



King's Research Portal

DOI:

[10.1177/01634437221104705](https://doi.org/10.1177/01634437221104705)

Document Version

Publisher's PDF, also known as Version of record

[Link to publication record in King's Research Portal](#)

Citation for published version (APA):

Chonka, P., Diepeveen, S., & Haile, Y. (2022). Algorithmic power and African indigenous languages: search engine autocomplete and the global, multilingual Internet. *Media, Culture and Society*.
<https://doi.org/10.1177/01634437221104705>

Citing this paper

Please note that where the full-text provided on King's Research Portal is the Author Accepted Manuscript or Post-Print version this may differ from the final Published version. If citing, it is advised that you check and use the publisher's definitive version for pagination, volume/issue, and date of publication details. And where the final published version is provided on the Research Portal, if citing you are again advised to check the publisher's website for any subsequent corrections.

General rights

Copyright and moral rights for the publications made accessible in the Research Portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognize and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the Research Portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the Research Portal

Take down policy

If you believe that this document breaches copyright please contact librarypure@kcl.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.

Algorithmic power and African indigenous languages: search engine autocomplete and the global multilingual Internet

Media, Culture & Society

1–20

© The Author(s) 2022



Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/01634437221104705

journals.sagepub.com/home/mcs**Peter Chonka** 

King's College London, UK

Stephanie Diepeveen

University of Cambridge, UK

Yidnekachew Haile

Royal Holloway University of London, UK

Abstract

Predictive language technologies – such as Google Search's Autocomplete – constitute forms of algorithmic power that reflect and compound global power imbalances between Western technology companies and multilingual Internet users in the global South. Increasing attention is being paid to predictive language technologies and their impacts on individual users and public discourse. However, there is a lack of scholarship on how such technologies interact with African languages. Addressing this gap, the article presents data from experimentation with autocomplete predictions/suggestions for gendered or politicised keywords in Amharic, Kiswahili and Somali. It demonstrates that autocomplete functions for these languages and how users may be exposed to harmful content due to an apparent lack of filtering of problematic 'predictions'. Drawing on debates on algorithmic power and digital colonialism, the article demonstrates that global power imbalances manifest here not through a lack of online African indigenous language content, but rather in regard to the moderation of content across diverse cultural and linguistic contexts. This raises dilemmas for actors invested in the multilingual Internet between risks of digital surveillance and effective

Corresponding author:

Peter Chonka, King's College London, S3.13, The Strand, London WC2R 2LS, UK.

Email: peter.chonka@kcl.ac.uk

platform oversight, which could prevent algorithmic harms to users engaging with platforms in a myriad of languages and diverse socio-cultural and political environments.

Keywords

Africa, algorithmic power, autocomplete, digital colonialism, digital media, language, search

Introduction

While search engines play a central role in curating and presenting users with information options from a seemingly limitless expanse of online data, algorithms like Google's are hardly neutral windows onto what is simply most popular 'out there' online (Noble, 2018). Search results are heavily influenced by commercial factors, including commissioned advertising and live markets for advertisers that commodify keywords and allow companies to bid for their association with particular types of searches. Commercial searches can reinforce racist and misogynist stereotypes, and present users with other forms of problematic content (Noble, 2013, 2018; Thornton, 2017). These concerns inform broader discussions around the moderation of online content and the role of human and algorithmic labour in these interventions (Gillespie, 2018). As debates continue about how online platform infrastructures allow or encourage the circulation of misleading or hateful content (Allcott and Gentzkow, 2017; Gray et al., 2020), pressure on companies like Google has, at times, encouraged their tweaking of algorithms to remove problematic results or specific 'autocomplete' search 'predictions'.

This research and public scrutiny has hitherto focussed overwhelmingly on Western, often English-language, online contexts. However, because Google/Alphabet dominates the global mobile operating system and browser markets (GlobalStats, 2022a, 2022b), its products – such as Google Search – are ubiquitous and used in hundreds of languages globally. This article offers a new perspective on how autocomplete algorithms interact with social and political relations within non-Western and non-English language settings, demonstrating how context- and language-specific analyses are required to grasp both the workings of search engine algorithms and their implications for socio-political discourse in different parts of the 'global South'. Here we draw on work challenging assumptions about 'universal features of communication with little cultural variation' (Pohjonen and Udupa, 2017: 1174) that often presuppose the universality of Western experiences. Google is a Western company that has effectively 'imposed itself' as a 'universal arbiter' (Asante, quoted in Anderson, 2012: 764) in processes of digital communication on the African continent. As such, we build on African-focussed critiques of digital colonialism (Kwet, 2019) to consider the implications of how its algorithmic power interacts with diverse African languages. This occurs in ways that are unanticipated, but also potentially intersect with the technology company's profit motives on a continent that is increasingly viewed as a primary market for future expansion (Oyedemi, 2021).

The article presents and analyses experimental data on how search engine autocomplete predictions respond to a sample of keywords related to political and social domains

in three African languages with distinct scripts: Amharic, Kiswahili and Somali. We demonstrate how Google's autocomplete algorithm does not appear to filter out 'predictions' for Somali and Amharic keywords that we would consider to be problematic given their (for example) misogynistic or sexualised character. Broader evidence suggests that some of these predictions would be removed from autocomplete predictions for search terms in English through combinations of algorithmic and human monitoring. The same keywords in Kiswahili tended to generate less similarly questionable predictions in our experiments. On one level, these findings raise questions on how these phenomena are being monitored by tech companies, political actors or concerned citizens in multi-lingual African contexts – that is, whether the companies responsible for such platforms have the capacity or motivation to undertake algorithmic tweaking across the full spectrum of languages that global users engage with them.

Regardless of linguistic variations in how search engine autocomplete mediates the relationship between users' prior searches and broader online behaviour, we argue that these technologies exert some level of agency in shaping social and political relations. This argument hinges on a problematisation of Google's claim that autocomplete works to provide 'predictions' as opposed to 'suggestions' for users' information seeking behaviours. Drawing from Miller and Record's (2017) analysis of the epistemological power of autocomplete, we argue that this prediction/suggestion distinction is untenable and highlights potential wider impacts of 'algorithmic power' (Bucher, 2018) on spaces for communication in multilingual and transnationally-connected African digital publics. The problems associated with the inscrutable 'black box' nature of algorithmic functions are well documented (O'Neil, 2017; Pasquale, 2016). We argue that concerns around transparency and accountability for the results of algorithmic operations – such as autocomplete suggestions on a search engine, or increasingly, 'predictive' text on other digital platforms – are heightened for 'smaller' global languages which interact with algorithms in even more opaque ways. A specific example illustrates this point. In our initial testing of Google search autocomplete for the gendered Somali keyword 'girl' (*gabadh*), the first suggestion for related search keywords was 'naked' (*qaawan*). At the time of writing this paper, however, that suggested keyword was no longer returned (See Figure 1).

