

International Journal of Learning, Teaching and Educational Research
Vol. 23, No. 12, pp. 148-173, December 2024
<https://doi.org/10.26803/ijlter.23.12.9>
Received Sep 1, 2024; Revised Oct 14, 2024; Accepted Dec 28, 2024

Currently Available GenAI-Powered Large Language Models and Low-Resource Languages: Any Offerings? Wait Until You See

Chaka Chaka* 

University of South Africa
Pretoria, South Africa

Abstract. A lot of hype has accompanied the increasing number of generative artificial intelligence-powered large language models (LLMs). Similarly, much has been written about what currently available LLMs can and cannot do, including their benefits and risks, especially in higher education. However, few use cases have investigated the performance and generative capabilities of LLMs in low-resource languages. With this in mind, one of the purposes of the current study was to explore the extent to which seven, currently available, free-to-use versions of LLMs (ChatGPT, Claude, Copilot, Gemini, GroqChat, Perplexity, and YouChat) perform in five low-resource languages (isiZulu, Sesotho, Yoruba, Māori, and Mi'kmaq) in their generative multilingual capabilities. Employing a common input prompt, in which the only change was to insert the name of a given low-resource language and English in each case, this study collected its datasets by inputting this common prompt into the seven LLMs. Three of the findings of this study are noteworthy. First, the seven LLMs displayed a significant lack of generative multilingual capabilities in the five low-resource languages. Second, they hallucinated and produced nonsensical, meaningless, and irrelevant responses in their low-resource language outputs. Third, their English responses were far better in quality, relevance, depth, detail, and nuance than their low-resource language only and English responses for the five low-resource languages. The paper ends by offering the implications and making the conclusions of the study in terms of LLMs' generative capabilities in low-resource languages.

Keywords: generative multilingual capabilities; hallucinations; large language models; low-resource languages; nonsensical, meaningless and irrelevant responses

* Corresponding author: *Chaka Chaka*, chakachaka8@gmail.com

1. Introduction

It is becoming increasingly obvious that English is the first and home language of the currently existing generative artificial intelligence-powered large language models (henceforth simply LLMs) (Snyder, 2023; Vashee, 2023). English is their default language. Here, Huang et al.'s (2023) dictum and main paper titled, "Not all languages are created equal in LLMs" (p. 1), is more than instructive. Of course, English does not dominate only LLMs' ecosystems, but also the Internet's ecosphere, in which languages such as French, Spanish, Portuguese, Arabic and Mandarin have their own share of dominance. The other European languages and/or Latin alphabet-based languages and certain non-European languages like Chinese and Indonesian (Akula et al., 2024; Snyder, 2023) have their own dominant share, too. All of the languages, which have a strong presence on the Internet, and which are concomitantly preferred by LLMs are positively regarded as high-resource languages. In contrast, all languages that have little or no Internet presence, and which are consequently not used by LLMs, are negatively referred to as low-resource languages. Most of these languages are marginalized Indigenous and subaltern languages. The majority of speakers of these languages are on the periphery not only of leading-edge AI developments but of many other technological developments currently taking place as well.

However, there is a contrast that needs to be drawn between the Internet and LLMs in terms of their preferred languages. The Internet was not *born* speaking English per se. It has been and continues to be provisioned with information, data, materials, and texts written mainly in English. In relation to higher education (HE), in particular, some of the pieces of information, data, materials, and texts comprise research data, journal articles, books, and monographs written in English by speakers of low-resource languages, which these speakers themselves or the publishers of their work make available online. Then, the Internet simply harvests and archives this published work, together with its biases and forms of misrepresentation in certain instances, as it is in its original language of publication, which happens to be English. This is not to deny that at times, Internet search engine algorithms tend to be biased towards certain provisioned information, while they are biased against some (see Cave & Dihal, 2020; Chaka, 2022; Lee et al., 2019; Lin et al., 2023). So, at issue for Internet search engines is the provisioned information available online, which may have its own bias, search engines' own algorithmic bias and the opaque realm in which search engines tend to operate.

Contrariwise, LLMs largely depend on the specific language through which their training data is made available and fed into them. The other factors that undergird them are their training data's quality and inclusiveness, their algorithm designs and their cross-domain generalisation (AIContentfy Team, 2023; Captain Words, 2024). Data quality has to do with the comprehensiveness and robustness of the training data while data inclusiveness is related to the diversity and representativeness of data not only in terms of language and dialects, but also concerning racial and ethnic demographics, geographies, gender, cultures, and value systems. Algorithms are often designed to be effective, reliable, accurate, and value-neutral (unbiased) (AIContentfy Team, 2023; Chaka, 2022, 2024a,

2024b; Rudolph et al., 2024). Cross-domain generalization pertains to the extent to which an LLM's dataset is generalizable and applicable to different domains of use that exist in real life (Wu et al., 2023). Allied to cross-domain generalization is the cross-linguistic generalization of LLMs. On one hand, this relates mainly to the extent to which LLMs have their training data in languages other than English. On the other, this factor has to do with LLMs' training datasets being available in other Internet-marginalized languages rather than just in English and other Internet-favoured languages. Cross-linguistic generalization allows LLMs to operate efficiently, reliably, and accurately across diverse languages. Elsewhere, Wu et al. (2023) refer to this language capability as a cross-linguistic applicability.

Against this background, this study set out to investigate the extent to which seven, available, free-to-use LLMs like ChatGPT (ChatGPT 3.5), Claude, Copilot, Gemini, GroqChat (Llama3-8b-8192), Perplexity and YouChat, perform in five low-resource languages, isiZulu, Sesotho, Māori, Yoruba, and the Mi'kmaq language (henceforth Mi'kmaq), in their generative multilingual capabilities. It also sought to explore the extent to which the responses generated by these seven LLMs in such low-resource languages are meaningful, sensible, and relevant, based on a common input prompt. In addition, the study wanted to compare the responses for the five low-resource languages with those for English. Requiring responses from LLMs based on prompts is what Hadi et al. (2023) refer to as a question-answering system. English was used as a benchmark high-resource language in the current study. In view of this, this study had the following research questions (RQs):

- To what extent do the seven, currently available, free-to-use LLMs such as ChatGPT, Claude, Copilot, Gemini, GroqChat, Perplexity, and YouChat, perform in the five low-resource languages, isiZulu, Sesotho, Māori, Yoruba, and Mi'kmaq in their generative multilingual capabilities?
- To what extent do these seven LLMs provide meaningful, sensible, and relevant responses from a common prompt in the five low-resource languages?
- To what extent do the responses generated by these seven LLMs in the five low-resource languages compare with the English responses generated by the same LLMs regarding the same common input prompt?

2. Argumentative Standpoint

While there is global hype about currently available, free-to-use LLMs and while some scholars tend to tout these LLMs as revolutionary and disruptive, especially for university knowledge generation and university teaching and learning, the current paper adopts a cautionary posture. In this evolving state of affairs, the sudden emergence of these LLMs has not only led to an AI arms race reminiscent of yester-years' space race, but it has also led to a sudden growth of armchair experts and gurus of AI in all its shapes and permutations. These armchair experts and gurus tend, at times, to occupy two polar sides of the AI equation – embracers and resisters (Luddites) of the new AI technology.

At the core of this cautionary posture is a degree of criticism and some scepticism of AI so that one does not blindly believe in AI technology to the point of embracing technologism. At the same time, this posture guards against stoic Luddism: stubbornly resisting or rejecting new technology for the sake of resisting or rejecting it. Rather, it maintains that the currently available LLMs need to be treated with caution concerning what they can and cannot do in HE, regarding what languages they cover and those they do not cover in their training data. Therefore, this cautionary critical-sceptical posture mainly has to do with the fact that most of the currently available LLMs often tend to hallucinate or make up information about the factual knowledge they generate (Hadi et al., 2023; Perkins, 2023; Popenici, 2023; Rudolph et al., 2023; Rudolph et al., 2024). This is the case even when that factual knowledge exists in high-resource languages such as English in which LLMs' training data is available and packaged. The cautionary critical-sceptical posture is also related to the fact that most of the currently available LLMs tend not to accommodate low-resource languages: they are biased towards high-resource languages in the same way as the Internet is biased towards these languages. So, whatever these LLMs may claim to be able to do needs to be treated with critical scepticism. This serves as the argumentative standpoint in this paper.

3. Related Literature

With the points highlighted above in mind, this paper provides a short, bespoke literature review of some of the scholarly papers that deal with LLMs and low-resource languages. As this is an emerging and evolving area, there are not many current studies that have focused on LLMs and low-resource languages. Therefore, three studies, Nguyen et al. (2023), Lankford et al. (2023), and Huang et al. (2023), which have relevance to this study, are briefly reviewed here. The first and last ones are preprints. For instance, Nguyen et al. (2023) point out that while LLMs have profound generative capabilities in high-resource languages, they nonetheless have constrained generative capabilities in low-resource languages owing to their inherent pre-training data asymmetry. So, to compensate for this deficiency, they collated in-context, synthetic, intra-lingual exemplars from varied datasets of high-resource languages and employed them to prompt LLMs to translate from given low-resource languages into English. This method, which they refer to as a linguistically-diverse prompting (LDP), was applied to 21 African and 13 Indic low-resource languages, and was used to perform translations and summarisation between these languages and English. It was then used to generate in-context, synthetic, intra-lingual exemplars to carry out generative tasks (translation and summarisation) in these target low-resource languages. Tapping into the ROOTs corpus, the LDP method was applied to the BLOOM model and InstructGPT (Nguyen et al., 2023).

Two of the results of this study are worth mentioning. First, the performance of LDP equalled supervised few-shot learning when zero supervision in English to and from 21 African and 13 Indic low-resource languages was employed. This LDP approach even outperformed in non-English-wide directions. Second, the approach outdid related English-pivoting methods in multilingual summarisation (Nguyen et al., 2023). Nguyen et al.'s (2023) use of LDP

demonstrates that current LLMs lack generative multilingual capabilities for low-resource languages. Importantly, it highlights the need to improvise or innovate in order to compensate for this inherent generative deficiency. It is this aspect – the generative multilingual deficiencies LLMs have when it comes to low-resource languages – that necessitates a critical-sceptical posture.

