Background Materials

Source	Why Relevant?
"Low-Resource Languages Jailbreak GPT-4" by the Montreal Ai Ethics institute (2024)	"(...) we can bypass GPT-4's safety guardrails easily with translations. By translating unsafe English inputs, such as "how to build explosive devices using household materials," into low-resource languages such as Zulu, we can obtain responses that get us to our malicious goals nearly 80% of the time. This cross-lingual vulnerability arises because safety research focuses on high-resource languages like English. Previously, this linguistic inequality in AI development mainly affected low-resource language speakers. Still, it poses **safety risks** for all users because anyone can exploit LLMs' cross-lingual safety vulnerabilities with publicly available translation services. Our work emphasizes the pressing need to embrace more holistic and inclusive safety research."
"Data-invisible groups and data minimization in the deployment of AI solutions: policy brief" by UNESCO (2023)	"While AI's deployment and uptake undoubtedly provide humanity with numerous opportunities to address global challenges, the data used for AI systems can create risks that must be addressed to avoid undesirable outcomes. In order to fully reap the benefits of AI and fulfill commitments to our common future, data and statistics must be generated and leveraged to ensure that everyone has a voice and is visible. Despite this noble pledge, AI's contemporary effective deployment is exposing many inequalities by creating **data invisible groups**. Traditionally underserved and vulnerable populations, such as persons with disabilities, refugees, migrants, and individuals in the LGBT community, make up the strata of data invisible groups. These individuals and communities will most likely remain invisible without creating inclusive data systems guided by principles for data minimization and sharing."
"Aya Dataset: An Open-Access Collection for Multilingual Instruction Tuning", by Cohere AI (2024)	"The factors underlying the construction of the datasets impact how models perform for users around the world. Models perform better on the distribution they are trained to mimic. This often introduces known **biases** towards languages and dialects not included during training and introduces **critical security flaws**. Datasets aren't simply raw materials that fuel breakthroughs but also **make the poor poorer and the rich richer**. Disparities in the access to technological resources pre-dates the advent of LLMs. However, as LLMs become more sophisticated and widely available, non-English languages will remain underrepresented and will likely become more so. The imbalance between languages has created a growing divide in the cost of using this technology as **marginalized languages require more tokens** and **incur higher latency for generations**, consigning speakers of low performing languages to lower quality technology. Often, speakers of low-resource languages do **not have the resources to improve NLP technology** for their language, facing a low-resource double bind with limited access to both compute and data."
How African NLP Experts Are Navigating the Challenges of Copyright, Innovation, and	"In the context of digital technology and software, the Global South inclusion project has often been underpinned by a requirement of openness. The intention has been to promote broader access and address and/or sidestep privacy and copyright issues arising from both the data needed to build AI systems and the datasets that are one outcome of building and using such systems.2 Essentially, the Global South inclusion project benefits from pushing for openness because in

Access, by Chijioke Okorie and Vukosi Marivate (2024)	many instances, once data has been made open, it allows for a sidestepping of privacy and copyright issues by users of such data. However, there are more factors related to the Global South inclusion project to consider and grapple with. First, builders of AI systems need to give greater consideration to the communities directly or indirectly providing the data used in commercial and noncommercial settings for AI development. These communities may include owners of traditional cultural expressions and traditional knowledge; data scientists and AI developers from African countries working on data collection, collation, curation, and annotation; linguists working on African languages; and users who provide or upload content (data) on African languages and practices on social media and other internet platforms.3 However, while openness in developing and deploying AI models offers transparency and shared learning, it can sometimes conflict with privacy and proprietary rights. By contrast, closed models prioritize proprietary information but can limit shared innovation."
"Canada needs a sovereign wealth fund – built by monetizing our personal data" by Kean Birch (The Globe and Mail, 2024)	Proposed idea: "Sovereign wealth fund built by monetizing [...] data".
"Towards a More Inclusive AI: Progress and Perspectives in Large Language Model Training for the Sámi Language," by Ronny Paul, Himanshu Buckchash, Shantipriya Parida, and Dilip K. Prasad (Silo AI, 2024)	
"Building Machine Translation Systems for the Next Thousand Languages," by Google Research	
"The Ghanaian founder	