It is only a secondary objective of this article to highlight such specific instances of problematic algorithmic predictions or suggestions. Our primary aim is to think through what these examples of the operation of Western-developed technologies can tell us about algorithmic power on the global multilingual Internet, and how this intersects with wider questions about 'digital', 'data' or 'techno colonialism' (Couldry and Mejias, 2019; Kwet, 2019; Madianou, 2019). Practically and specifically, our findings inform our call for greater transparency around how companies, whose platforms are increasingly ubiquitous worldwide, manage the global reach of this power. In the example above, we do not (and currently cannot) know whether the disappearance of this sexualising Somali term was the result of manual or automated moderation by Google, or changes in users' search behaviour. Accessible information on what (if any) multilingual moderation is being undertaken is vital for global language communities seeking to understand the impact of such algorithms on public discourse and advocate for platform accountability.

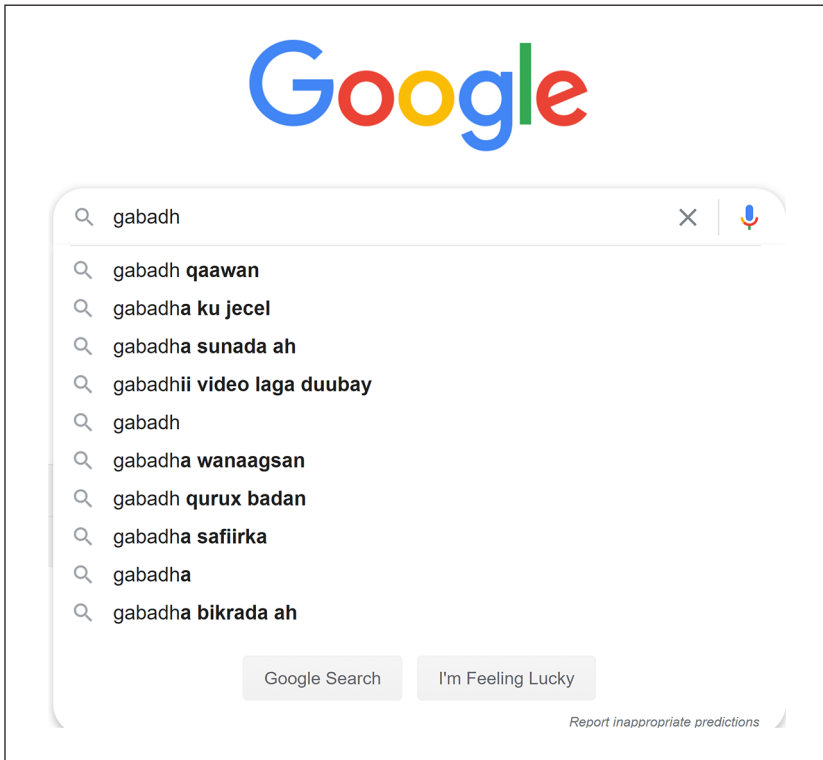


Figure 1. Screenshot of Google autocomplete for Somali search term 'gabadh' (girl), 17 September 2020.

We develop our argument in the following steps. First, we explain what Google autocomplete is and situate our comparative multilingual experimentation in relation to existing scholarship on this predictive technology. We contend that autocomplete has not been adequately problematised in scholarship on algorithmic power and bias, particularly concerning global multilingual platform use. We then explain our method for gathering and comparing data on autocomplete and East African languages with keywords relating to gender and national politics. We present results in relation to their linguistic, political and communicative contexts. Our analysis highlights sexualised and misogynistic autocomplete for local language search keywords for 'girl'/'woman' in the Somali and Ethiopian (Amharic) contexts. We then discuss a more ambiguously problematic example relating to autocomplete results of 'clan'-related keywords associated with political figures in Somalia. We consider the complexities of the local discursive landscape and how this algorithmic feature of the digital public is informed by – and contrasts with – norms of political debate in the region. As location-based personalisation of autocomplete results appears to be minimal (and users may be located outside of the region), this draws attention to the possible influence of transnational diasporic digital activity on public spheres of debate in these languages.

Ultimately, the article highlights how global linguistic diversity conditions forms of algorithmic power. It questions the extent to which the use of ‘smaller’ languages insulates users from practices of data extractivism, while potentially exposing them to other forms of digital harm in contexts where content monitoring/moderation is limited. Our multilingual autocomplete case study brings to the fore potential tradeoffs between promoting linguistic diversity online and exposing global users to increased digital surveillance. We conclude by posing questions that require attention within specific socio-linguistic contexts about the desirability of (and responsibility for) policing ‘inappropriate’ search predictions in African languages.

Autocomplete, algorithmic power and global linguistic diversity

In operation on major search engines since around 2004, autocomplete tools dynamically predict what a user is looking for as they type a query into a search engine (Miller and Record, 2017). Our empirical analysis focuses on Google Search, the globally predominant (bar China) and most used search engine across the African continent (GlobalStats, 2021). Autocomplete functions are not limited to Google Search, featuring on other applications within the Google ecosystem (e.g. YouTube) and other search engines and social media platforms. What autocomplete presents users with – in the process of typing a query – is determined by web indexing methods. These have tended to use techniques such as ‘vector-space modelling’ where sets of keywords are represented as vectors in a high (>3) dimensional space and relevant predictions/suggestions are gathered from similar vectors as per the ‘distance’ between them (Hiemstra, 2009). Although their proprietary nature precludes full transparency, these statistical analyses appear to be based on various factors including the number of queries entered by other users, click-throughs on results, user location and use history, the number, diversity and quality of hits, search language, the explicit occurrence of certain terms in search hits and broader user trends (BBC News, 2012; Hiemstra, 2009; Sullivan, 2011). Google states that autocomplete is based on what someone types, their search history (if logged into Google) and what other people are searching for. Google’s autocomplete policy limits what can feature in autocomplete predictions, potentially removing (for some languages) content that is violent, dangerous, sexually explicit, hateful against groups or disparaging to individuals (Google, n.d.). It allows users to report predictions they find inappropriate, thus also informing algorithmic filtering of content.

Autocomplete functions can facilitate greater efficacy and speed, potentially enhancing usability for different groups, for example, older users (Doubé and Beh, 2012), and people with dyslexia (Berget and Sandnes, 2016). Some researchers have examined how autocomplete can become more precise in its predictions, and better reflect a user’s concerns at a point in time (Bar-Yossef and Kraus, 2011). Others have adopted more normative lines of questioning, interrogating the legality of (and liability for) information presented through predictions (Karapapa and Borghi, 2015; Olteanu et al., 2020) or how such predictions – across platforms and timescales – could steer patterns of user inquiry (Robertson et al., 2019). Baker and Potts (2013) demonstrated how autocomplete predictions can reproduce racist, biased and stereotyped discourses

in English-language searches. The appearance of harmful and objectionable content in autocomplete has increased public and legal scrutiny on companies such as Google, and successful legal action has been taken in instances of defamation of individuals or groups (Hammer, 2013).