The second study is Lankford et al. (2023). This study dealt with the impact of LLMs on efficient machine translation (MT) outputs related to two low-resource languages, Marathi and Irish. Its overall objective was to address the shortcomings LLMs have in delivering high-quality MT outputs for these languages by developing a tool called adaptMLLM, with MLLM standing for multilingual large language models. To realise its objective, the study focused on fine-tuning pre-built MLLMs to improve MT on two low-resource language pairs. These pairs involved English to Marathi (ENG ↔ MR) and English to Irish (ENG ↔ GA). When the adaptMLLM system was compared and benchmarked against the baselines from the LoResMT2021 Shared Task, it generated improved translation outputs. The improved translation outputs were noticed bi-directionally in ENG ↔ MR pairs and in ENG ↔ GA pairs, respectively (Lankford et al., 2023). Lankford et al.'s (2023) adaptMLLM, like Nguyen et al.'s (2023) LDP above, represents an attempt at improvising and innovating LLMs to atone for generative deficiencies LLMs have in low-resource languages. This, again, emphasises the need for adopting a critical-sceptical posture when dealing with LLMs' performance in low-resource languages.

The third study that has some relevance for this paper is Huang et al. (2023). This study set out to experiment with a cross-lingual-thought prompting (XLT) method intended to iteratively enhance the multilingual capabilities of LLMs across high-resource and low-resource languages. XLT is a general method for prompting that is able to trigger cross-lingual and logical reasoning to improve task performance in diverse languages. This method was used to evaluate seven representative benchmarks dealing with understanding (e.g. natural language inference and paraphrasing), reasoning (e.g. arithmetic reasoning and common-sense reasoning), and generation (e.g. question answering, summarisation, and machine translation) tasks related to high-resource and low-resource languages. In all, the experiment involved 27 different languages, among which English, French, German, Spanish, Italian, Russian, Japanese, and Chinese Simplified represented some of the high-resource languages. Low-resource languages were represented by languages such as Swahili, Bengali, Tamil, Galician, Urdu, Telugu, Javanese, Haitian Creole, and Southern Quechua. Two LLMs, GPT-3.5-Turbo and Text-Davinci-003, were used in the experiment. Notably, this experiment demonstrated that XLT significantly improved the performance of different multilingual tasks and markedly minimised the gap between the best performance and the mean performance of each task in the various languages. Crucially, XLT generated more than 10 points of mean improvement in open-domain question-answering and arithmetic reasoning (Huang et al., 2023). In the same way as the two studies discussed above, this XLT experiment underscores the kind of improvisation and innovation needed to compensate for the multilingual generative shortcomings LLMs have for low-resource languages.

Once more, this calls for a critical-sceptical posture when one deals with LLMs' performance in low-resource languages.

The matter of cross-linguistic or multilingual capabilities of LLMs is crucial and pressing since scholars like Snyder (2023) and Qin et al. (2024) argue that there are over 7,000 global languages. Therefore, the need for MLLMs in the exponentially growing LLM ecosystem cannot be overemphasised. Of course, no illusions must be harboured about MLLMs, either. Like their currently available, high-resource language-biased LLM cousins, MLLMs too will have their shortcomings. Chief among these shortcomings, but by no means the only one, is hallucination. LLM hallucination is a practice in which LLMs misinterpret objects or patterns and produce responses that seem to be factually true, but which are inaccurate, erroneous or nonsensical (Aboze, 2023; Huang et al., 2023; IBM, 2024; Patil & Gudivada, 2024; Guerreiro et al., 2023). It occurs when LLMs tend to make up facts or simply churn out untruths, which are embedded in plausible-looking statements. As IBM (2024) opines, at face value, this phenomenon appears to be counterintuitive as it is often associated with humans or animals, and hardly with AI tools. In fact, it is difficult for one to think of an LLM that currently cannot hallucinate, notwithstanding the continuing advancements of LLMs. For instance, Qin et al. (2024) contend that MLLMs are often prone to hallucination. Other scholars such as Dale et al. (2023) and Guerreiro et al. (2023) have flagged the practice of hallucination in multilingual machine translation. Similarly, Aharoni et al. (2024) have explored the hallucination generated by multilingual machine summarisation. This calls into question the blind adoption and usage in certain quarters of academia of LLMs, in both high-resource languages and low-resource languages, when actually hallucination is a characteristic feature for both LLMs and MLLMs. This is where the critical-sceptical posture adopted in this paper comes into the picture, and not a blind, ardent and blanket adoption and usage of LLMs.

4. Method

This study focused on the use cases of the seven, currently available LLMs from the point of view of an end-user, who, in this case, was the author of this paper. As mentioned above, these LLMs comprised ChatGPT, Claude, Copilot, Gemini, GroqChat, Perplexity, and YouChat. At the time when the study was conducted, the first four LLMs had a persistent online presence on both Google and Bing search engines, while the last three did not. The overall aim, then, was to select well-known and less-known LLMs in this study for diversification purposes.

The study did not improvise and innovate by employing an experimental tool to compensate for the generative multilingual deficiencies these LLMs might have pertaining to the five low-resource languages, which were part of its use cases. Instead, it compared the seven LLMs' responses to a common lesson plan prompt in five low-resource languages (isiZulu, Sesotho, Yoruba, Māori, and Mi'kmaq (also known as Mi'kmawí'simk) with their counterpart English responses. In view of this, this study was exploratory in nature as it set out to explore a phenomenon or an aspect that has not yet been extensively studied (Chaka, 2024b). This phenomenon is the generative multilingual capabilities of the seven LLMs in the

five aforesaid low-resource languages versus their language capabilities in English (a high-resource language) based on a given common lesson plan prompt.

4.1 Materials and Data Collection Procedure

The process of collecting data for this study took place between 23 March 2024 and 29 March 2024, and again on 23 April 2024. The five low-resource languages mentioned above were selected to represent three Indigenous languages in Africa (isiZulu, Sesotho, and Yoruba) and to represent two more Indigenous languages (Māori and Mi'kmaq) outside of Africa.

After selecting the LLMs, a bespoke lesson plan prompt (hereafter, the common prompt or the prompt) was formulated and entered into each of the seven LLMs mentioned above. The common prompt was phrased as follows:

Write me a class lesson in (*a specific name of one of the five low-resource languages/or English*) on the following topic, "Southern multilingualism". Make sure the lesson has all the necessary sections and teaching aids, a class activity and an assignment. Also, provide a useful tip for students about Southern multilingualism in the context of AI-powered large language models.

The first LLM to be queried with this prompt was Copilot (see Figure 1) and the last was ChatGPT. No regenerate prompt or re-prompting was used for all seven LLMs. For each of the seven LLMs, the output response (henceforth the response) generated from the input prompt mentioned above was translated into English using Google Translate if it had been generated exclusively in the target low-resource language as spelt out in the prompt (see Figure 1). Where necessary, Bing Translate, Machine Translation.com, and Rytr were used for translation purposes as well. The latter, Rytr, was used for translating Mi'kmaq as the other three online machine translation tools could not translate it as a marginalized, minority, Indigenous language. If a response was generated in English, counter to the instruction in the prompt, it was left untranslated.

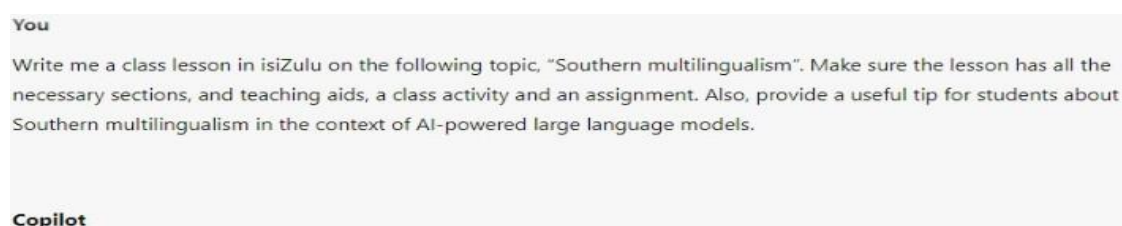


Figure 1: A screenshot of the lesson plan prompt for isiZulu as displayed on Copilot

There was a corresponding English prompt version used. In all, the data for this study comprised the following datasets: the main low-resource language datasets; the English dataset; and the English Google-translated version of each low-resource language dataset if it (the dataset) had been generated exclusively in the given target low-resource language.

4.2 Data Analysis

Content analysis, comprising manifest content analysis and latent content analysis, was used to analyse the datasets collected for this study. Manifest content analysis entails developing and constructing context from raw words or phrases related to the text at hand to derive literal meanings from those words or phrases. By its nature, it is, as its name indicates, a deductive analysis, conducted from macro-elements to micro-elements. It focuses on the surface structures of the text. By contrast, latent content analysis is an inductive analysis that starts from the micro-elements of a text and culminates in analysing the macro-elements of the text at hand. It deals with the deeper, hidden structures of the text (see Delve & Limpaecher., 2022; Kleinheksel et al., 2020). Both forms of analysis involve quantitative analysis in varying degrees.

Pertaining to the current study, the manifest content analysis focused on the raw responses generated by the seven LLMs based on the common prompt mentioned earlier. For example, in keeping with this form of analysis, the surface structures of all the raw responses generated by the seven LLMs were analysed as such without adding any layer of meaning to them. This entailed dealing with the straightforward, surface meanings directly observable from the responses (e.g. nonsensical and incomprehensible responses) together with the attendant literal implications of such responses. Contrarily, latent content analysis went beyond the surface nonsensical and incomprehensible meanings of these responses and attributed such meanings to, among other things, the phenomenon of hallucination. The latter is not manifest in the LLMs' responses: it was latently inferred. In other words, it is a latent inference. The manifest content or data is presented under the findings followed by a discussion of the findings. This latter section offers the latent content and its underlying structures.

