K. (2021). *Atlas of ai*. Yale University Press.

Eubanks, V. (2018). *Automating inequality*. St. Martin's Press.

References II

Koenecke, A., Nam, M., Lake, E., et al. (2020).Racial disparities in automated speech recognition. *PNAS, 117*(14), 7684–7689.

Pasquale, F. (2015). *The black box society*. Harvard University Press.

Sap, M., et al. (2019).The risk of racial bias in toxicity detection. *ACL.*

Tatman, R. (2017).Gender and dialect bias in youtube's automatic captions. *ACL Workshop on Ethics in NLP.*

Xia, M., et al. (2020).Understanding dialect bias in toxicity classifiers. *EMNLP.*