Miller and Record (2017) argue that the responsibility for autocomplete results should lie primarily with search providers. They present six features of autocomplete that underpin its power to influence users, and underscore the difficulties of relying on user accountability or automated filtering. These features relate to (1) the fact that popular searches (reflected in autocomplete) are not necessarily useful for the searcher; (2) the *automatic generation* of suggestions; (3) the *feedback loop* that occurs when autocomplete visibility of search queries makes it more likely that terms will be searched for again; (4) users' *involuntary exposure* to suggestions; (5) the dynamic and interactive way in which autocomplete can *affect a user's inquiry* towards new paths and (6) its potential *impact on a user's belief formation* 'in a real time and responsive manner' (Miller and Record, 2017: 1950). They also question whether algorithms can appropriately identify and deal with problematic or offensive autocomplete results given cultural specificities (Miller and Record, 2017: 1954). These reflect wider issues for platform content moderation (Gillespie, 2018) and the necessity for culturally-competent human labour to understand offence, nuance, humour and satire. Much of this (difficult, degrading and low paid) work is outsourced globally by digital platforms (Dwoskin et al., 2019). However, companies' investment in such human labour may not be keeping pace with the quantity and diversity of online content globally, a fact highlighted by the failure of platforms such as Facebook to adequately monitor user activity for incitement of violence. This has been evidenced in English-language contexts, where moderation capacity is higher, and – with arguably more devastating impacts – in conflicted contexts such as Myanmar where the company has not had enough moderators who understand local socio-linguistic settings (Rajagopalan et al., 2018).

For autocomplete, human moderation appears to occur alongside automated filtering, for instance through the manual identification of 'blacklists' of terms which should not appear as predictions. However, we know much less about the extent of human moderation for global multilingual search use. Miller and Record (2017) highlight the importance of cultural context, but do not specifically draw attention to the linguistic challenges that exist across a multilingual world of search engine users. We do not know who is doing this work for different languages, particularly languages that are not (globally) dominant but have sizeable digital communities. There are broader questions here about how effectively algorithms function with different languages given the (machine) readability of multilingual databases of prior searches, as well as wider online content. Data harvesting on online trends also influences autocomplete and it is unclear how visible such trends are in non-dominant global languages.

Baker and Potts (2013) concluded their study by suggesting the need for empirical research into how this might occur in languages other than English. Subsequent research suggests that language does intersect with and shape algorithmic functioning, and algorithms' relationship to users, though again with limited study of non-European languages. Moe (2019) examines ranking of YouTube video search results around Islam in three Scandinavian languages, finding changes in ranking by language

between popular videos. Existing scholarship on algorithmic prediction that extrapolates universal claims about their operation from Anglo/European linguistic contexts and legal systems may fail to account for the mediating role of place and language in algorithmic operation, or their intersection with local socio-cultural/political contexts and global power imbalances.

Addressing this gap, our study focuses on normative aspects of indigenous African languages' interaction with autocomplete algorithms. More broadly, we contribute to critiques of geographical/linguistic biases in how digital platforms are studied. These approaches have accompanied efforts to internationalise or decolonise media and communication studies (Arora, 2019; Mignolo, 2007; Waisbord and Mellado, 2014; Willems and Mano, 2016) and highlight dynamic and diverse uses of indigenous (African) languages on digital platforms (Salawu, 2018). Nonetheless, limited attention has been given to how digital platform algorithms operate in (and with) different linguistic settings, despite prior work on the role of language in mediating post-colonial social, cultural and political worlds (Laitin, 1977; Mazrui, 2019). Mpfu and Salawu's (2020) study of the linguistic localisation of the Google Search engine for users in Zimbabwe highlights dynamics of linguistic colonialism in the Anglo-centrism of cyberspace. The translation of elements of the Google Search interface into African languages demonstrates the company's nod towards global multilingual inclusivity. Mpfu and Salawu argue that this is redundant in the wider context of language policy and practices that privilege English globally. Focussing on the language of the interface, their study shows how linguistic colonialism manifests in a lack of content or functionality for African languages on online platforms. As our analysis of different African contexts demonstrates, the issue here is less about an absence of indigenous language content, but rather the ways in which a search engine algorithm (regardless of how its interface is translated) processes African language keywords that users are entering to find content in those languages.

When technologies such as search engines engage with global multilingual content, they exert forms of algorithmic power that are often connected to state or commercial data surveillance and extractivism (Zuboff, 2019). These phenomena have started to be explored with explicit reference to coloniality and post/de-colonial critiques (Coudry and Mejias, 2019). Oyedemi (2021) interrogates the profit motives of Western companies such as Google in their apparently benevolent provision of 'free' Internet access in rapidly growing African markets, home to millions of users whose data can be monetised through its sale to advertisers. It is necessary here, we argue, to bring together debates in the broader digital colonialism literature on the marginalisation of smaller languages across the Anglo-centric Internet, with consideration of developments in Natural Language Processing (NLP) that may accelerate platforms' data extraction from users. Zaugg's (2019) discussion of digital innovation to promote greater online functionality for languages such as Amharic raised potential dilemmas for activists promoting multilingualism online. As she puts it, 'while the benefits of these supports for digitally-disadvantaged language communities are clear, the reality is that standardised script use, standardised spelling and NLP systems in particular increase a language community's legibility for digital surveillance' (Zaugg, 2019: 227). As we demonstrate with specific case studies, unequal global relations of power do not manifest in the *absence* of

linguistic diversity online, but rather through platforms' (unpredictable) interaction with multilingual content that users are inputting, and the wide range of potential impacts that this can have on individual users and digital publics.

Epistemological technologies such as autocomplete are now a ubiquitous feature of search engines, which themselves shape contemporary (digital) cultures worldwide (Halavais, 2009). Andersen (2018) argues that logics of archiving, ordering and searching that underpin search engines, algorithms and databases represent a form of what Couldry and Hepp (2018) call 'deep mediatisation'. Although users appropriate these tools for their own purposes, the 'ideology' of search impacts their communicative actions, for instance, in the mundane expectation that access to databased information is constantly available. Of course, the mediation of access (through search engines) is not neutral, and a 'critical sociology' (Gillespie, 2016) of 'algorithmic culture' (Hallinan and Striphos, 2016) is required. Again, however, the bulk of this work starts from epistemological positions grounded within the same (offline, geographical) cultures where these algorithms were developed – primarily, 'the West'. As certain algorithms spread from their Silicon Valley origins to non-Western contexts, intercultural encounters occur. Kotliar (2020: 919) notes that the 'spatial trajectories through which algorithms operate and the distances and differences between the people who develop such algorithms and the users their algorithms affect [often] remain overlooked'.

