

Digital Policy Hub – Working Paper

# Multilingualism as Design: Building Digital Architectures

**Anna Legault**

Winter term – 2025–2026 cohort

## About the Hub

The Digital Policy Hub at CIGI is a collaborative space for emerging scholars and innovative thinkers from the social, natural and applied sciences. It provides opportunities for undergraduate and graduate students and post-doctoral and visiting fellows to share and develop research on the rapid evolution and governance of transformative technologies. The Hub is founded on transdisciplinary approaches that seek to increase understanding of the socio-economic and technological impacts of digitalization and improve the quality and relevance of related research. Core research areas include data, economy and society; artificial intelligence; outer space; digitalization, security and democracy; and the environment and natural resources.

The Digital Policy Hub working papers are the product of research related to the Hub's identified themes prepared by participants during their fellowship.

## Partners

Thank you to Mitacs for its partnership and support of Digital Policy Hub fellows through the Accelerate program. We would also like to acknowledge the many universities, governments and private sector partners for their involvement allowing CIGI to offer this holistic research environment.



## About CIGI

The Centre for International Governance Innovation (CIGI) is an independent, non-partisan think tank whose peer-reviewed research and trusted analysis influence policy makers to innovate. Our global network of multidisciplinary researchers and strategic partnerships provide policy solutions for the digital era with one goal: to improve people's lives everywhere. Headquartered in Waterloo, Canada, CIGI has received support from the Government of Canada, the Government of Ontario and founder Jim Balsillie.

## About the Author

Anna Legault is a Digital Policy Hub master's fellow pursuing an M.A. in digital humanities at McGill University. With a background in linguistics and translation in United Nations and Canadian governmental contexts, she is a practitioner-academic of interdisciplinary language work. As part of the Digital Policy Hub fellowship, she is conducting research on emerging translation technologies and localized language models.

Copyright © 2026 by Anna Legault

The opinions expressed in this publication are those of the author and do not necessarily reflect the views of the Centre for International Governance Innovation or its Board of Directors.

Centre for International Governance Innovation and CIGI are registered trademarks.

67 Erb Street West  
Waterloo, ON, Canada N2L 6C2  
cigionline.org

# Multilingualism as Design: Building Digital Architectures

Anna Legault

## Bottom Line Up Front

Multilingualism is a design approach to artificial intelligence (AI) that safeguards the pluralism of knowledge in the digital sphere. Emerging language technologies, in particular large language models (LLMs), construct a spatial map of meaning in their neural networks using word-embedding techniques. These encodings represent the relationship between words and concepts through numerical values of distance and direction. LLM-generated text reflects the conceptual priorities of the language it was trained on, meaning multilingualism is a structural question when LLMs influence data flows. The 2003 United Nations Educational Scientific and Cultural Organization (UNESCO) recommendations on access to cyberspace declare digital multilingualism as essential to information rights. Upholding these norms today requires that linguistic diversity, and the underlying geometries of knowledge held in language, shape the AI outputs that contribute to contents of the internet.

### Key Points

- LLM-powered language technologies encode, compute and generate large swaths of available information online. Digital infrastructures are influenced by their rapid adoption.
- Natural languages hold knowledge about how ideas relate to one another through diverse category systems. Vectorization and word-embedding attempts to represent and operationalize this knowledge through computational methods.
- Where LLMs generate digital language, multilingualism is a structural need in addition to a question of representation.
- Digital governance that fosters universal access to cyberspace supports the diversification of both model geometries and of actors designing language technologies.