5. Findings

The first part of the findings presented in this section relates to the number of words each LLM generated for each of its responses to the same common prompt that was inputted to it for each low-resource language and English as a high-resource language (see Table 1). The second part of the findings provides sample responses that were generated by the seven LLMs. Where necessary, the Google-translated versions of these sample responses or the English versions of these sample responses as translated by the other three MT tools are offered. The two parts are presented as instances of manifest content. As shown in Table 1, the seven LLMs generated responses with varying word counts for each of the five low-resource languages and English. When the seven LLMs are taken together, YouChat generated the highest total word count (3,601 words) for its six language responses, which was boosted by Yoruba response (1,122 words). It was followed by ChatGPT (2,790 words), with Claudia having generated the lowest total word count of 1,966 words. A notable exception is Gemini, which generated a zero response for isiZulu by disclaiming that "I'm still learning languages ...". Still, it managed to dwarf Claudia into the last position. Save for Gemini's isiZulu's zero word count, the lowest word count for any of these seven LLMs is that of the response generated by Perplexity for Yoruba (n=224 words).

Table 1: Seven LLMs, five low-resource languages and their responses, English and its responses, and response word counts

LLMs	Low-resource language 1 (IsiZulu)	Low-resource language 2 (Sesotho)	Low-resource language 3 (Yoruba)	Low-resource language 4 (Māori)	Low-resource language 5 (Mi'kmaq)	High-resource language (English)	LLM-specific sub-total
ChatGPT	398 words (Exclusively in isiZulu)	403 words (Exclusively in English)	477 words (Exclusively in English)	499 words (Exclusively in English with Māori headings & sub-headings and their translated English versions).	425 words (Exclusively in English, with Mi'kmaq headings & sub-headings and their translated English versions, and with English vocabulary items translated in Mi'kmaq).	588 words (Including six additional resources – books). NB: Lesson plan divided into time slots.	2,790
Claudia	244 words (Exclusively in isiZulu)	475 words (Exclusively in Sesotho)	303 words (Exclusively in Yoruba)	370 words (Exclusively in Māori)	233 (Exclusively in Mi'kmaq)	341 words	1,966
Copilot	246 words (Exclusively in isiZulu, excluding 8 sources together with their links at the end).	413 words (Exclusively in English, with two sources listed in the References section at the end).	282 words (Exclusively in English with time slots for each section, excluding 8 sources together with their links at the end).	439 words (Exclusively in English, with three sources listed in the references section at the end).	379 words (Exclusively in English, excluding 5 sources together with their links at the end).	359 words (with three sources listed in the references section at the end). NB: Lesson plan divided into time slots	2,118
Gemini	0 words (I'm still learning languages ...)	467 words (Sesotho response with translated English versions; some English-translated versions were nonsense translations).	382 words (In Yoruba with English-translated versions, except for the two clauses at the beginning of the lesson).	419 words (In English including titles of three additional resources and their Internet links, with Māori headings, sub-headings and portions and their translated English versions).	345 words (In English with Mi'kmaq headings and sub-headings, and Mi'kmaq concepts translated in English).	425 words (Lesson Title: Southern Multilingualism: A Symphony of Languages Grade Level: 6-8). NB: Time slots.	2,038
GroqChat	474 words (In English, with the lesson topic eight vocabulary words and the signing off translated in isiZulu)	285 words (Exclusively in English)	240 words (Exclusively in English)	378 words (In English except the lesson topic)	412 words (In English with vocabulary items (n = 5) translated in Mi'kmaq). NB: Disclaimer: Please note that I am not a native Mi'kmaq speaker ...	451 words. (NB: Assessment tasks weighted into percentages). NB: Time slots.	2,240
Perplexity	425 words (Exclusively in isiZulu)	776 words (Exclusively in Sesotho, and largely repetitive nonsense Sesotho paragraphs (except for sections/headings that do have unrepeated nonsense Sesotho paragraphs).	224 words (In English with the Introduction and vocabulary items translated into Yoruba).	301 words (In English except for the signing off, Kia kaha! (Stay strong!))	296 words (Exclusively in English)	269 words	2,291
YouChat	323 words (Mainly in isiZulu except for the last section titled, Uhwazi lwesifundo)	598 words (Mixture of Sesotho, Setswana and Sepedi) NB: Lesson sections divided into time slots).	1,122 words (Exclusively in Yoruba)	779 words (Exclusively in Māori)	349 words (Largely in English except for the translated vocabulary items (n = 5 and clauses (n = 2)).	430 words (NB: Lesson plan sections divided into time slots).	3,601
Language-specific sub-total	2,110	3,417	2,748	3,185	2,439	2,863	17,044 16,762

With regard to the five low-resource languages, Sesotho had the highest total word count of 3,417 words across the seven LLMs. It was followed by Māori that generated a total word count of 3,185 words. IsiZulu produced the lowest total word count (2,110 words) since it had a zero response from Gemini. English had a total word count of 2,883 words, which was the third-highest total word count. Of the five low-resource languages, Yoruba had the highest and lowest word counts, 1,122 words and 224 words, barring isiZulu, which generated a zero response from Gemini.

Lesedi la Naha ya Bokgoni ya Sesotho - Setšo sa Bolelo ya Kganakgang ya Naha ya Boraro
 Diphetogo: Ka letsatsi la tlase, lilemo tse fetileng tsa kgoro, baithuti ba tla nka tlhahlobo ya setšo sa bolelo ya kganakgang ya naha ya boraro le boemo ba yona ka mafatleng a Sesotho. Ba tla fumanwa le dikeletso le tsa boemo ba tsa bolelo ya kganakgang ya naha ya boraro ka mosebetsi wa AI wa boitsebiso.
 Itekanelo: 45 ntle metsotso

<p>Lesedi la Naha ya Bokgoni ya Sesotho - Setšo sa Bolelo ya Kganakgang ya Naha ya Boraro Diphetogo: Ka letsatsi la tlase, lilemo tse fetileng tsa kgoro, baithuti ba tla nka tlhahlobo ya setšo sa bolelo ya kganakgang ya naha ya boraro le boemo ba yona ka mafatleng a Sesotho. Ba tla fumanwa le dikeletso le dibeao</p>	<p>Lesedi Nabo Bokko Sesotho - Third National Pronunciation Culture Changes: On the last day, in the previous years of the course, students will take an examination of the linguistic culture of a third country and its position in the Sesotho world. They will also receive advice and reports on the status of third-party research through AI data</p>
--	--

Figure 2: A sample of YouChat's Sesotho response with a mixture of Setswana words (in red) and Sepedi words (in yellow), including nonsense constructions (in turquoise) at the top half, and the Google translation of this response at the bottom half

In terms of the responses generated in the target low-resource languages as per the common prompt, the following LLMs were able to do so, with the respective target low-resource language(s) in parentheses: ChatGPT (IsiZulu); Claudia (IsiZulu, Sesotho, Yoruba, Māori, and Mi'kmaq); Copilot (IsiZulu); Perplexity (IsiZulu and Sesotho); and YouChat (Yoruba and Māori). In this regard, four LLMs were able to generate their responses exclusively in isiZulu (see Table 1). In contrast, two LLMs generated their responses exclusively in Sesotho, Yoruba, and Māori. Only one LLM (Claudia) generated a response in Mi'kmaq (see Table 1), while another LLM (YouChat) produced a Sesotho response that had a mixture of Setswana and Sepedi¹ words (see Figure 2). This response also had a lot of nonsense constructions (see the text in turquoise in Figure 2). In addition, one LLM (Gemini) generated responses for Sesotho and Yoruba, which included their corresponding English-translated versions. For Yoruba, two clauses at the beginning of the lesson had no English translations. Moreover, one LLM (YouChat) produced an isiZulu response, but whose last section, **Ulwazi lwesifundo [Course information]**, was exclusively in English (see Tables 1 and 2). To this end, Claudia produced the most low-resource language responses (n=5), followed by Perplexity and YouChat with two low-resource language responses each (see Table 1).

Table 2: Types of responses produced by the seven LLMs for the five low-resource languages

Low-resource language-only responses	Mixed-language response (Sesotho)	Responses with corresponding English-translated versions	Response produced in isiZulu, but whose last section was exclusively in English	Responses produced English, with headings, sub-headings, and vocabulary items or concepts translated in relevant low-resource languages	Responses generated only in English	Total responses
11	1	2	1	9	11	35

Five LLMs produced English responses which had only headings, sub-headings, and vocabulary items or given concepts translated into the relevant low-resource languages. As illustrated in Table 1, these LLMs with the relevant low-resource language in parenthesis in each case included GroqChat (IsiZulu), Perplexity (Yoruba), ChatGPT (Māori), Gemini (Māori), ChatGPT (Mi'kmaq), Gemini (Mi'kmaq) (also see Figure 3), GroqChat (Mi'kmaq), and YouChat (IsiZulu and Mi'kmaq). Most of these instances were for Mi'kmaq (n=4), followed by those for isiZulu (n=2) and Māori (n=2), with GroqChat, ChatGPT, and Gemini featuring twice, each, in these instances.

Egwepanetulti'i (Useful Tip):

As large language models (AI systems that can process and generate language) become more prevalent, it's important to consider Southern multilingualism in their development. This means ensuring these models can handle and understand the unique languages and dialects spoken in our region. By advocating for Southern multilingualism in AI, we can help create technology that better reflects the diversity of our communities.

Key Terms (Mi'kmaq - English):

- Wegwalqun - South
- Kekitenuti'i - Language, speaking
- Mi'kmaqi - Mi'kmaq language

Wabakulti (Extension Activities):

- Interview an elder or community leader about their experiences with Southern multilingualism.
- Create a short video or podcast episode about the importance of Southern multilingualism.

I hope this lesson plan helps you explore the fascinating topic of Southern multilingualism in your Mi'kmaq class!