Our empirical study shows how algorithmic cultures can mediate – in sometimes quite jarring and (apparently) uncontrolled ways – discourses that are grounded in specific political and cultural-linguistic settings, and which the (Western) platform has conceivably limited comprehension of. We illustrate this by discussing implications of autocomplete's foregrounding of keywords relating to 'clan' (or 'tribe') in Somali-language political discourse. The impacts of such 'deep mediatisation' can be productively incorporated into debates about identity formation in contested and conflicted environments. In relation to our geographical area of interest, long-standing debates around the salience of ethnicity may need to consider how the changing (algorithmic) media environment affects people's access to information and impacts on identity construction. However, this is not a one-way street. Studies of 'algorithmic culture' can also benefit from a greater appreciation of the potential for user inputs rooted in diverse cultural-linguistic norms or complex transnational political environments to feed back into these technologies' operation. Afrocentric approaches can contribute here to scholars' attempts to articulate more nuanced accounts of algorithms that appreciate their power in different societies but avoid overlooking human agency, as people use, appropriate, understand and interpret (some of) their affordances (Andersen, 2020). Autocomplete algorithms matter to meaning making, and play a role in shaping social and political imaginaries and relations. In the contexts we examine, these impacts will only increase with further digital penetration of the African continent and the spread of predictive language technologies across platforms.

Exploring African language autocomplete

We problematise the algorithmic power of autocomplete in three African-language contexts through empirical experimentation on how these algorithms interact with a sample

of search terms in Amharic, Kiswahili and Somali. Each of these languages are important in the East/Horn of Africa and constitute *lingua franca* in Ethiopia, Kenya/Tanzania and Somalia. Each are from different linguistic families and together have tens of millions of speakers in the region, with more in global diaspora communities. Although digital platforms such as Google and Facebook are increasingly used by people in this large and diverse region, such technologies were developed elsewhere and not necessarily designed with such global users and global languages in mind (Arnauld, 2019). Our everyday use of Google Search with these languages initially indicated that many keywords produced autocomplete suggestions in both English and these languages. Each has different orthographical features which affect autocomplete returns. This is both a finding of this paper (which we detail below) and a factor that shaped the comparative and iterative choice of keywords we used to explore this phenomenon.

Amharic is a constitutionally-recognised national language in multi-ethnic Ethiopia and can be regarded as a national *lingua franca*. It is spoken by at least a third of Ethiopia's 115 million population, and in global diasporic communities. Amharic is a left-to-right written language with its own Ethiopic script or alphabet called 'abugida'. This contains 33 letters, each taking 7 forms. Prior to the 12th century, Amharic mainly served as a court language, whereas the related Ge'ez which operated as a language of literature used primarily by clerics of the Ethiopian Orthodox church (Meyer, 2006). Amharic became a main language of literature in the 19th Century (Meyer, 2006). According to Eko (2007: 24), the digital history of Ethiopic script can be traced to the efforts of US-based Ethiopian computer scientists in the 1980s, their translation of Ge'ez into code and fonts, and the creation of the first Ethiopic word processor. Zaugg (2017) outlines how late 20th century efforts enabled the inclusion of the Ethiopic script in Unicode-ISO/IEC 1046, the digital standards allowing scripts to appear on devices and software/applications. As Amharic uses its own distinctive Ethiopic script, it can be assumed that the autocomplete returns provided by Google Search are influenced by the search behaviour of other Amharic speakers, located either in Ethiopia or elsewhere.

Kiswahili is a Bantu language, which is spoken widely in East Africa. It dates to at least the 10th century (Mazrui and Mazrui, 1995), and is a national language in the DRC, Kenya, Tanzania and Uganda. Kiswahili is distinct in that the number of speakers for whom it is a second language far outnumbers first-language speakers. Outside of Tanzania, where Kiswahili is the official language for business and parliament, it is often spoken in combination with other African languages and English. Kiswahili has a strong presence of Arabic and English loan words through historical Indian Ocean trade connections and European colonisation in East Africa. Like Somali, Kiswahili now uses a Latin orthography; though Arabic script was used prior to the colonial encounter (Mazrui and Mazrui, 1993). Unlike Somali, Kiswahili names of people tend to be spelt the same as in English. Therefore, autocomplete for a search on the Kenyan president (Uhuru Kenyatta) would therefore be influenced by search behaviour of both Kiswahili-speaking Kenyans and global users. Searches for figures with higher international profiles would be more likely to be undertaken by global audiences, therefore increasing the prevalence of English language autocomplete predictions and making it harder to pinpoint information-seeking practices of particular linguistic communities.

Somali is a Cushitic language spoken widely across Somalia, Djibouti and parts of Ethiopia and Kenya. It is also used across the widespread global diaspora. Modern ‘standard’ Somali (*Maxaa tiri*) is an amalgamation of various northern and central Somali dialects (Lamberti, 1986) and was entrenched as the Somali *lingua franca* by Somalia’s military regime in the 1970s. The Somali territories of the Horn of Africa have a rich oral literary heritage, however it was not until 1972 that a written orthography for the language was formalised and used in mass literacy campaigns (Laitin, 1977). There are various minoritised languages spoken within Somalia and regional variations in northern and southern *Maxaa tiri*. Nonetheless, this ‘standard’ Somali is widely understood across the politically fragmented territories and is used across global Somali media networks (Chonka, 2019). Somali orthography uses a Latin script convenient for devices and platforms designed in and for other global markets. Somali-spelt search terms in our testing returned Somali language auto-complete, interspersed with occasional English suggestions/predictions. In the case of names, transliterations have different levels of similarity between English and Somali spellings (Mohamed/Maxamed; or Farmajo/Farmaajo). The similarity of Somali search terms with English language inputs has a significant bearing on auto-complete returns.

Observation that orthographic and spelling features of the languages influenced autocomplete informed the design of our testing. As such, we did not design a fully systematic comparative test of autocomplete for the same keywords across the languages. Instead, we developed three sets of gender and politicised keywords relevant in each context. This was informed by our initial encounters with multilingual autocomplete and early identification of certain notable (and potentially) problematic dynamics. For Amharic the gendered keywords were: ‘ሴቶች’ (women), ‘ሴት’ (woman), ‘ልጅ ገረድ’ (Girl), ‘ወንዶች’ (Men) and ‘ወንድ’ (Man). We also tested several keywords associated with (then) current or contentious issues. For Somali, the term ‘girl’, specifically the northern dialect spelling (*gabadh*), was tested as the southern spelling (*gabar*) resembled various non-Somali keywords and returned non-Somali autocomplete results. Political keywords focussed on the Somali spelling of names of male and female politicians. For Swahili, gendered nouns were tested (girl/girls, *msichana/wasichana*). The names of Kenyan politicians were tested on their own, and with a verb or conjunction (both in English and Kiswahili). Also tested were the terms *uislamu* (Islam), *ukristo* (Christianity), *ufisadi* (corruption) and *virusi vya corona* (coronavirus).

The Amharic, Kiswahili, and Somali tests were conducted in the UK, whereas the Somali test was also conducted in Somalia to understand the impact of location. For each test, terms were typed into Google Search and a screenshot was taken when autocomplete predictions appeared. The researchers used private browser settings and remained logged out of Google to limit the impact of prior searches. Data was recorded and translated into English. In some cases, changes in autocomplete results were noticed after initial tests. At no point in the testing were specific search terms (or any of their associated autocompletes) clicked through to return search results. Were we to complete searches, this would influence future autocomplete results, which would have been ethically undesirable.