## Recommendations

- **Develop an updated set of standards that integrate norms on multilingualism and AI:** Update multilingualism and access recommendations to account for the influence of AI on digital architectures. Shaping emerging digital infrastructures to support prevailing information rights requires structural capacity for multilingualism. Developing updated standards provides public and private actors with a clear direction for orienting emerging technologies toward epistemic pluralism.
- **Audit models procured for public service according to a linguistic alignment protocol:** Prioritize models evaluated for multilingual obligations and/or language-specific computation aligned with the use case. The mapping of semantic meaning contributes to delimiting what kind of output an LLM can generate. As a result, responsible public deployment of LLM tools involves an evaluation of their computational priorities around language.
- **Prioritize funding community-led interventions to the language technology ecosystem:** Community-led LLM initiatives such as participatory curation of training data, linguistic resource digitization and fine-tuning existing generative pre-trained transformers fosters models that align with the cultural protocols of a language community. Multilingual design may be an approach to individual model development but also a strategy to diversify the larger digital ecosystem via many language-specific models.



# Introduction

The creation of the internet formed new digital architectures for ideas to flow worldwide with far-reaching implications for information rights, such as those enumerated in the United Nations Declaration of Human Rights: “everyone has the right to freedom of opinion and expression; this right includes freedom to hold opinions without interference and to seek, receive and impart information and ideas through any media and regardless of frontiers.”<sup>1</sup> In consideration of how these rights manifest in emerging digital infrastructures, in 2003 the United Nations Educational Scientific and Cultural Organization (UNESCO) created the Recommendation concerning the Promotion and Use of Multilingualism and Universal Access to Cyberspace, which remains relevant to standards setting today. This document provided a normative standard for member states such as Canada to safeguard equity among groups in the digital era. Building capacity for multilingualism featured heavily in the recommendations, as it reads in part that “linguistic diversity in the global information networks and universal access to information in cyberspace are at the core of contemporary debates and can be a determining factor in the development of a knowledge-based society.”<sup>2</sup> The standards set forth in the document are grouped into four key recommendations: developing multilingual content and systems, facilitating access to networks and services, developing public domain content and reaffirming the equitable balance between the interests of rights holders and the public interest.

Since 2003, what it means to create access to cyberspace has rapidly evolved. Generative artificial intelligence (AI) has increased the complexity of engaging with cyberspace by building vast infrastructures of cyberspace itself. Large language model (LLM)-based technologies encode, redistribute, translate and generate much of the text available online. Not only does language determine who is represented in the digital sphere, but linguistic information in the form of LLM training data shapes the output of what can be generated by a given model. Language technologies increasingly function as foundational layers of digital architecture, building the networks that allow for large-scale data flows. This working paper frames multilingualism not only as a measure of diversity but as a design approach to building the technologies that construct digital systems. It considers what it means to strive for the UNESCO standards set in 2003 in light of the rapid evolution of LLM technologies and their influence on digital infrastructures. Multilingualism and universal access to cyberspace are possible in a technological landscape shaped by machine learning; however, they require governance informed by how language technologies author spatial maps of meaning in the digital sphere.

---

1 *Universal Declaration of Human Rights*, GA Res 217A (III), UNGAOR, 3rd Sess, Supp No 13, UN Doc A/810 (1948), art 19.

2 *Recommendation concerning the Promotion and Use of Multilingualism and Universal Access to Cyberspace*, ESC Res 32 C, UNESCO, 32nd Sess (2003), online: <<https://unesdoc.unesco.org/ark:/48223/pf0000133171>>.

# Language Building Digital Architectures

## Geometries of Meaning

From a sociolinguistic point of view, natural language is already a spatial phenomenon. Meaning is co-created between people and their environments. The words in a language differentiate ideas in part by associating certain things with one another and dissociating others. This is part of a culturally specific system of organizing the many variables of reality in order to make sense of them. Knowing which concepts are associated, and how they are associated, is just one aspect that can contribute to the rich inheritance of cultural worldview.