Figure 3: A sample of Gemini's Mi'kmaq's English response with English-translated headings, sub-headings, and keywords

Furthermore, the following LLMs generated the responses for the low-resource languages in parentheses in English: Copilot (Sesotho, Yoruba, Māori, and Mi'kmaq) (also see Figure 4 for the Yoruba sample response); GroqChat (Sesotho, Māori, and Yoruba); ChatGPT (Sesotho and Yoruba); and Perplexity (Māori and Mi'kmaq). Of these, Copilot produced the most English responses for four low-resource languages, followed by GroqChat (n=3). Finally, the English responses had a total word count of 2,863 words. Five of the six English responses had lesson plans with time slots for their respective lesson plan sections, with one lesson plan having its assessment tasks weighted in percentages (see Table 1).

Copilot

Yoruba Lesson: Southern Multilingualism

Objective:

- Understand the concept of Southern multilingualism and its implications.
- Appreciate the linguistic diversity in the Global South.
- Explore language resources beyond monolingual and multilingual orientations.

Teaching Aids:

- Visual aids: Charts, diagrams, or slides to illustrate key points.
- Handouts: Printed materials with relevant vocabulary and examples.
- Audio recordings: For pronunciation practice.

Lesson Structure:

1. Introduction to Southern Multilingualism (10 minutes)

- **Discussion:** What do you understand by the term “multilingualism”?
- **Explanation:** Briefly introduce the concept of Southern multilingualism. Mention that it challenges the dominance of monolingualism and highlights the linguistic richness of the Global South.
- **Tip:** Encourage students to think beyond binary language categories.

Figure 4: A sample of Copilot’s Yoruba English response from a PDF file generated by Copilot

6. Discussion

In this section, the deeper, underlying and latent structures embedded in the manifest data presented above as well as the other samples of data that could not be presented above are unpacked and discussed. Thematically, these structures are as follows: hallucinations and nonsense phrases or clauses; lack of generative multilingual capabilities; and low-resource language-only responses versus low-resource language English responses versus English language responses.

6.1 Hallucinations and Nonsense Phrases or Clauses

One classic example of a hallucinated response is the response generated by ChatGPT in isiZulu for the lesson plan of this low-resource language. The English-translated version of this sample response is nowhere near perfect, aside from having some flaws in its original isiZulu version. Its very topic is not only bizarre when juxtaposed with the concept of *Southern multilingualism*, as captured in the prompt, but it is also a nonsense topic even when seen from Google Translate’s perspective (see Figure 5). In other words, it has nothing to do with Southern multilingualism: it is a lesson plan topic for something totally different. Its last portion, **Ulimi Oluningi ngesiZulu {The Many Languages in Zulu}** demonstrates how at the point of generating a lesson plan topic in isiZulu, ChatGPT started hallucinating about **Freedom Lesson and Writing in Zulu About South Africa**, both of which have no relevance to Southern multilingualism. The same applies to the points mentioned under the lesson plan’s sections, especially under **Speech (Questions)** and **Reading Material (Examples)**, which have instances of nonsense phrases and clauses.

ChatGPT	Freedom Lesson: Writing in Zulu About South Africa - "The Many Languages in Zulu"
<p>Isifundo SeNkululeko: Ukubhala NgesiZulu Ngezwe YaseNingizimu Afrika - "Ulimi Oluningi NgesiZulu"</p> <p>Lokho Okufundelwe:</p> <p>Isizinda</p> <ul style="list-style-type: none"> Inkulumo Izinto Okufundelwe Ukufundiswa <p>Isizinda Esilandelayo</p> <ul style="list-style-type: none"> Ukufunda Isicelo SeNkululeko Indlela Yokwenza Isicelo <p>Inkulumo (Umbuzo) Ukubhala ngesiZulu kwenza kube nothwazi oluningi futhi olunamandla. Lokhu kungakwazi ukwenza ukuthi abantu bakwazi ukuxhumana nokubona izinto eziningi. Indlela esebenza ngayo kwezifundo zethu.</p> <p>Izinto Okufundelwe (Izibonelo)</p> <ol style="list-style-type: none"> Ukugcina izilimi eziningi Kulo lonke izinguquko, umuntu kumele akwazi ukugcina izilimi eziningi. Lesi simo ukubonakalisa ukuthi kuhlenganisa nolimi oluningi kulenze nje ukunikizela kakhulu. Ukuvumelanisa izilimi Indlela esebenza ngayo ukuthi ukubhala ngesiZulu kuzohlukanisa ukubhala ngesiNdebele noma ngesiXhosa. Ukuvumelanisa izilimi kuyoba nokwazi ukuzivikela ngokomzimba futhi nokwenza abantu bazigcine. 	<p>What You've Learned:</p> <p>Background</p> <ul style="list-style-type: none"> Speech Learning Materials Education <p>The Next Domain</p> <ul style="list-style-type: none"> Reading Request for Freedom How to Apply <p>Speech (Question) Writing in Zulu makes for more and more powerful information. This can enable people to communicate and see more things. How it works in our courses.</p> <p>Reading Material (Examples)</p> <ol style="list-style-type: none"> Maintaining multiple languages In all changes, one must be able to maintain multiple languages. This situation proves that it includes a lot of language just waiting to give too much. Synchronization of languages The way it works is that writing in Zulu will separate writing in Ndebele or Xhosa. Synchronizing languages will be able to defend oneself physically and to make people save themselves. Keeping on time The main time is for people to be able to learn and understand many languages. Timeliness will be able to show how good it is to maintain multiple languages.

Figure 5: A sample of an isiZulu response generated by ChatGPT (left) and its English-translated version by Google Translate (right)

Another instance of hallucination relates to YouChat's Sesotho response, which was produced with a mixture of Setswana and Sepedi words (see Figure 2). While Setswana and Sepedi do share certain words in common, and while the three languages do share common words such as *baithuti* [students/learners] and *metsofso* [minutes] as in Figure 2, Sesotho hardly employs *ga* and *go* either as single forms or as parts of words. So, mixing up forms of these languages in this way evinces a form of hallucination on the part of this LLM (YouChat) as it tended to be at sixes and sevens as to what a Sesotho orthography should be like vis-à-vis orthographies of Setswana and Sepedi. Instances of hallucination are also in the form of nonsense constructions this response has. One example of such a hallucination in this figure is the lesson plan topic itself, **Lesedi la Naha ya Bokgoni ya Sesotho - Setšo sa Bolelo ya Kganakgang ya Naha ya Boraro**, whose Google Translate version is as nonsensical and hallucinatory as its original Sesotho version. Two other examples are *Ka letsatsi la tlase*, which Google Translate translated as *On the last day*, when actually the phrase, *ka tlase*, refers to *below* or *beneath/underneath* in English, and *mafatleng a Sesotho* translated as *in the Sesotho world* by Google Translate, which is a meaningless construction as there is nothing called the Sesotho world. Both these examples, together with the first one, have nothing to do with Southern multilingualism. One more example of a nonsensical and hallucinatory Sesotho response is the one generated by Perplexity, which had a lot of repetitive nonsense Sesotho paragraphs (see Table 1 and Figure 12).

In both cases, ChatGPT's isiZulu response and YouChat's Sesotho response represent factual fabrication, which is part of factuality hallucination. In this form of hallucination, an LLM fabricates or invents non-existent facts that cannot be verified against real-world knowledge (Huang et al., 2023; see Banerjee et al., 2024). For example, the lesson plan topic of the isiZulu response, **Freedom Lesson: Writing in Zulu About South Africa - The many Languages in Zulu**, and its sub-

heading, **Synchronization of languages**, are fabricated facts that lack sense and coherence when judged against the prompt and their own flow of logic. The same can be said about YouChat's Sesotho response, which, as said earlier, is not only an admixture of Sesotho, Setswana, and Sepedi but also has a factually fabricated lesson topic and invented nonsense phrases. The fabricated lesson topic and its invented nonsense phrases lack coherence and logic and are irrelevant to the prompt.

There are further instances of hallucination and nonsense phrases and clauses in the responses produced in the other three low-resource languages. Two classic examples are the Yoruba responses generated by Claude and YouChat, even though only Claude's response and its English translations by three translation tools, Google Translate, Bing Translator, and Machine Translation.com (see Figure 6), will be used due to space constraints.

Àmi pàtáki,
Èkò Onf: Iwáðibináwó Ìpinlẹ̀ Gúúsù
Ifẹ̀síkò Ibèrè

- Awáðibináwó jẹ́ igbésí àwọn ènìyàn láti ká àwọn èdè miiran lóbi lókiki àjòbí sí ní àwọn èdè miiran tori gbogbo àwọn àjòbí ní.
- Ìpinlẹ̀ Gúúsù ní apapọ̀ àwọn ìpinlẹ̀ tó wà ní ihà gúúsù Amérikà nìbẹ̀.
- Àwọn ilú pàtáki bí New Orleans, Atlanta, Miami àti Houston ní ibi àwọn olùkòwè òpólòpò èdè wà.

Ìkórán:

- Àwọn Èyà Ìtán:
 - Ìtán àwọn àwọn iran àti igbálódí àwọn èdè yìí sí Ìpinlẹ̀ Gúúsù láti òrán rẹ̀
 - Èyí tùn fà àmi àwọn èdà òpólòpò nìbẹ̀ mọ́.
- Àwọn ètò Ìtán àti Ìwúlò:
 - Ìtán iyatò Ìyanu àwọn èdè yìí lágbádí àwọn àrùn apáádí iwọn, òrun ibiti wón wà, átẹ̀gbẹ̀be ojúdún.
 - Àwọn orisun iwúlò àti isẹ́jòba tó se pàtáki fún àwọn èdè yìí
- Èrò Ìjómràwón:
 - Àwọn ètò imúráwón àti isẹ́ àmòdí fún àwọn èdè apáádí lórílẹ̀-èdè nàà
 - Àwọn agbáyé igbésí ayé àwọn omọ ilú láti bimọ̀ èdè ibilẹ̀ àti àwọn èdè mirán

Important sign,
Today's Lesson: Southern State Crime

Initiation

- **Awadibinawo** is the life of people to learn other languages and to be famous in other languages because they are all natives.
- The Southern States are the sum of the states in the southern part of America.
- Major cities such as New Orleans, Atlanta, Miami and Houston are home to many multilingual writers.