Search date	Location	Search term entered (Amharic)	Search term (English translation)	Autocomplete suggestion/prediction (in order)	English translation of prediction
2020.09.28	UK	ሴት	Woman	ሴትን ለማርካት	To satisfy a woman
				ሴት	Woman
				ሴትን ቶሎ ለማርካት	To satisfy a woman fast
				ሴት ልጅ ስትውድህ	When a woman loves you
				ሴትን ለመብዳት	To f**k a woman
				ሴትን ፍቅር ለማሰዳዝ	Making a woman love you
				ሴትን በሰልክ መጥበስ	Hacking a woman (for love) via telephone
				ሴት ልጅ ስታፈቅር	When a woman loves
				ሴትን መብዳት	F**king a woman
				ሴትን መላም	Kissing a woman
2020.09.28		ልጃገረድ	Girl	ልጃገረድ	Girl
				ልጃገረድ ሴት	Girl woman
				ልጃገረድ አምስ	Girl vagina
				ልጃገረድ አስከ አምስት	Girl up to five
				ልጃገረድ translation	Girl "translation"
				ልጃገረድ ትርጉም	Girl meaning
				ልጃገረድ ፍቅር	Girl love
				ልጃገረድ መገናድ	Girl way/road
				ባላላ ልጃገረድ	Naughty girl
				የቤት ልጃገረድ	House girl

Figure 2. Autocomplete prediction for gendered search terms in Amharic.

Gendered keywords in autocomplete

Our first set of results indicates that autocomplete fails to operate according to Google’s own policies to remove harmful predictions. Sexually explicit, vulgar or profane keywords (with an exception for human anatomy or education) and hateful keywords against groups (e.g. gender, race or ethnic origin, religion, age, nationality) are among the autocomplete suggestions/predictions that Google states it tries to exclude, through manual and automated filtering and user identification. In searching for gender-related terms, it became clear that the parameters set for autocomplete results varied between the three languages, with more sexually explicit content appearing in the Amharic script. Here, almost all suggestions associated with the keywords: ‘women’, ‘woman’ and ‘girl’ were sexualised and misogynistic. Figure 2 provides an example of the autocomplete suggestions for keywords: ‘woman’ and ‘girl’ in Amharic, in the original script and translated into English. The results for woman are sexually explicit, including ‘to f**k a woman’ and ‘f**king a woman’. The results for girl are less explicit but still include some sexual connotations, for example, ‘naughty girl’ and ‘girl vagina’.

The search in Somali also included sexualised results for girl, for example ‘naked’ or ‘virgin’ or ‘recorded on video’. However, ‘naked’, which was the top suggestion when the search was conducted in September 2020, was absent by June 2021. Whereas Figure 2 shows autocomplete for two different gendered search terms, Figure 3 shows results for one gendered Somali keyword at different times (and changes in the results):

The Kiswahili search did not result in any sexually explicit predictions/suggestions (Figure 4). Results referred instead to development initiatives for girls, translation requests and a song title. Predictions in other languages were quite common in the

Search date	Location	Search term entered (Somali)	Search term (English translation)	Autocomplete suggestion/prediction (in order)	English translation of prediction
2020.09.17	UK	<i>Gabadh</i>	Girl	<i>qaawan</i>	naked
				<i>ku jecel</i>	[that] loves you
				<i>sunada ah</i>	[circumcised according to the] Sunnah
				<i>video laga duubay</i>	recorded on video
				<i>wanaagsan</i>	good
				<i>qurux badan</i>	very beautiful
				<i>safiirka</i>	the ambassador
				<i>a</i>	the
				<i>bikrada ah</i>	virgin
2021.06.07	UK	<i>Gabadh</i>	Girl	<i>ku jecel</i>	[that] loves you
				<i>wanaagsan</i>	good
				<i>a</i>	the
				<i>xijaaban</i>	[wearing] hijab
				<i>bikrada ah</i>	virgin
				<i>safiirka</i>	the ambassador
				<i>ii ilaahay u waxyooday</i>	[the girl] revealed by God
				<i>maayo</i>	mermaid
				<i>a ku rabto</i>	[the girl who] wants you

Figure 3. Autocomplete for gendered search terms in Somali.

Search date	Location	Search term entered (Swahili)	Search term (English translation)	Autocomplete suggestion/prediction (in order)	English translation of prediction
2021.04.21	UK	<i>Msichana</i>	Girl	<i>msichana empowerment</i>	girl empowerment growth
				<i>kuria</i>	girl initiative
				<i>msichana initiative</i>	girl initiative
				<i>msichana</i>	girl
				<i>msichana in english</i>	girl in English
				<i>Msichana Wa Elimu (song by Daudi Kabaka)</i>	Educated/educational girl (Song by Daudi Kabaka)
				<i>msichana meaning</i>	girl meaning
				<i>msichana wa elimu mp3 download</i>	Educated girl mp3 download
				<i>msichana wa elimu lyrics</i>	lyrics of Educated girl
				<i>msichana wa taekwondo</i>	Taekwondo girl
				<i>msichana uganda</i>	Ugandan girl

Figure 4. Autocomplete for gendered search terms in Kiswahili.

Kiswahili search. For example, *Mke* (wife) returns suggestions of places and people in English, including an Airport in Milwaukee; MKE Akaragucu (football team) and Mike Tyson. The last of these predicts an (English) spelling mistake, rather than the Kiswahili word. This shows how algorithmic results may misalign with the intentions of a Kiswahili-speaker, but still operate within the confines of Google’s autocomplete content moderation policies.

Given cultural nuances, algorithms' ability to identify and deal with problematic or offensive content has been questioned elsewhere (Miller and Record, 2017). Our findings suggest that there are issues even with content moderation where strong shared norms exist about what could constitute harmful content, potentially due to the limited attention paid by Google to 'learning' African languages and scripts, or by fewer users inputting into the algorithm. This reveals both diversity and uncertainty in how content moderation works across languages and contexts. The algorithm is clearly working differently across languages, particularly Amharic with its own script and a relatively small number of Google users using the script. This even occurs where content clearly aligns with Google's definition of problematic content as that which is sexually explicit.

A possible response to these issues is to call for more attention to content moderation and algorithm training in different languages by the search provider. For sexually explicit content, an algorithmic fix, if invested in across all languages, could help remove obviously harmful/explicit suggestions. Indeed, as our analysis has shown that some terms have disappeared over time, this type of algorithmic moderation may already be taking place. However, the lack of a public record available on which terms are being filtered and removed in autocomplete (and for which languages), raises questions about transparency and the ability of members of global language communities to hold to account the decision-making underpinning algorithmic power. While we return to this in the conclusion, the following section highlights difficulties inherent in managing lists of problematic keywords across multiple cultural-linguistic contexts. Here, the potentially 'easy' fix of using more data from more languages to train algorithms to filter out inappropriate suggestions could obscure other challenges of content moderation related to specific socio-political settings.