Any system in which variables are oriented in relationship to one another can be mapped spatially. In languages, associations between terms are encoded in many geometric forms; beyond the linear relation of opposites and synonyms, the concepts in a category can be “‘graduated oppositions’ (burning, hot, tepid, cold, icy),” “‘implications’ (demobilized, immobilized)” or “‘relations of complementarity’ (husband, wife; theory, practice; hunger, thirst; sell, buy),” to name only a few examples in English (Perusset 2023). Hierarchy or gradients of association are examples of patterns potentially encoded into the knowledge system of a language. These patterns are spatial, and with any attempt to represent language through computation, accurate geometric association between terms is integral to cultural accuracy.

Nick Seaver (2017, 5) made the assertion that algorithms are culture: “not singular technical objects ... [but] rather unstable objects, culturally enacted by the practices people use to engage with them.” Algorithmic systems such as language models can be analyzed, in turn, not as singular technical objects but as highly complex amalgamations of cultural encodings. Representing a language through numerical embeddings, to the extent that a model can generate useful text, is the result of countless decisions made during training about how to encode meaning and draw relationships between words, sub-word units and concepts. LLMs generating and translating language online is a non-neutral interaction between the geometries of meaning set during training and the practices people use to engage with them. Especially when a model is trained on monolingual data, the geometry it then reproduces is specific to that language. Multilingual LLMs map and represent multiple languages, yet discrepancies in training data availability impact equitable quality among them.

Analyzing the priorities set during training, and the geometries thus performed by a given model, can be difficult when closed algorithmic systems are treated like a “black box ... knowable only through the relationship between inputs and outputs” (ibid.). Peeling back the layers to analyze how LLMs encode language offers insights into how systems of meaning are spatially organized in the computation itself. Geometry can already be seen as part of meaning making when considering how natural languages draw relationships between ideas. The internal architecture of a language model numerically organizes language so that some ideas appear closer, more relevant or more compatible than others. When these technologies contribute to the language that is present online, multilingualism is as much about the conceptual geometry in digital architectures as it is about representation.

## Contextual Word Embeddings

Approaches to designing the internal mechanisms of LLMs have continuously evolved, but the process of word embedding remains a throughline that enables the technology. As Charles Zhang et al. (2025, 1) write, “by mapping words into high-dimensional spaces where semantically related words are situated near each other, embeddings support nuanced language interpretation and have become essential to NLP applications such as machine translation, sentiment analysis, and information retrieval.” Earlier language models assigned fixed vectors (a quantity that has direction) to each word that represented its relative association to all other words in the language. These maps were built on an interpretation of meaning; how similar (close) or dissimilar (far) terms were from one another, based on co-occurrence in the training data. With this approach, meaning as captured by a given model is not stored in definitions but mapped as distance, clusters and directional associations. As previously found, the “conceptual proximity between words is an artefact of social values in the learning environment” (Legault 2026). Fixing these proximities through vectors created a technology that performed a given set of social values informed by the underlying geometry of the language it was trained on.

The innovation of a contextualized approach to word embedding offers new potential for LLM performance but also a challenge regarding the evaluation of its semantic geometry. Instead of each word having a fixed position, contextual word embedding can incorporate added information to a word’s numerical embedding depending on surrounding text. Sub-word units, sentences and even document-level information can all factor into the weighting of a word. Transformer-based systems such as BERT (Bidirectional Encoder Representations from Transformers) contain many layers that compute contextual information, influencing the value of a word embedding in a given instance according to the fine-tuning of the model’s layers. While this malleable mapping somewhat obscures how a model encodes its ontological proximities, strategies such as visualization techniques exist to decipher them. Zhang et al. (2025, 17) argue that “improving interpretability and explainability of embeddings is another critical area for future research. Although embeddings effectively capture semantic relationships, their abstract nature often makes it difficult to understand how specific representations are derived. This lack of transparency limits their applicability in domains requiring explainable AI.” LLM technologies continue to evolve but while word-embedding maps linguistic information spatially, interpretable geometry remains relevant to accuracy and language-specific conceptual organization. Systems trained primarily on high-resourced language corpora may better model context for those languages than for under-resourced languages, further underscoring the need for transparency and multilingual approaches.