Infection:

- Historical Features:
 - The history of these generations and the spread of these languages to the Southern State from its case
 - This also causes the sign of multiple copies.
- Historical and Useful programs:
 - The story of the difference/miracle of these languages in the midst of the hellish diseases, the heaven where they are, the year's table.
 - Sources of utility and government that are important to these languages
- Summary:
 - Development and implementation programs for indigenous languages in the country
 - The stories of the life of the citizens to learn the native language and other languages

Figure 6: A sample of Claude's Yoruba response and its English translation by Google Translate

As is evident from Figure 6, this response, together with its English translation by Google Translate, has a lot of factual fabrication as propounded by Huang et al. (2023) and Banerjee et al. (2024). For example, the lesson plan topic, the first three bullet points, and the information under **Infection** are fabricated facts that have nothing to do with the prompt and with Yoruba in Nigeria. In addition, this response consists of a mishmash of illogical and senseless ideas that have no relevance to Southern multilingualism. A case in point is the way the sections for this lesson plan have been framed and the bullet points listed under them. For instance, the statements translated as *This also causes the sign of multiple copies* and *The story of the difference/miracle of these languages in the midst of the hellish diseases, the heaven where they are, the year's table* are illogical, senseless, and hallucinatory.

An example, which is the nadir of the illogical and senseless hallucination of this response, is the reference to both America's southern states and the cities cited as *home to many multilingual writers*, an aspect which has nothing to do with Yoruba spoken in Nigeria. The same illogical and senseless hallucination is aptly captured by the English translations of the same statements offered by Bing Translator and Machine Translation.com (see Figure 7). Overall, this response and its translated English versions provide ample evidence of an LLM hallucinating in trying to respond to a prompt.

- A special sign.**
Title: The State of the South
Early Learning
- Finance is a person's responsibility to read other languages or to be popular in other languages because it is all social.
 - The Southern State is the sum of the states on America's south side there.
 - Key cities like New Orleans, Atlanta, Miami and Houston are the locations of multilingual writers.
- References:**
- Historical Features:
 - The history of the generations and modernization of these languages to the Southern State from its case
 - This also attracts the attention of many copies there.
 - History and Values programs:
 - The story of the difference/miracle of these languages through the diseases of hell size, the heaven where they are, the day of the year.
 - Sources of usefulness and governmental work necessary for these languages
 - The concept of imprisonment:
 - Assessment programs and services for hell languages in the country
 - The lives of citizens to learn local languages and other languages
- Keynote.**
Today's Lesson: Promiscuity in the Southern State
Beginning Research
- Promiscuity is the life of people to study other languages, and the tribe has other languages because they are all natives.
 - The Southern States are the sum of the states in the southern part of America.
 - Major cities like New Orleans, Atlanta, Miami and Houston are home to many multilingual writers.
- Ikoran:**
- Historical Aspects:
 - History of generations and the spread of these languages to the Southern State from its case
 - This also brings the sign of many copies there.
 - Historical and useful programs.
 - The story of the difference/miracle of these languages in the middle of the hellish diseases, the heaven where they are, the calendar of the year.
 - Sources of utility and government that are important for these languages.

Figure 7: A sample of Bing Translator's (top half) and Machine Translation.com's (bottom half) English-translated versions of Claude's Yoruba response

A measure of atonement for hallucinated responses is provided by a Māori response generated by YouChat, whose three English translations by Google Translate, Bing Translator, and Machine Translation.com are displayed in Figures 8 and 9. This response has some aspects of Southern multilingualism in words such as *multilingual* or *multilingualism* (Bing translation) and *the South* in the lesson plan topics of the three translated English versions, with the phrase, *the South* having an aura of the Global South. However, some of the aspects of these translated versions are more about how to learn Māori (e.g. language skills and language competence) using computers and Māori language applications than about what Southern multilingualism entails. Most crucially, though, there are instances of hallucination exemplified by phrases like *multilingual culture*, *intellectual technologies*, *psychological technology*, *spend cohesion*, and *over-ability utilisation of loudspeakers*. These phrases lack sense, and barring the first one, have little to do with Southern multilingualism.

Te Rārangi Whārangi: Te Whakapuakitanga o te Kaupapa - "Te Ahurea Reo Maha o te Tonga"
Whakarāpopotanga: Ko te kaupapa o tēnei whārangi whakapuakitanga, ko te "Ahurea Reo Maha o te Tonga".

Tohutohu Whaiwhakaaro mō ngā Ākonga: Hei āwhina i ngā ākonga ki te whakapau kaha i tō rātou mātauranga mo te ahurea reo maha, ka whakapau kaha i te whakapākehātanga reo me ngā whakapā ngāwari i roto i ngā hangarau ā-hinengaro, ka whakarārangahia ēnei tohutohu:

- Whakamahi i ngā whārangi reo Māori, pēnei i te YourChat, hei whakapau kaha i te mātauranga reo.
- Whakamahi i ngā rorohiko hei whakapau kaha i te mātanga reo, pēnei i te whakapau kaha i te whakapākehātanga reo.
- Whakamahi i ngā pūtōrino, ngā pukapuka, me ngā taupānga reo Māori hei whakapau kaha i te mātauranga reo.
- Whakapau kaha i te whakapākehātanga reo i ngā wāhanga reo Māori, pēnei i te whakapau kaha i te whakapākehātanga reo i roto i te rorohiko.

Whakamutunga: Kua whakakitahia e tēnei whārangi whakapuakitanga i ngā āhuatanga o te kaupapa "Ahurea Reo Maha o te Tonga". Kua whakatakotia e ia ngā whārangi whakapuakitanga, ngā whakapau kaha āwhina

Google Translate

Page List: Statement of the Project – "The Multilingual Culture of the South"
Summary: The purpose of this disclosure page is "The Multilingual Culture of the Tonga".

Counseling for Students: To help students make the most of their education for multilingual culture, make the most of language interaction and easy connections in intellectual technologies, the following tips are listed:

- Use Māori language sites, such as YourChat, to promote language education.
- Use computers to improve language skills, such as improving language skills.
- Use tapes, books, and Māori language apps to strengthen reo education.
- Use language skills in Māori language areas, such as using language skills in computers.

Conclusion: This disclosure page has summarized the aspects of the "Tonga Multilingual Culture" project. He has set up announcement pages, helpful efforts

Figure 8: A sample of Google Translate's English version of YouChat's Māori response

A Mi'kmaq response generated by Claude was translated using Rytr, a GenAI chatbot (see Figure 10). Google Translate misrecognised it as Quechua (an Indigenous language spoken in Peru) and hallucinated, too, in trying to translate it into English (see Figure 11), while the other two online translation tools could not translate it.

Bing Translator

The Page List: The Declaration of the Project - "The Many Language Culture of the South"
Summary: The subject of this declaration page is the "Multicultural Culture of the South".

Thinking Counsel for Students: To help students spend their knowledge strongly on multilingualism, spend linguistic impairment and ease of contact in psychological technology, these guidelines **can** doubt:

- Use Maori language pages, such as YourChat, to consume language education strongly.
- Use computers to spend cohesion in language competence, such as over-ability utilisation of loudspeakers.
- Use letters, books, and Maori language apps to rigorously utilise language education.
- Strongly utilises the loudspeaker in Maori parts of languages like over-utilisation of language or language **implicitly** in computer.

Conclusion: This express page **confis** the properties of

Machine Translation.com

Page List: Statement of Purpose - "The Multilingual Culture of the South"
Summary: The topic of this statement page is "The Multilingual Culture of the South".

Counseling for Students: To help students make the most of their education for multilingual culture, make the most of language communication and easy communication in intellectual technologies, the following tips are listed:

- Use Māori reo sites, such as YourChat, to make an effort in language education.
- Use computers to improve language skills, such as improving language skills.
- Use tapes, books, and Māori language apps to strengthen reo education.
- Use language skills in Māori language areas, such as using language skills in computers.

Conclusion: This disclosure page has summarized the aspects of the "Tonga Multilingual Culture" project. He has set up announcement pages, helpful efforts

Figure 9: A sample of Bing Translator's English translation of YouChat's Māori response (top half) and Machine Translation.com's version of the same sample (bottom half)

What is evident from Figure 10 is that the lesson plan topic and its related sections, together with the bullet points under each of these sections are purely related to Mi'kmaq as a language and to some of the Mi'kmaq ways of life. In this way, it can be said that Southern multilingualism is localised and seen through the prism of Mi'kmaq. Put differently, Mi'kmaq serves as a bedrock for looking at Southern multilingualism, even though this is not explicitly mentioned. Either way, the nuance of Southern multilingualism such as how Mi'kmaq is related to and is an

instance of a Southern multilingualism or how it is related to other Southern multilingualisms (Chaka, 2024c; Heugh, 2021) are lost as they are not mentioned in the response.