Political keywords in autocomplete: (in)visible identities

While many autocomplete returns for the local-language search terms 'girl' or 'woman' were clear in their sexualised or misogynistic character, other areas of our testing highlighted more ambiguously contentious suggestions related to specific dynamics within states affected by conflict. Our experiments with Somali autocomplete for searches for names of political figures consistently returned suggestions relating to clan or ethnicity. This section considers the significance of such identity markers and the potential influence of transnational algorithmic power on popular discourse and identity formation in a region affected by protracted conflict. When testing search predictions in Somali for the names of 10 male politicians, 6 returned clan related keywords, primarily the noun for 'clan' in Somali (*qabiil*). These autocomplete returns thus present users with the (arguably increased) opportunity to search for an individual's clan. One of the predictions returned the actual name of the clan of the politician, and another the name of a clan group with whom that historical figure had a contentious relationship.

Somali society is often understood as being structured around clan groups which trace their lineages back through common male ancestors. The Somali Horn of Africa is also a region affected by political fragmentation and – in some places – ongoing armed conflict. Clan divisions have undoubtedly played a part in the Somali civil war, state collapse in the early 1990s and current political divides (Ahmed, 1995; Besteman, 1998; Kapteijns,

2012; Samatar, 1992). Space precludes a full explanation that draws from this literature to show the evolution and manipulation of the concept of clan from the colonial period into the rise and fall of the unified post-independence Republic of Somalia. However, certain points are salient for our analysis and relate to the context of modern Somali politics and the information environment where Somalis debate political identity. ‘Clannism’ is frequently condemned by Somalis as being a problem that affects the (global) Somali community and feeds into local conflict. It is often a taboo for clan identities to be discussed explicitly and openly in local news media. Yet, in some Somali political structures, clan identities are partially institutionalised – such as in Somaliland’s Upper House of clan elders, or in the ‘4.5’ system that continues to be used to divide up seats in Somalia’s parliament into clan quotas.

With clan/ethnic difference in the Somali territories not being marked by linguistic diversity in the same way (or extent) as in its multi-ethnic neighbouring states (including Kenya and Ethiopia) it is conceivable that people use search engines to access information relating to individuals’ clan identity – making a so-called ‘invisible’ difference (Lewis, 2004) visible. The fact that these predictions appear at all indicates that some people have used them in searches – although it is unclear how many. It is also conceivable that their appearance for subsequent users may lead to higher frequencies of searches for those keywords, even in instances where the user was not initially considering that query.

It is also impossible to know from our testing whether the users who enter these types of keywords are based in the Somali territories (where Google Search usage is potentially lower) or are Somali-speaking users in the global Somali diaspora. The latter scenario would raise a question about how global Somali internet users’ online activity outside of the region may impact the information environment in the Horn of Africa itself. For the Somali keywords, an earlier trial of the test with politicians’ names involved one of the authors in the UK and a colleague in Somalia to account for the potential impact of location. No significant differences were noticeable between the suggestions returned. This indicates a potentially limited amount of location-based personalisation of search engine returns. This could reflect the limited harvesting of local language data from Somali internet users. On one hand, this might be considered to be somewhat emancipatory – from the algorithmic hyper-individualisation of content (based on harvesting of data from users and targeted content) that is argued to affect Western public spheres through its role in creating ‘filter bubbles’ or ‘echo chambers’ of polarised debate. However, a lack of personalisation for autocomplete means that particular predictions will be visible for wider audiences, potentially extending the reach of problematic keywords.

Overall, this data shows how potentially sensitive political keywords are made visible in search in a setting affected by conflict and political instability, dynamics which are themselves influenced by such markers of identity. This phenomenon does not appear to have been discussed in the region, and these are questions which we would argue require engagement from those people who may be most greatly affected. The influence of digital diasporas (Brinkerhoff, 2009) on ‘homeland’ conflict dynamics is an emerging topic of research more widely (Ogunyemi, 2017; Osman, 2017). This often focuses on provocative debates or content generated outside of the region and fed back into the Somali territories through social media platforms. Moving beyond contentious content itself, our

case shows a potential algorithmic impact of external information seeking practices on platforms used outside and within the region. At the time of writing, the ethnic-federal makeup of Ethiopia is challenged by armed conflict. Ethnic cleavages have widened, and the role of social media in the spread of hate speech and mis/disinformation has been frequently noted (Wilmot et al., 2020). Although the linguistic context differs with Somalia (where ethnicity is less clearly tied to language), digital platforms are a key theatre in Ethiopia's ethnicised conflict. Assessing how predictive language technologies may intersect with social media manipulation and the spread of polarising content is beyond the scope of this paper but demands future attention.

Conclusion

A burgeoning literature details the impacts of different forms of algorithmic power on aspects of social, cultural and political life worldwide. Nonetheless, the ways in which (Western) platform algorithms interact with a world of linguistic diversity on the global Internet are rarely scrutinised in detail. This is particularly true for languages and regions marginalised in the global context of technology production. Studies in digital culture that consider and critique global power imbalances often illuminate the Anglo-centrism of the Internet, demonstrating how this disadvantages particular language communities and reflects dynamics of digital colonialism. Our article has brought research agendas on algorithmic power and digital colonialism together, and in doing so, highlights a different set of problems and dilemmas. Our experimentation was not premised on the lack of indigenous language content online in East Africa. Instead, our case studies highlight how users in and beyond the region are using languages such as Amharic, Somali and Kiswahili to search for content that exists in these languages – and that which our previous research has demonstrated contributes to East Africa's vibrant multilingual digital publics (Chonka, 2019; Diepeveen, 2021).

In these contexts, algorithmic power – which reflects and reinforces wider geo-political and historic power imbalances – is clearly visible. Building on critiques of technologies such as autocomplete that problematise tech platforms' distinctions between predictive and suggestive results, we have detailed how automated and unsolicited presentation of possible search keywords may have a recursive impact on future practices of information seeking. However, our objective has not simply been to contribute additional case studies for diversity's sake, but instead highlight issues that emerge when such algorithms (trained primarily on Western language data) process non-Western language inputs. Partly this relates to content moderation, and the capacity, interest or willingness of a platform like Google to police autocomplete results that are returned by indigenous language search inputs. We have highlighted examples of highly problematic keywords that would likely be removed in English or European language contexts. The fact that these suggestions appear in autocomplete is evidence that users who speak these languages have searched for those terms. However, making these terms repeatedly visible to other users amplifies their impact in the digital public and recursively contributes to the increased likelihood that they will appear in future searches.