## Orienting Ideas

By encoding the relationships between words, LLMs are using numbers to orient ideas in relation to one another. Sara Ahmed (2006) cites many philosophical works to approach the question of what it means to be oriented. The distinction between location and position, or “objective” space and relative space, is explored to suggest that orienting anything requires reference to other variables and a perception of the relationship between them. Word embedding has expanded the horizons of how language can be computed. When LLMs numerically graph the distance between concepts, they are

fixing an orientation and defining the relationship. Whether a model uses static vectors or contextual word embeddings, the output it generates is just one perception of mapped meaning. Algorithmically producing language necessitates conceptual choices regarding how ideas relate, approximating a performance of ontology. This inherently carries exclusions of orientation and position of alternative perceptions. Languages carry distinct conceptual geometries, thus multilingualism online relies on digital architectures built not only with multiple vocabularies but also with the capacity for diverse orientations.

## Findings from 2025 Language Technologies for All

The development and deployment of language technologies is having a global impact on both digital and linguistic landscapes. Following its 2003 recommendations, UNESCO (2025) held the 2025 Language Technologies for All conference to assess the convergence of AI and multilingualism, citing that “the ability to communicate, learn, and participate online depends heavily on language [yet] most of the world’s linguistic diversity remains excluded from the digital realm.” The proceedings and submissions from the conference offer reports from communities around the world and are openly accessible. The following is a scan of this data for themes and local examples of LLM applications offering insights on the creation of policy geared toward a multilingual digital ecosystem.

### LLM Training

Building the deep neural network required for an LLM has conventionally relied on vast quantities of training data. This is especially the case for multilingual LLMs and contextually variable models. Since frontrunning generative pre-trained transformer (GPT) tools scrape the web for available training data, the languages with the most representation online provide the base infrastructure for how they learn to weight and redistribute information. This results in LLMs whose word embeddings are geared to those languages (often English). Even if the resulting generated language gets translated, the underlying map of how concepts relate to one another may reflect the semantic proximities contained in the training data. Politeness systems, kinship terms, ecological vocabularies and ways of knowing about animism and interconnectedness are all examples of proximities that do not neatly translate across languages.

The full pipeline of LLM development, from pre-training to fine tuning to system prompting, may not be feasible (or welcome) for each natural language. However, the 2025 conference included many examples of localized LLMs built not from the internet at scale but rather from culturally specific datasets curated according to community protocols. EthioNLP, for example, is improving and fine tuning LLMs for the linguistically diverse context of Ethiopia: key strategies include a consolidation of resources on 83 regional languages, dataset creation and model building (Yimam 2025, 61). Heritage Lab “AI!” develops community-driven language technologies including translation programs with LLMs that are fine tuned according to Inuit cultural knowledge and Indigenous language sovereignty (Macdonald and Mehdi 2025, 85). The Centre for

the Luxembourgish Language is using “small, context-rich corpora, allowing for the advancements in computational linguistics for Luxembourgish that are shaped by deep language expertise and cultural embeddedness” (Morse 2025, 3). These examples demonstrate the potential of smaller data and localized interventions at varying points along the pipeline. A plurality of use-specific and geometrically precise models can contribute to structural linguistic diversity in the digital realm. Supporting these initiatives and procuring models aligned to use cases incentivize the research and development of technologies.

## Computation Is Not a Lingua Franca

The results of the 2025 conference reflect a tension between the idea of AI to function as a potential “universal translator” (Fusacchia 2025, 16) and the possible homogenizing effect of generalizing knowledge. Computational approaches such as “language-agnostic” models contain numerical representations for concepts applied across many languages (Chen and Zhang 2024). Since some concepts are interoperable between languages, this approach can be highly useful to expand the linguistic reach of technology where training data is low. Computation itself may be a universal system but with word embeddings formed by cultural information, a language-agnostic “latent space” is not necessarily a lingua franca (Tezuka and Inoue 2025). The findings report that “neuron activation patterns of LLMs exhibit similarities when processing the same language” but also that neuron activation patterns are similar “when processing sentences with the same semantic meaning in different languages” (Zeng et al. 2025, 33). Equal access to accurate language translation technologies may be facilitated by a negotiation between context-specific knowledge and generalized technical approaches.