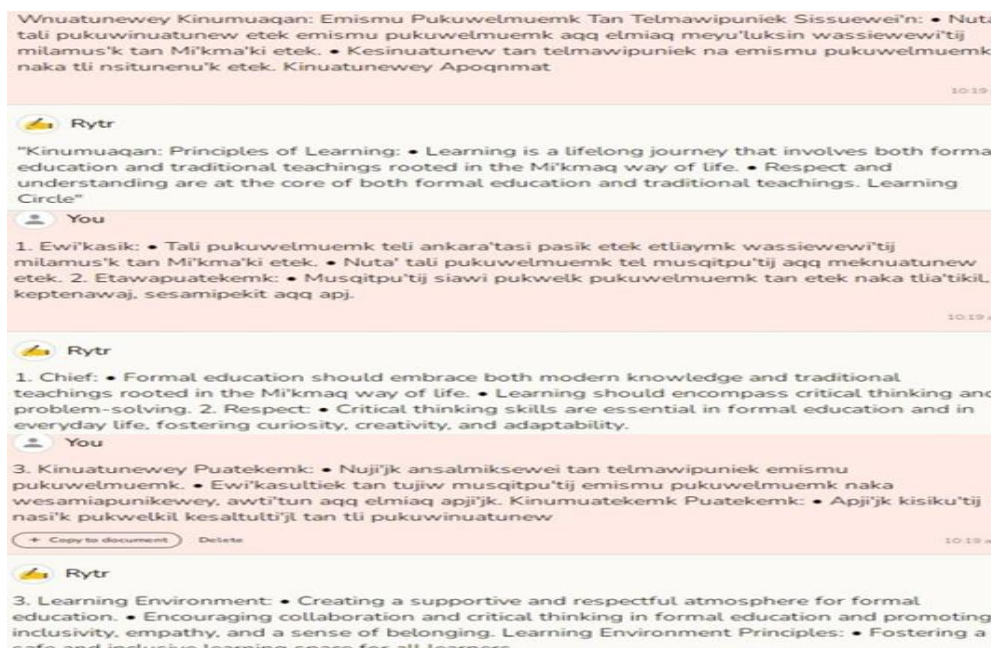


Figure 10: Rytr's English translation of Claude's Mi'kmaq response

Pertaining to the instances of hallucination related to the low-resource language responses generated by the LLMs as discussed above, the studies cited earlier flagged such hallucinations even in MLLMs and other AI machine translation tools. For example, Qi et al. (2024) argue that MLLMs have a proclivity to hallucinate, while Aharoni et al. (2024), Dale et al. (2023), and Guerreiro et al. (2023) have raised concerns about how multilingual machine translation tools tend to hallucinate. In the same vein, Chaka (2023) points out how three generative AI chatbots, ChatGPT, Chatsonic and YouChat hallucinated English responses to four English prompts related to selected areas of applied English language studies (AELS). Leffer's (2024) title, *AI chatbots will never stop hallucinating*, which she has coined for her paper and her view that LLMs tend to hallucinate even when confronted with mundane prompts aptly sum up the propensity of LLMs to hallucinate (also see Associated Federated Press [AFP], 2024; Kalai & Vempala, 2024 for another view on LLMs' hallucinations). Moreover, it is also the case that LLMs' hallucinations are related to social, cultural, and demographic biases built into LLMs' training data or to LLMs' lack of representative training data (Ferrara, 2023; Huang et al., 2023). It is for this reason that this study adopted a cautionary critical-sceptical posture regarding LLMs' performance in low-resource languages.

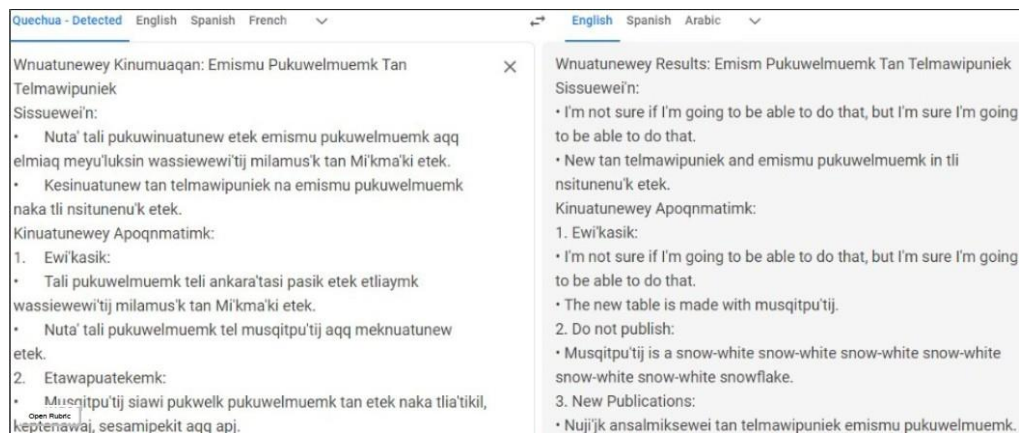


Figure 11: Google Translate's English translation of Claude's Mi'kmaq response

6.2 Lack of Generative Multilingual Capabilities

Some of the nonsense phrases or clauses and repetitive nonsense phrases or clauses including the irrelevant and flawed responses generated by the LLMs as demonstrated in the preceding section, reflect the lack of generative multilingual capabilities these LLMs have in low-resource languages. This is a concern raised by Qin et al. (2024), arguing that the current MLLMs, and by analogy, the current LLMs, tend to display poor performance in low-resource languages. But, in the current study, some of the LLMs did not possess any relevant knowledge of some of the investigated low-resource languages. The YouChat's Sesotho response (see Figure 2), the ChatGPT's isiZulu response (see Figure 5), Claude's Yoruba response (see Figures 6 and 7), and some aspects of the YouChat's Māori response (see Figures 8 and 9) are the classic examples. Concerning YouChat's Sesotho response, the resultant three-language mixed response demonstrates this LLM's dearth or absence of generative multilingual capabilities in distinguishing between a purely Sesotho response and standalone Setswana and Sepedi responses.

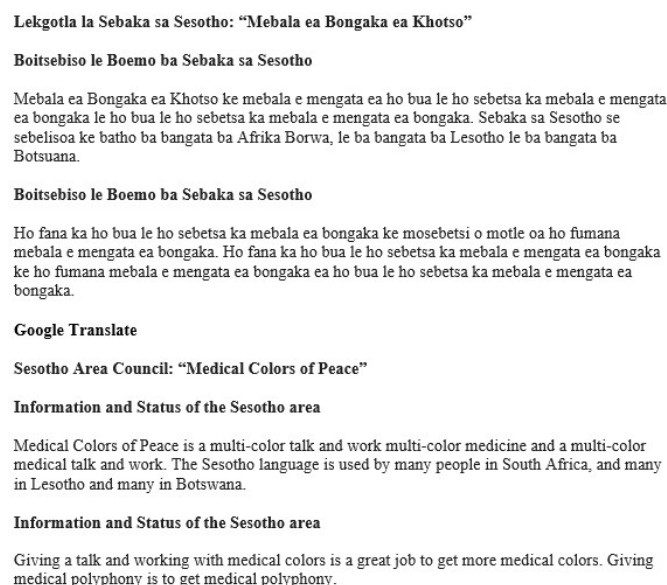


Figure 12: A sample of Perplexity's Sesotho response (top half) and its Google Translate English version (bottom half)

Allied with this Sesotho response is another Sesotho response generated by Perplexity (see Figure 12). This response, which Microsoft Word's Set Proofing Language feature detects as Sesotho² (the South African version) when actually it is a Lesotho version of Sesotho) displays numerous instances of hallucination as exemplified by the lesson plan topic and its two sections. However, as this aspect has been discussed in the preceding section, the focus here is on repetitive nonsense phrases or clauses. For example, the first clause under the first section and the two clauses under the second section are nonsense clauses. Moreover, the second section, together with the information under it, is repeated verbatim in the other 10 sections of this lesson that Perplexity generated. This means that the 776 words that this response has (see Table 1) are 776 words of repetitive, nonsensical and meaningless information. This phenomenon of spewing out repetitive, nonsensical and meaningless sentences serves as one of the indicators of this LLM's lack of generative multilingual capabilities in Sesotho as a low-resource language.

As pointed out above, the habit of churning out nonsense phrases or clauses is evident in Claude's Yoruba response as well (see Figures 6 and 7). From the lesson plan topic to its section headings and their attendant bullet points, this response is replete with nonsensical and meaningless information. For instance, there is a huge disjuncture in the lesson topic as captured in its three English translation versions. This is apart from the bizarreness and senselessness of these translated versions. The same bizarreness and senselessness of ideas is manifest in the two sections of the lesson and their related bullet points. This strange and senseless construction of ideas, which tends to upset even non-Yoruba speakers, is symptomatic of the lack of generative multilingual capabilities that LLMs such as Claude have in a low-resource language like Yoruba.

As is the case with Perplexity's Sesotho response which had 776 words of repetitive, nonsensical, and meaningless information, Claude's Yoruba response with a word count of 1,122 words (see Table 1) is a response with 1,122 words of bizarre and senseless ideas. This means that the number of words this response has does not reflect any quality³ response. Ironically, this generative multilingual deficiency tends to play itself out even in machine translation tools such as Google Translate, Bing Translator, and Machine Translation.com as depicted by Figures 6 and 7. Furthermore, pockets of such a generative multilingual deficiency are evident in YouChat's Māori response in relation to phrases such as *intellectual technologies*, *psychological technology*, and *over-ability utilisation of loudspeakers* (see Figures 8 and 9). It is only a machine, and not a human being, that can thoughtlessly churn out a phrase like the last one. Most crucially, the lack of generative multilingual capabilities of the LLMs analysed in this study is further reflected by how some of them generated their responses exclusively in English when the prompt instructed them to do so in each specified low-resource language (see Tables 1 and 2). Huang et al. (2023) call the inability of an LLM to follow instructions as instruction inconsistency.

As highlighted earlier, the cross-linguistic or multilingual capabilities of LLMs are critical as LLMs, at times, tend to lack generative multilingual capabilities even in high-resource languages. One of the results of this generative multilingual deficiency is the generation of incomprehensible outputs. Even in cases of low-resource languages where LLMs have been fine-tuned with innovative tools such as adaMLLM (Lankford et al., 2023) and with innovative methods like a linguistically-diverse prompting (LDP) (Nguyen et al., 2023) and a cross-lingual-thought prompting (XLT) (Huang et al., 2023), such tools and methods do not completely eliminate the generative multilingual deficiencies LLMs have. Based on the instances of nonsensical and meaningless responses discussed above, it seems that LLMs' generative multilingual incapability is more intense and concerning in low-resource languages. This is what the next section focuses on.