This paper does not advocate for a technical fix to filter keywords. The contextual nuances of the clan-related autocomplete results reveal the difficulty of automating the

identification of contextually specific results. There are already specific incentives for companies such as Google to develop their natural language processing (NLP) capabilities, and from anecdotal observation it is clear that tools like Google Translate for languages such as Somali are becoming increasingly effective. While such developments are visible, they must contend with the effects of past inattention and challenges in developing NLP for non-Latin languages that do not easily align with UNICODE standards and operating systems (Arnauldo, 2019: 102–103). Furthermore, recent developments in this field are driven by commercial imperatives to expand global markets for user data extraction, as opposed to content moderation, although increased NLP capacity may enhance companies' ability to perform the latter when they are put under public pressure to do so. There is an emerging tension here between increased inclusion and digital legibility for global languages (to promote digital multilingualism and potentially enable algorithmic protection for users from certain digital harms), set against the prospect of expanded data extractivism and commercial/political surveillance (Zaugg, 2019). For us, it is evident that a lack of African language processing and oversight for autocomplete exposes users to potential digital harms, such as the increased likelihood of encountering problematic or offensive suggestions. Equally, improving algorithmic linguistic sensitivity for tools like autocomplete may expose users to alternative future harms. These could include the increased capacities for state or corporate surveillance and the extraction of data for the curation of search results, newsfeeds and targeted advertisements – all of which are linked in Western contexts to losses of individual agency and autonomy online.

Our contextualised experimentation illustrates the complex political, cultural and social terrain upon which (Western) algorithmic power can make unanticipated and jarring interventions. Considering this, we might question the very legitimacy of autocomplete's unrestrained functioning across the global multilingual Internet. If it is unfeasible for the developers of this technology to ensure that harmful content cannot appear as predictive suggestions across the spectrum of digital languages, then does this risk outweigh the purported benefits of the tool? Much as we see the value of engaging in this debate, we are realistic about the prospects of an abolitionist approach to technologies like autocomplete, even if they so clearly exemplify the implications of algorithmic power for individual autonomy of thought and (information seeking) behaviour. At a minimum, our results point to the need for greater transparency around how these technologies interact with user inputs in various global settings to gauge the *extent* to which content moderation is happening and – crucially – *how* it is being undertaken. Our experimentation showed that certain prominent and problematic autocomplete suggestions had disappeared over the course of a year. For the language communities most directly affected by such results, it is currently impossible to ascertain whether such terms were removed because of automated filtering, manual content moderation, factors that reduced their relevance in searches (e.g. a drop in use by users), or other reasons. Especially where there may be legitimate debate about (in)appropriate content, it is essential that people can find out which terms have been removed and for what reasons, in order to support informed engagement and advocacy by affected linguistic groups. Tech companies may consider this level of transparency to be onerous or impracticable, but this would again raise the question as to whether the purported benefits of such technologies can be realistically guaranteed to outweigh their potential harms. In our case study, the

current information asymmetry that exists between African users and Google merely reflects wider power imbalances between the continent and a range of external tech companies whose algorithmic power is – as we have demonstrated – shaping contemporary multilingual digital publics around the world.

Acknowledgment

The authors are grateful to Mahad Wasuge who assisted with testing autocomplete in Somalia and discussing translations of particular results. The authors would also like to thank the two anonymous peer reviewers for their constructive comments and suggestions.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iD

Peter Chonka  <https://orcid.org/0000-0001-9871-1377>

References

- Ahmed AJ (1995) *The Invention of Somalia*. Trenton: Red Sea.
- Allcott H and Gentzkow M (2017) Social media and fake news in the 2016 election. *Journal of Economic Perspectives* 31(2): 211–236.
- Andersen J (2018) Archiving, ordering, and searching: Search engines, algorithms, databases, and deep mediatization. *Media Culture & Society* 40(8): 1135–1150.
- Andersen J (2020) Understanding and interpreting algorithms: toward a hermeneutics of algorithms. *Media, Culture & Society*, 42(7–8), 1479–1494.
- Anderson R (2012) Molefi Kete Asante: The Afrocentric idea and the cultural turn in intercultural communication studies. *International Journal of Intercultural Relations* 36(6): 760–769.
- Arnauddo D (2019) Bridging the deepest digital divides: A history and survey of digital media in Myanmar. In: Punathambekarand A and Mohan S (eds) *Global Digital Cultures: Perspectives From South Asia*. Ann Arbor, MI: University of Michigan Press, pp.96–121.
- Arora P (2019) *The Next Billion Users: Digital Life Beyond the West*. Cambridge: Harvard University Press.
- Baker P and Potts A (2013) ‘Why do white people have thin lips?’ Google and the perpetuation of stereotypes via autocomplete search forms. *Critical Discourse Studies* 10(2): 187–204.
- Bar-Yossef Z and Kraus N (2011) Context-sensitive query auto-completion. In: *Proceedings of the 20th international conference on world wide web*. Hyderabad: International World Wide Web Conference Committee (IW3C2), pp. 107–116.
- BBC News (2012) Google autocomplete: The internet where you are. *BBC News*, 7 September. Available at: www.bbc.co.uk/news/world-us-canada-19258912 (accessed 1 June 2021).
- Berget G and Sandnes FE (2016) Do autocomplete functions reduce the impact of dyslexia on information-searching behavior? The case of Google. *Journal of the Association for Information Science and Technology* 67(10): 2320–2328.
- Besteman C (1998) Primordialist blinders: A reply to I. M. Lewis. *Cultural Anthropology* 13(1): 109–120.
- Brinkerhoff JM (2009) *Digital Diasporas: Identity and Transnational Engagement*. Cambridge: Cambridge University Press.

- Bucher T (2018) *If. . .Then: Algorithmic Power and Politics*. Oxford: Oxford University Press.
- Chonka P (2019) News media and political contestation in the Somali territories: Defining the parameters of a transnational digital public. *Journal of Eastern African Studies*, 13(1), 140-157.
- Couldry N and Hepp A (2018) *The Mediated Construction of Reality*. Oxford: Wiley.
- Couldry N and Mejias UA (2019) *The Costs of Connection: How Data is Colonizing Human Life and Appropriating It for Capitalism*. Stanford, CA: Stanford University Press.
- Diepeveen S (2021) *Searching for a New Kenya: Politics and Social Media on the Streets of Mombasa*. Cambridge: Cambridge University Press.
- Doubé W and Beh J (2012) Typing over autocomplete: Cognitive load in website use by older adults. In: *Proceedings of the 24th Australian computer-human interaction conference*. Melbourne: OzCHI'12, pp. 97-106.
- Dwoskin E, Whalen J and Cabato R (2019) Content moderators at YouTube, Facebook and Twitter see the worst of the web. *Washington Post*. Available at: <https://www.washingtonpost.com/technology/2019/07/25/social-media-companies-are-outsourcing-their-dirty-work-philippines-generation-workers-is-paying-price/> (accessed 26 April 2021).
- Eko L (2007) Africa: Life in the margins of globalisation. In: Artz L and Kamalipour YR (eds) *The Media Globe Trends in International Mass Media*. Lanham, MD: Rowman and Littlefield, pp.7-32.
- Gillespie T (2016) #trendingtrending: when algorithms become culture. In: Seyfert R and Roberge J (eds) *Algorithmic Cultures: Essays on Meaning, Performance and New Technologies*. London: Routledge, pp. 52-75.
- Gillespie T (2018) *Custodians of the Internet: Platforms, Content Moderation, and the Hidden Decisions That Shape Social Media*. New Haven, CT: Yale University Press.
- GlobalStats (2021) Search engine market share Africa. Available at: <https://gs.statcounter.com/search-engine-market-share/all/africa> (accessed 8 June 2021).
- GlobalStats (2022a) Mobile operating system market share worldwide. Available at: <https://gs.statcounter.com/os-market-share/mobile/worldwide> (accessed 21 March 2022).
- GlobalStats (2022b) Browser market share worldwide. Available at: <https://gs.statcounter.com/browser-market-share> (accessed 21 March 2022).
- Google (n.d.) How Google autocomplete predictions work. Available at: <https://support.google.com/websearch/answer/7368877> (accessed 30 May 2021).
- Gray J, Bounegru L and Venturini T (2020) 'Fake news' as infrastructural uncanny. *New Media & Society* 22(2): 317-341.
- Halavais A (2009) *Search Engine Society*. London: Wiley.
- Hallinan B and Striphas T (2016) Recommended for you: The Netflix prize and the production of algorithmic culture. *New Media & Society* 18(1): 117-137.
- Hammer D (2013) Google settles anti-defamation suit in Israel. *Law.co.il*. Available at: <https://www.law.co.il/news/2013/11/28/settlement-reached-in-Google-auto-complete-defamation-suit/> (accessed 8 May 2021).
- Hiemstra D (2009) Information retrieval models. In: Goker A and Davies J (eds) *Information Retrieval: Searching in the 21st Century*. London: Wiley, pp.1-20.
- Kapteijns L (2012) *Clan Cleansing in Somalia: The Ruinous Legacy of 1991*. Philadelphia: University of Pennsylvania Press.
- Karapapa S and Borghi M (2015) Search engine liability for autocomplete suggestions: Personality, privacy and the power of the algorithm. *International Journal of Law and Information Technology* 23(3): 261-289.
- Kotliar DM (2020) Data orientalism: On the algorithmic construction of the non-Western other. *Theory and Society* 49(5-6): 919-939.