A submission from Finland shared that “neither a common language nor AI-based translations can substitute for the creation and dissemination of knowledge in various languages” (Pölönen 2025, 69). Language communities have different cultural protocols meaning not all linguistic information is invited to be represented computationally in language-agnostic layers. As described by Te Taka Keegan regarding generative AI approaches to te reo Māori, linguistic encoding must abide by data sovereignty principles defined by the language community (Keegan 2025, 79). Digital approaches to multilingualism may be aided by AI translation as long as standards on data protection are culturally informed.

## Evaluating Accurate Geometry

In light of a rapid adoption of LLM technologies, the mapped meaning they encode holds the potential to influence knowledge distribution online. Evaluating these maps will be an important part of ushering in multilingual data flows. Open systems provide the opportunity for language-specific evaluation of word-embedding values. Pass/fail metrics for closed-system outputs can also provide some framework to evaluate semantic proximities. One example is the “geo-anchored approach” of Project Vaani, which provides a benchmark to evaluate the cultural knowledge of LLMs for the speech landscape of India (Talukdar 2025, 16). Technologies built for language families are more likely to balance interoperable geometry among dialects with cultural alignment. SEA-HELM (Southeast Asian Holistic Evaluation of Language Models) is one example of a holistic multicultural evaluation metric geared toward a family of languages (Montalan et al. 2025, 9).

# Conclusion

According to Cecily Raynor (2026), “the question before us is not how to make AI systems more multilingual: it is how linguistic plurality reshapes the distribution of epistemic power within global AI architectures.” Governance that promotes a multilingual design of these architectures will be key in striving for the standards set out in 2003. In a technological landscape where encoded meaning contributes to shaping digital infrastructures, “developing multilingual content and systems” requires many models shaped by distinct linguistic and cultural geometries rather than one universal architecture. “Facilitating access to networks and services” still relies on conventional infrastructures of broadband internet, and increasingly on community-led partnerships that co-develop systems according to cultural protocols for how languages may or may not be computed. “Developing public domain content” could entail augmenting transparency regarding word embedding and encoded meaning. Encouraging open source and interoperable solutions democratizes model training, inviting evaluation and iterative design. In this context, “reaffirming the equitable balance between the interests of rights-holders” and the public interest becomes a call for cohesive digital governance that decentralizes the authorship of data flows.

# Recommendations

- **Develop an updated set of standards that integrates digital multilingualism and AI:** Multilingualism in the global information networks is an arbiter of access. Numerically representing semantic meaning through word embedding now contributes to the architecture of data flows. As a result, creating standards that integrate linguistic diversity and digital governance is required. Multilingualism is a structural need in addition to a question of representation. Standards that are attuned to linguistic and epistemic pluralism in AI provide jurisdictional bodies with a clear direction for shaping architectures in line with information rights.
- **Audit models procured for public service according to a linguistic alignment protocol:** Standardize pass/fail metrics for the multilingual obligations of models. Prioritize model procurement that aligns technical geometries with the use case of the task. Evaluating language technologies along these lines incentivizes both transparency around encoded meaning, whether through training data disclosure or word-embedding values, and the development of interpretable systems.
- **Prioritize funding community-led interventions to the language technology sector:** Community-led LLM initiatives — such as the participatory curation of training data, linguistic resource digitization and fine-tuning of broader GPTs — foster models that align with cultural protocols. Support for this type of work intervenes in the pipeline of LLM development, diversifying the sector to include large and small actors. Multilingual design can be an approach to building an individual language model but also a strategy for cultivating a diverse ecosystem of language-specific systems.