6.3 Low-resource Language-Only Responses Versus Low-Resource Language English Responses Versus English Language Responses

When the low-resource language responses produced in the respective low-resource languages are compared with the English responses produced for these languages, the latter set of responses (the low-resource language English responses) is, collectively, better than their counterparts in terms of quality, relevance, depth, detail, and nuance. For instance, the ChatGPT's Sesotho response, which is titled **Southern Multilingualism**, states as part of its lesson objective that students should be able to understand the concept of Southern multilingualism, identify its *characteristics*, and *appreciate its importance in the context of the Southern African region*. In addition, it mentions in its introduction things such as a *brief overview of multilingualism* and the *importance of language diversity in Southern Africa*. This is a far cry from the two Sesotho responses generated by YouChat (see Figure 2) and Perplexity (see Figure 12) discussed earlier.

The same can be said about the Copilot's Yoruba response generated in English (see Figure 4), which is qualitatively better than Claude's Yoruba response counterpart produced exclusively in Yoruba (see Figures 6 and 7). To contextualize its better quality, its lesson plan titled, **Southern Multilingualism**, has these two aspects, *appreciate the linguistic diversity in the Global South* and *explore language resources beyond monolingual and multilingual orientations*, as part of its lesson plan objective. In its lesson plan introduction, under the sub-section **Explanation**, it refers to *Mention that it (Southern multilingualism) challenges the dominance of monolingualism and highlights the linguistic richness of the Global South*. Moreover, in the lesson plan portions not displayed in Figure 6, this response has a section titled, **Defying Monolingual Norms**, and has a task, *Discuss how AI-powered large language models handle multilingual data*, under a section titled, **AI and Multilingualism**. All of these aspects are some of the essential elements of Southern multilingualism – there is no gainsaying that one of the focal points of Southern multilingualism, a concept associated with the Global South, is to problematize and resist the hegemony of monolingualism (Chaka, 2024c; Heugh, 2021). Most significantly, the task, *Discuss how AI-powered large language models handle multilingual data*, speaks to one of the aspects mentioned in the prompt: *Also, provide a useful tip for students about Southern multilingualism in the context of AI-powered large language models* (see Figure 1). This response is in stark contrast

not only to Claude's Yoruba response counterpart but also to the other responses produced in the respective low-resource languages discussed above.

The points highlighted above about Claude's Yoruba response apply in varying degrees to the Mi'kmaq response generated exclusively in English by Perplexity (see Table 1). Titled, **Mi'kmaq Language Class Lesson: Southern Multilingualism**, the response says the following in its introduction: *Southern multilingualism refers to the phenomenon of individuals or communities speaking multiple languages in the southern regions. This is a rich and diverse aspect of language and culture that we will delve into.* Towards the end, under its **Useful Tip for Students** section, it states, *In the context of AI-powered large language models, understanding Southern multilingualism can provide valuable insights into how these models can better support and represent diverse languages and dialects.* This statement deals with the last aspect of the prompt as discussed in the preceding paragraph (also see Figure 1). Coincidentally, what is captured by this statement resonates with one of the things the current study attempted to investigate even though its Southern multilingualism is subsumed under low-resource languages. In what has become a standard practice for most LLMs, no sources have been cited for the statements used in this response, including its definition of Southern multilingualism. And, its phrasal verb, *delve into*, which it has used in the introduction, ranks number 11 among the top 100 most commonly used AI words (AI Phrase Finder, 2024; see Gray, 2024). Notwithstanding, this response is qualitatively better than Claude's Mi'kmaq-only response counterpart (see Figure 10). In addition, like Claude's Yoruba-only response above, it fares better than the other low-resource language responses dealt with thus far.

Barring one response, the English responses had time slots allocated to their lesson plans (see Table 1). One of these, the GroqChat's response, even had its assessment tasks weighted into percentages as follows: participation in class discussions and activities (20%); written essay (40%); class participation and engagement (20%); and quality of skit or dialogue (20%). Even though some of the low-resource language responses, especially those produced in English by these LLMs, had lesson plan time slots, none of them had their assessment tasks allocated percentage weightings. For example, in a like-like comparison, GroqChat produced all its low-resource language responses in English, except for minor variations here and there (see Table 1). Of these responses, none had percentage weightings allotted to its assessment tasks. Except for its isiZulu response, which had time slots for its various sections, and excluding its Māori response, which had a global 60 minutes mentioned under **Duration** at the beginning of the lesson plan, the other three low-resource languages (Sesotho, Yoruba, and Mi'kmaq) had no time slots for their lesson plans. In addition, the English lesson plan and the isiZulu and Māori responses had grade or educational levels mentioned, while the other three low-resource languages had no grade or educational levels specified.

When the English responses are compared with the two sets of responses discussed above, their better quality is noticeable: they are qualitatively better concerning relevance, depth, detail, and nuance. Three responses generated by ChatGPT, Claude, and Copilot (see Table 1) are used for illustrative purposes. For

example, titled, **Southern Multilingualism**, the ChatGPT's response mentions new aspects of Southern multilingualism, which the two sets of responses dealt with above did not touch on. These include the following statements: *Influence of colonization and migration on Southern multilingualism*; *Impact of indigenous languages and African languages on Southern multilingualism* and *Reflect on the implications of Southern multilingualism in the context of AI-powered large language models*. This response ends with the following useful tip for students:

When interacting with AI-powered large language models like OpenAI's GPT-3, it's important to be aware of the limitations and biases that can arise due to the dominance of certain languages and dialects in the training data. Southern multilingualism offers a rich diversity of languages and cultures that can enhance the development and application of AI technologies.

All of these aspects are qualitatively different from the two sets of responses discussed earlier in terms of both their contextual relevance and their nuanced approach to Southern multilingualism. For its part, Claude's response introduces new pieces of information to Southern multilingualism such as an *Overview of the multilingual landscape in the southern regions (Africa, Asia, South America)* and *When using AI-powered large language models like Claude, be aware that these models are trained on vast amounts of text data, which may include biases and inaccuracies regarding linguistic diversity and minority languages*. These pieces of information add a new dimension to Southern multilingualism which is missing in the other responses. Similarly, Copilot's response introduces a different tack to the Southern multilingualism lesson plan with aspects like *Emphasising that it refers to the use of multiple languages in the southern regions of sub-Saharan Africa* and *Introduce the idea of decolonial approaches to multilingualism*. While these two aspects are equally relevant to Southern multilingualism, the latter adds decoloniality, which is one of the characteristic features of Southern multilingualism (see Chaka, 2024c). The better quality of English responses vis-à-vis the other sets of responses (the low-resource language-only responses and the low-resource language English responses) resonates with Lorandi and Belz's (2023) observation that LLMs such as ChatGPT excel in English prompts and tasks as this reflects how high-resource languages like English dominates LLMs' training datasets. This is the point taken a step further by Navigli et al. (2023) who argue that the selection and creation of training data for the current LLMs is biased towards high-resource languages and ignores low-resource languages. They also contend that this training data bias and imbalance manifests itself in richer quality and quantity of responses for high-resource languages as opposed to responses for low-resource languages, which are often poorer in quality. The current study has not only demonstrated how the low-resource language-only responses were poor in quality and relevance as compared to both the low-resource language English responses and the English language responses, but it has also illustrated how such responses had the high quantity (the high word counts) of hallucinated, nonsensical and meaningless information.

7. Implications

This study has implications for the use of LLMs to generate responses from prompts that are exclusively in low-resource languages such as the ones which the study investigated. The first implication is that LLMs hallucinate responses in low-resource languages. This means that if academics and educators of low-resource languages think that currently available LLMs will be of any value to them, they must think twice. The second implication relates to nonsense phrases or clauses. The LLMs investigated in this study generated nonsensical and senseless responses in the five low-resource languages on which the study focused. Here the picture is not rosy or promising at all, especially if the two Sesotho responses (Figures 2 and 12) and the Yoruba response (Figures 6 and 7) and their respective translated English versions are anything to go by. If this is the case, then, low-resource language academics and educators should expect nonsensical and senseless responses from the currently available LLMs, unless something radically changes.

The third implication is that the currently available LLMs significantly lack generative multilingual capabilities in low-resource languages. In fact, they seem to have better monolingual generative capabilities in English as a high-resource language. This aspect is exemplified by the low-resource language responses produced exclusively in English (n=11) by four of the seven LLMs investigated (e.g. GroqChat, ChatGPT, Copilot, and Perplexity) (see Table 1). The fourth and last import is that the designers and data trainers of LLMs need to know that the currently available LLMs are heavily biased towards a high-resource language such as English, while they grossly marginalise low-resource languages like the ones investigated in this study.

8. Conclusion

This study had three focal points stated in its three research questions mentioned earlier. Overall, the study found that the seven LLMs have a significant lack of generative multilingual capabilities in the five low-resource languages investigated by the study. As a result of this generative multilingual deficiency, the seven LLMs hallucinated when they were prompted to generate responses exclusively in these five low-resource languages. The hallucinations were more profound and pervasive in isiZulu, Sesotho, and Yoruba responses. Allied to hallucinated responses is the fact that these LLMs spewed out nonsensical, meaningless, and irrelevant responses in their low-resource language outputs. Such nonsensical, meaningless, and irrelevant responses were more pronounced and telling in the YouChat's Sesotho response, the Perplexity's Sesotho response, the ChatGPT's isiZulu response, and Claude's Yoruba response, and in some aspects of the YouChat's Māori response.

Moreover, the study discovered that the English language-specific responses or the responses generated by all seven LLMs in English as a language were far better in quality, relevance, depth, detail, and nuance than the low-resource language-only responses and the English responses generated for the five low-resource languages. This aspect highlights how the seven, currently available LLMs investigated in this study are heavily skewed towards a high-resource language

such as English in their training data, while they grossly under-represent low-resource languages. This calls for more representative, cross-lingual and more inclusive training datasets for these LLMs than the ones they currently have, which are heavily biased towards high-resource languages like English. This is what future research needs to consider. Finally, one of the shortcomings of the current study is that it focused only on five low-resource languages, two of which were from the same region.

Notes

1. Sepedi is also known as Northern Sotho.
2. Here, it is worth mentioning that Google South Africa's search engine has, under its African language offerings, a Lesotho Sesotho version, and not a South African Sesotho version.
3. Quality response refers to an effective, reliable, and accurate LLM response as measured against its prompt.