- Kwet M (2019) Digital colonialism: US empire and the new imperialism in the Global South. *Race & Class* 60(4): 3–26.
- Laitin D (1977) *Politics, Language, and Thought: The Somali Experience*. Chicago, IL: University of Chicago Press.
- Lamberti M (1986) *Die Somali-Dialekte: Eine vergleichende Untersuchung*, vol. 5. Hamburg: Buske Verlag.
- Lewis I (2004) Visible and invisible differences: The Somali paradox. *Africa* 74(4): 489–515.
- Madianou M (2019) Technocolonialism: Digital innovation and data practices in the humanitarian response to refugee crises. *Social Media + Society* 5(3): 1–13.
- Mazrui AA (2019) *The Political Sociology of English Language: An African Perspective*. The Hague: Mouton.
- Mazrui AA and Mazrui AM (1995) *Swahili State and Society: The Political Economy of an African Language*. London: Currey.
- Mazrui AM and Mazrui AA (1993) Dominant languages in a plural society: English and Kiswahili in post-colonial East Africa. *International Political Science Review* 14(3): 275–292.
- Meyer R (2006) Amharic as lingua franca in Ethiopia. *Lissan: Journal of African Languages and Linguistics* 20(1/2): 117–132.
- Mignolo WD (2007) Delinking. *Cultural Studies* 21(2–3): 449–514.
- Miller B and Record I (2017) Responsible epistemic technologies: A social-epistemological analysis of autocompleted web search. *New Media & Society* 19(12): 1945–1963.
- Moe H (2019) Comparing platform ‘ranking cultures’ across languages: The case of Islam on YouTube in Scandinavia. *Social Media + Society* 5(1): 1–10.
- Mpofu P and Salawu A (2020) African language use in the digital public sphere: Functionality of the localised Google webpage in Zimbabwe. *South African Journal of African Languages* 40(1): 76–84.
- Noble SU (2013) Google search: Hyper-visibility as a means of rendering black women and girls invisible. *InVisible Culture* 19: 1–21.
- Noble SU (2018) *Algorithms of Oppression: How Search Engines Reinforce Racism*. New York: New York University Press.
- Ogunyemi O (2017) Introduction. In: Ogunwemi O (ed.) *Media, Diaspora and Conflict*. London: Palgrave Macmillan, pp.1–16.
- Olteanu A, Diaz F and Kazai G (2020) When are search completion suggestions problematic? In: *Proceedings of the ACM on human-computer interaction*, 4(CSCW2), vol. 171. New York: ACM, pp. 1–25.
- O’Neil C (2017) *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. New York: Penguin.
- Osman I (2017) *Media, Diaspora and the Somali Conflict*. London: Palgrave Macmillan.
- Oyedemi TD (2021) Digital coloniality and ‘next billion users’: The political economy of Google Station in Nigeria. *Information Communication & Society* 24(3): 329–343.
- Pasquale F (2016) *The Black Box Society: The Secret algorithms That Control Money and Information*. Harvard: Harvard University Press.
- Pohjonen M and Udupa S (2017) Extreme speech online: An anthropological critique of hate speech debates. *Journal of International Communication* 11: 19.
- Rajagopalan M, Thuy VL and Naing SA (2018) How Facebook failed the Rohingya in Myanmar. *BuzzFeed News*. Available at: <https://www.buzzfeednews.com/article/meghara/facebook-myanmar-rohingya-genocide> (accessed 11 June 2020).
- Robertson RE, Jiang S, Lazer D, et al. (2019) Auditing autocomplete: Suggestion networks and recursive algorithm interrogation. In: *Proceedings of the 10th ACM conference on web science*. New York: ACM, pp. 235–244.

- Salawu A (2018) *African Language Digital Media and Communication*. London: Routledge.
- Samatar AI (1992) Destruction of state and society in Somalia: Beyond the tribal convention. *The Journal of Modern African Studies* 30(4): 625–641.
- Sullivan D (2011) How Google Instant's autosuggestions work. *Search Engine Land*. Available at: <https://www.searchengineland.com/how-google-instant-autocomplete-suggestions-work-62592> (accessed 6 June 2021).
- Thornton P (2017) Geographies of (con)text: Language and structure in a digital age. *Computational Culture* 6: 1–29.
- Waisbord S and Mellado C (2014) De-westernizing communication studies: A reassessment. *Communication Theory* 24(4): 361–372.
- Willems W and Mano W (2016) Decolonizing and provincializing audience and internet studies: Contextual approaches from African vantage points. In: Willems W and Mano W (eds) *Everyday Media Culture in Africa*. London: Routledge, pp.15–40.
- Wilmot C, Tveteraas E and Drew A (2020) Dueling information campaigns: The war over the narrative in Tigray. *Media Manipulation*. Available at: <https://mediamanipulation.org/case-studies/dueling-information-campaigns-war-over-narrative-tigray> (accessed 1 December 2021).
- Zaugg I (2017) *Digitizing ethiopic: Coding for linguistic continuity in the face of digital extinction*. PhD Thesis, American University.
- Zaugg I (2019) Digital surveillance and digitally-disadvantaged language communities. In: *Proceedings of the language technologies for all (LT4All)*. Paris: UNESCO, pp. 227–230.
- Zuboff S (2019) *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. London: Profile Books.