challenging Google", by Rest of the World	
"No Language Left Behind: Scaling Human-Centered Machine Translation," by Facebook's NLLB Team	Driven by the goal of eradicating language barriers on a global scale, machine translation has solidified itself as a key focus of artificial intelligence research today. However, such efforts have coalesced around a small subset of languages, leaving behind the vast majority of mostly low-resource languages. What does it take to break the 200 language barrier while ensuring safe, high quality results, all while keeping ethical considerations in mind? In No Language Left Behind, we took on this challenge by first contextualizing the need for low-resource language translation support through exploratory interviews with native speakers. Then, we created datasets and models aimed at narrowing the performance gap between low and high-resource languages. More specifically, we developed a conditional compute model based on Sparsely Gated Mixture of Experts that is trained on data obtained with novel and effective data mining techniques tailored for low-resource languages. We propose multiple architectural and training improvements to counteract overfitting while training on thousands of tasks. Critically, we evaluated the performance of over 40,000 different translation directions using a human-translated benchmark, Flores-200, and combined human evaluation with a novel toxicity benchmark covering all languages in Flores-200 to assess translation safety. Our model achieves an improvement of 44% BLEU relative to the previous state-of-the-art, laying important groundwork towards realizing a universal translation system.
IrokoBench: A New Benchmark for African Languages in the Age of Large Language Models, by Adelani et al (June, 2024)	Despite the widespread adoption of Large language models (LLMs), their remarkable capabilities remain limited to a few high-resource languages. Additionally, many low-resource languages (e.g., African languages) are often evaluated only on basic text classification tasks due to the lack of appropriate or comprehensive benchmarks outside of high-resource languages. In this paper, we introduce IrokoBench—a human-translated benchmark dataset for 16 typologicallydiverse low-resource African languages covering three tasks: natural language inference (AfriXNLI), mathematical reasoning (AfriMGSM), and multi-choice knowledge-based QA (AfriMMLU). We use IrokoBench to evaluate zero-shot, few-shot, and translate-test settings (where test sets are translated into English) across 10 open and four proprietary LLMs. Our evaluation reveals a significant performance gap between high-resource languages (such as English and French) and low-resource African languages. We observe a significant performance gap between open and proprietary models, with the highest performing open model, Aya-101 only at 58% of the best-performing proprietary model GPT-4o performance. Machine translating the test set to English before evaluation helped to close the gap for larger models that are English-centric, like LLaMa 3 70B. These findings suggest that more efforts are needed to develop and adapt LLMs for African languages.
"The AI Language Gap", by Cohere for AI (June, 2024)	More than 7000 languages are spoken around the world today, but current, state-of-art AI large language models cover only a small percentage of them and favor North American language and cultural perspectives. This is in part because many non-English languages are considered "low-resource," meaning they are less prominent within computer science research and lack the high-quality datasets

	necessary for training language models. This language gap in AI has several undesirable consequences: 1) Many language speakers and communities may be left behind as language models that do not cover their language become increasingly integral to economies and societies. 2) The lack of linguistic diversity in models can introduce biases that reflect Anglo-centric and North American viewpoints, and undermine other cultural perspectives. 3) The safety of all language models is compromised without multilingual capabilities, creating opportunities for malicious users and exposing vulnerable users to harm. There are many global efforts to address the language gap in AI, including Cohere For AI's Aya project — a global initiative that has developed and publicly released multilingual language models and datasets covering 101 languages. However, more work is needed. To contribute to efforts to address the AI language gap, we offer four considerations for those working in policy and governance around the world: 1) Direct resources towards multilingual research and development. 2) Support multilingual dataset creation. 3) Recognize that the safety of all language models is improved through multilingual approaches. 4) Foster knowledge-sharing and transparency among researchers, developers, and communities.
Artificial Intelligence, Regulation, and the Right to Development - Thematic study by the Expert Mechanism on the Right to Development	