## Acknowledgements

Thank you to Dianna H. English and Meaghan Dietrich at the Digital Policy Hub for the opportunity to do work that explores the humanness of language in an everchanging digital world. I am grateful to the 2025–2026 cohort of fellows for a year of rich transdisciplinary collaboration. This working paper significantly benefited from the expertise and guidance of Amal Hussein (peer review), Dr. Cecily Raynor (supervision) and Dr. Matthew da Mota (CIGI senior fellow).

## Works Cited

- Ahmed, Sara. 2006. *Queer Phenomenology: Orientations, Objects, Others*. Durham, NC: Duke University Press. <https://doi.org/10.1215/9780822388074>.
- Chen, Xiao and Chirui Zhang. 2024. "Language-agnostic Zero-Shot Machine Translation with Language-specific Modeling." In *Proceedings of the 2024 International Joint Conference on Neural Networks*, 1–8. <https://doi.org/10.1109/IJCNN60899.2024.10649983>.
- Fusacchia, Alessandro. 2025. "Towards a Truly Universal Translator?" In *LT4All 2025 Advancing Humanism through Language Technologies Book of Abstracts*, 16. [www.lt4all2025.eu/wp-content/uploads/2025/02/LT4All-2025\\_Book-of-Abstracts\\_Final.pdf](http://www.lt4all2025.eu/wp-content/uploads/2025/02/LT4All-2025_Book-of-Abstracts_Final.pdf).
- Keegan, Te Taka. 2025. "Testing Māori Data Sovereignty Principles with AI Models." In *LT4All 2025 Advancing Humanism through Language Technologies Book of Abstracts*, 79. [www.lt4all2025.eu/wp-content/uploads/2025/02/LT4All-2025\\_Book-of-Abstracts\\_Final.pdf](http://www.lt4all2025.eu/wp-content/uploads/2025/02/LT4All-2025_Book-of-Abstracts_Final.pdf).
- Legault, Anna. 2026. "Language Translation in the AI-Augmented Era." Digital Policy Hub Working Paper. [www.cigionline.org/static/documents/DPH\\_Fall2025\\_Paper\\_Legault.pdf](http://www.cigionline.org/static/documents/DPH_Fall2025_Paper_Legault.pdf).
- Macdonald, Natasha Ita and Ali Mehdi. 2025. "Ail: Building Inclusive Pathways to Indigenous Language Learning and Knowledge Access." In *LT4All 2025 Advancing Humanism through Language Technologies Book of Abstracts*, 85. [www.lt4all2025.eu/wp-content/uploads/2025/02/LT4All-2025\\_Book-of-Abstracts\\_Final.pdf](http://www.lt4all2025.eu/wp-content/uploads/2025/02/LT4All-2025_Book-of-Abstracts_Final.pdf).
- Montalan, Jann Railey, Jian Gang Ngui, Yosephine Susanto and William Chandra Tjhi Tjhi. 2025. "Sama-sama: Bringing Southeast Asia Together to Build Better LLMs." In *LT4All 2025 Advancing Humanism through Language Technologies Book of Abstracts*, 9. [www.lt4all2025.eu/wp-content/uploads/2025/02/LT4All-2025\\_Book-of-Abstracts\\_Final.pdf](http://www.lt4all2025.eu/wp-content/uploads/2025/02/LT4All-2025_Book-of-Abstracts_Final.pdf).
- Morse, Christopher. 2025. "Redefining Progress in the Margins: Lessons from Big Tech & Small Data." In *LT4All 2025 Advancing Humanism through Language Technologies Book of Abstracts*, 3. [www.lt4all2025.eu/wp-content/uploads/2025/02/LT4All-2025\\_Book-of-Abstracts\\_Final.pdf](http://www.lt4all2025.eu/wp-content/uploads/2025/02/LT4All-2025_Book-of-Abstracts_Final.pdf).
- Perusset, Alain. 2023. "How Post structural Semiotics Models Categories: From Greimasian Semantics to Categorial Semiotics." *Signata* 14. <https://doi.org/10.4000/signata.4556>.
- Pölonen, Janne. 2025. "Reform of Research Assessment, Lingua Franca, and AI-Based Translation." In *LT4All 2025 Advancing Humanism through Language Technologies Book of Abstracts*, 69. [www.lt4all2025.eu/wp-content/uploads/2025/02/LT4All-2025\\_Book-of-Abstracts\\_Final.pdf](http://www.lt4all2025.eu/wp-content/uploads/2025/02/LT4All-2025_Book-of-Abstracts_Final.pdf).