NB: The ethical clearance certificate for the current study was granted by the College Research Ethics Committee with the following registration and reference numbers, respectively: NHREC Registration #: Rec-240816-052; and CREC Reference #: 35288353_CREC_CHS_2024.

9. References

- Aboze, B. J. (2023). *Risks of large language models: A comprehensive guide*. <https://www.deepchecks.com/risks-of-large-language-models/>
- Associated Federated Press. (2024). *ChatGPT faces Austria complaint over 'uncorrectable errors'*. <https://www.taipeitimes.com/News/biz/archives/2024/04/30/2003817178>
- Aharoni, R., Narayan, S., Maynez, J., Herzig, J., Clark, E., & Lapata, M. (2024). *Multilingual summarization with factual consistency evaluation*. <https://arxiv.org/pdf/2212.10622.pdf>
- AIContentfy Team. (2023). *Evaluating the effectiveness of AI detectors: Case studies and metrics*. <https://aicontentfy.com/en/blog/evaluating-of-ai-detectors-case-studies-and-metrics>
- AI Phrase Finder. (2024). *The 100 most common AI words*. <https://aiphrasefinder.com/common-ai-words/>
- Akula, B., Andrews, P., Ayan, N. F., Barrault, L., Bhosale, S., Costa-jussa, M. R., Cross, J. ... Youngblood, A. (2024). *200 languages within a single AI model: A breakthrough in high-quality machine translation*. <https://ai.meta.com/blog/nllb-200-high-quality-machine-translation/>
- Banerjee, S., Agarwal, A., & Singla, S. (2024). *LLMs will always hallucinate, and we need to live with this*. <https://arxiv.org/pdf/2409.05746>
- Chaka, C. (2022). Digital marginalization, data marginalization, and algorithmic exclusions: a critical southern decolonial approach to datafication, algorithms, and digital citizenship from the Souths. *Journal of e-Learning and Knowledge Society*, 18(3), 83-95. <https://doi.org/10.20368/1971-8829/1135678>
- Chaka, C. (2023). Generative AI chatbots - ChatGPT versus YouChat versus Chatsonic: Use cases of selected areas of applied English language studies. *International Journal of Learning, Teaching and Educational Research*, 22(6), 1-19. <https://doi.org/10.26803/ijlter.22.6.1>

- Chaka, C. (2024a). Reviewing the performance of AI detection tools in differentiating between AI-generated and human-written texts: A literature and integrative hybrid review. *Journal of Applied Learning & Teaching*, 7(1), 1-12. <https://doi.org/10.37074/jalt.2024.7.1.14>
- Chaka, C. (2024b). Accuracy pecking order – How 30 AI detectors stack up in detecting generative artificial intelligence content in university English L1 and English L2 student essays. *Journal of Applied Learning & Teaching*, 7(1), 1-13. <https://doi.org/10.37074/jalt.2024.7.1.33>
- Chaka, C. (2024c). Multilingualism, translanguaging, diversity, equity, social justice, and activism: A tenuous nexus and misrepresentations? *International Journal of Language Studies*, 18(1), 7-28. <https://doi.org/10.5281/zenodo.10468173>
- Captain Words. (2024). *Testing AI detection tools – Our methodology*. <https://captainwords.com/ai-detection-tools-test-methodology/>
- Cave, S., & Dihal, K. (2020). The Whiteness of AI. *Philosophy & Technology*, 33, 685-703. <https://doi.org/10.1007/s13347-020-00415-6>
- Dale, D., Voita, E., Lam, J., Hansanti, P., Ropers, C., Kalbassi, E., Gao, C., Barrault, L., & Costa-jussà, M. R. (2023). *HalOmi: A manually annotated benchmark for multilingual hallucination and omission detection in machine translation*. <https://arxiv.org/pdf/2305.11746.pdf>
- Delve, H. L., & Limpaecher, A. (2022). *Qualitative content analysis: Manifest content analysis vs. latent content analysis*. <https://delvetool.com/blog/manifest-content-analysis-latent-content-analysis>
- Ferrara, E. (2023). *Should ChatGPT be biased? Challenges and risks of bias in large language models*. <https://arxiv.org/pdf/2304.03738.pdf>
- Gray, A. (2024). *ChatGPT “contamination”: Estimating the prevalence of LLMs in the scholarly literature*. <https://arxiv.org/pdf/2403.16887>
- Guerreiro, N. M., Alves, D. M., Waldendorf, J., Haddow, B., Birch, A., Colombo, P., & Martins, A. F. T. (2023). *Hallucinations in large multilingual translation models*. <https://arxiv.org/abs/2303.16104>
- Hadi, M. U., al tashi, q., Qureshi, R., Shah, A., muneer, a., Irfan, M., Zafar, A., Shaikh, M. B., Akhtar, N., Wu, J., & Mirjalili, S. (2023). *Large language models: A comprehensive survey of its applications, challenges, limitations, and future prospects*. https://d197for5662m48.cloudfront.net/documents/publicationstatus/181139/preprint_pdf/edf41a1f2a93aadb235a3c3aff2dcf08.pdf
- Heugh, K. A. (2021). Southern multilingualisms, translanguaging and transknowledging in inclusive and sustainable education. In P. Harding-Esch & H. Coleman (Eds.), *Language and the sustainable development goals* (pp. 37-47). British Council.
- Huang, H., Tang, T., Zhang, D., Zhao, W. X., Song, T. Xia, Y. & Wel, F. (2023). *Not all languages are created equal in LLMs: Improving multilingual capability by cross-lingual-thought prompting*. <https://arxiv.org/abs/2305.07004>
- Huang, L., Yu, W., Ma, W., Zhong, W., Feng, Z., Wang, H., Chen, Q., Peng, W., Feng, X., Qin, B., & Liu, T. (2023). *A survey on hallucination in large language models: Principles, taxonomy, challenges, and open questions*. <https://arxiv.org/pdf/2311.05232>
- IBM. (2024). *What are AI hallucinations?* <https://www.ibm.com/topics/ai-hallucinations>
- Kalai, A. T., & Vempala, S. S. (2024, revised version). *Calibrated language models must hallucinate*. <https://arxiv.org/abs/2311.14648>
- Kleinheksel, A. J., Rockich-Winston, N., Tawfik, H., & Wyatt, T. R. (2020). Demystifying content analysis. *American Journal of Pharmaceutical Education*, 84(1), 127-137.
- Lankford, S., Afli, H., & Way, A. (2023). adaptMLLM: Fine-tuning multilingual language models on low-resource languages with integrated LLM playgrounds. *Information*, 14, 638. <https://doi.org/10.3390/info14120638>

- Lee, N. T., Resnick, P., & Barton, G. (2019). *Algorithmic bias detection and mitigation: Best practices and policies to reduce consumer harms*. <https://www.brookings.edu/research/algorithmicbias-detection-and-mitigation-best-practices-and-policies-to-reduce-consumer-harms/>
- Leffer, L. (2024). *AI chatbots will never stop hallucinating*. <https://www.scientificamerican.com/article/chatbot-hallucinations-inevitable/>
- Lin, C., Gao, Y., Ta, N., Li, K., & Fu, H. (2023). Trapped in the search box: An examination of algorithmic bias in search engine autocomplete predictions. *Telematics and Informatics*, 85, 102068. <https://doi.org/10.1016/j.tele.2023.102068>
- Lorandi, M., & Belz, A. (2023). *Data-to-text generation for severely under-resourced languages with GPT-3.5: A bit of help needed from Google Translate*. <https://aclanthology.org/2023.mmnlg-1.9.pdf>
- Navigli, R., Conia, S., & Ross, B. (2023). Biases in large language models: Origins, inventory, and discussion. *ACM Journal of Data and Information Quality*, 15(2), 1–21. <https://doi.org/10.1145/3597307>
- Nguyen, X. P., Aljunied, S. M., Joty, S., & Bing, L. (2023). *Democratizing LLMs for low-resource languages by leveraging their English dominant abilities with linguistically-diverse prompts*. <https://arxiv.org/abs/2306.11372>
- Patil, R., & Gudivada, V. (2024). A review of current trends, techniques, and challenges in large language models (LLMs). *Applied Sciences*, 14, 2074. <https://doi.org/10.3390/app14052074>
- Perkins, M. (2023). Academic integrity considerations of AI large language models in the post-pandemic era: ChatGPT and beyond. *Journal of University Teaching & Learning Practice*, 20(2), 7. <http://dx.doi.org/10.53761/1.20.02.07>
- Popenici, S. (2023). The critique of AI as a foundation for judicious use in higher education. *Journal of Applied Learning and Teaching*, 6(2), 378–384. <https://doi.org/10.37074/jalt.2023.6.2.4>
- Qin, L., Chen, Q., Zhou, Y., Chen, Z., Li, Y., Liao, L., Li, M., Che, W., & Yu, P. S. (2024). *Multilingual large language model: A survey of resources, taxonomy and frontiers*. <https://arxiv.org/abs/2404.04925>
- Rudolph, J., Tan, S., & Tan, S. (2023). ChatGPT: Bullshit spewer or the end of traditional assessments in higher education?. *Journal of Applied Learning and Teaching*, 6(1), 342-363. <https://doi.org/10.37074/jalt.2023.6.1.9>
- Rudolph, J., Ismail, M. F., & Popenici, S. (2024). Higher education's generative artificial intelligence paradox: The meaning of chatbot mania. *Journal of University Teaching and Learning Practice*, 21(6). <https://doi.org/10.53761/pzd17z29>
- Snyder, A. (2023). *AI's language gap*. <https://studies.www.axios.com/2023/09/08/ai-language-gap-chatgpt>
- Vashee, K. (2023). *Making generative AI effectively multilingual at scale*. <https://blog.modernmt.com/making-generative-ai-multilingual-at-scale/>
- Wu, J., Yang, S., Zhan, R., Yuan, Y., Wong, D. F., & Chao, L. S. (2023). *A survey on LLM-generated text detection: Necessity, methods, and future directions*. <https://arxiv.org/pdf/2310.14724.pdf>