- Raynor, Cecily. 2026. "Regional GPTs, Linguistic Pluralism, and the Reconfiguration of Global AI Power." YouTube video, 1:55:01. The 3rd International Symposium on Humanities and Culture. [www.youtube.com/live/Vjxl9qasq-U](http://www.youtube.com/live/Vjxl9qasq-U).
- Seaver, Nick. 2017. "Algorithms as culture: Some tactics for the ethnography of algorithmic systems." *Big Data & Society* 4 (2): 1–12. <https://doi.org/10.1177/2053951717738104>.
- Talukdar, Partha Pratim. 2025. "Towards Linguistically and Culturally Inclusive LLMs." In *LT4All 2025 Advancing Humanism through Language Technologies Book of Abstracts*, 16. [www.lt4all2025.eu/wp-content/uploads/2025/02/LT4All-2025\\_Book-of-Abstracts\\_Final.pdf](http://www.lt4all2025.eu/wp-content/uploads/2025/02/LT4All-2025_Book-of-Abstracts_Final.pdf).
- Tezuka, Hinata and Naoya Inoue. 2025. "The Transfer Neurons Hypothesis: An Underlying Mechanism for Language Latent Space Transitions in Multilingual LLMs." In *Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing*, 31742–92. Suzhou, China: Association for Computational Linguistics. <https://doi.org/10.18653/v1/2025.emnlp-main.1618>.
- UNESCO. 2025. "UNESCO Launches Global Roadmap for Multilingualism in the Digital Era." UNESCO. November 10. [www.unesco.org/en/articles/unesco-launches-global-roadmap-multilingualism-digital-era](http://www.unesco.org/en/articles/unesco-launches-global-roadmap-multilingualism-digital-era).
- Yimam, Seid Muhie. 2025. "Advancing Language Technology for Ethiopia's Diverse Linguistic Landscape through the EthioNLP Collaborative Effort." In *LT4All 2025 Advancing Humanism through Language Technologies Book of Abstracts*, 61. [www.lt4all2025.eu/wp-content/uploads/2025/02/LT4All-2025\\_Book-of-Abstracts\\_Final.pdf](http://www.lt4all2025.eu/wp-content/uploads/2025/02/LT4All-2025_Book-of-Abstracts_Final.pdf).
- Zeng, Hongchuan, Senyu Han, Lu Chen and Kai Yu. 2025. "Converging to a Lingua Franca: Evolution of Linguistic Regions and Semantics Alignment in Multilingual Large Language Models." In *LT4All 2025 Advancing Humanism through Language Technologies Book of Abstracts*, 33. [www.lt4all2025.eu/wp-content/uploads/2025/02/LT4All-2025\\_Book-of-Abstracts\\_Final.pdf](http://www.lt4all2025.eu/wp-content/uploads/2025/02/LT4All-2025_Book-of-Abstracts_Final.pdf).
- Zhang, Charles, Benji Peng, Xintian Sun, Qian Niu, Junyu Liu, Keyu Chen, Ming Li, Pohsun Feng, Ziqian Bi, Ming Liu et al. 2025. "From Word Vectors to Multimodal Embeddings: Techniques, Applications, and Future Directions for Large Language Models." Preprint, arXiv. <https://doi.org/10.48550/arXiv.2411.05036>.