



DECOLONIZING DATA, ONE LANGUAGE AT A TIME

Claudia Magallanes Blanco
Sabelo Mhlambi
Nanjala Nyabola

Nick Couldry
Toussaint Nothias
Kathleen Siminyu

2023



2023 Claudia Magallanes Blanco, Nick Couldry,
Sabelo Mhlambi, Toussaint Nothias, Nanjala Nyabola, Kathleen Siminyu.
Attribution and share alike.

Spanish translation: **Silvia Coca Córdova**
Kiswahili translation: **Alfred Mtawali**

pacscenter.stanford.edu/digital-civil-society-lab
tierracomun.net/



INTRODUCTION

At a time of growing anxiety about the ubiquity of digital technologies, the idea that we are witnessing the age of digital and data colonialism is gaining traction. In recent years, scholars and activists have increasingly relied on the notion of coloniality to understand the social impact of data-centric technologies, discuss their harms, and imagine liberating and empowering alternatives.

Central to this area of research, design, and advocacy is the question of languages. In March 2022, the Digital Civil Society Lab at Stanford University and the Tierra Común network held a virtual conversation with speakers bringing various perspectives and insights to the topic. Their work invites us to reconsider the place of languages in the development of technologies but also in the public framing of these technologies. Together, the speakers explored why language gaps in digital technologies are so significant to address and how they are entrenched in a colonial legacy. They discussed many creative ways communities have been responding to these language-related challenges. And they reflected on what decolonial perspectives on language and data mean in practice.

This document is the outcome of this conversation. Rather than releasing a video recording of the event, we have prepared an edited version of the transcript. Our goal is to introduce some generative friction in how we disseminate information to invite people to slow down and engage more thoughtfully with the conversation. You won't immediately hear the many accents in our discussion, but you will encounter words in different languages. In addition, we are also releasing a Spanish and Kiswahili translation of the transcript. The choice of these two languages reflects the Latin American and African emphasis in the experiences and projects discussed by the speakers.

We appreciate you taking the time to engage with this document and hope it will help you imagine countervailing visions of what a more just digital world can look like.

Toussaint Nothias

NANJALA NYABOLA:

Thank you all so much for being here today. I love and relish the opportunity to speak about what we've been working on for the last year and a half now. A lot of it is really thanks to the Digital Civil Society Fellowship from Stanford, which made it possible to pursue something that had been on my mind – and probably to anybody who has been working in tech, and certainly digital rights, for the last 10-15 years.

The genesis of the **Kiswahili Digital Rights Project** was my own work in digital rights. As a multilingual person who grew up speaking three languages and studied a couple more through school and through university, I realized the gap that existed between the linguistic support available for African languages and the linguistic support available to other languages—European languages—that have far fewer speakers and far fewer of a geographical reach. In terms of investments to create a linguistically rich context, the gap was just huge. And I'm not really speaking about technical gaps, which I think the other speakers are probably more proficient in. I'm really talking about it in practical terms. How does language function as an opportunity, or as a space, where people are invited to engage with technology on their own terms?

Anybody who grows up multilingual will tell you that you are a slightly different person in the different languages that you speak. I tell people all the time, I think I am funnier in Kiswahili. Maybe I am not funny at all, but you can bring about different sides of your

personality, of your perspectives on society, of your perspectives on different things, because the language makes certain words available. One language makes certain words available that another language doesn't necessarily make available. For example, in my family we grew up speaking three languages; you will use all three in one sentence, because the word for a specific action exists in the mother language that doesn't exist in Kiswahili, and the verb exists in Kiswahili but it doesn't exist in English. It's very typical for multilingual people to experience the world differently because of this linguistic facility. And that was the personal genesis of this project—realizing that I was unable to express my own interest in digital rights, even in Kiswahili, which is a taught language, a researched and well supported language. There is an academy of Kiswahili, much like there is an academy of French and German. There are research institutes in multiple universities across East Africa and across North America and Europe. And yet there was this big gap for expressing very foundational concepts of digital rights in Kiswahili.

In the 10-15 years that I have been working in digital rights, writing and organizing around these various issues, there would always be this wall that we would hit upon. For example, if you were doing a campaign around the Huduma Namba [Kenya's national digital ID program, or "service number," mandated by the government], which was the precipitating immediate event, we were trying to explain to people why the government should not be able to collect your data for digital ID and without your informed consent and without telling you what it is that they are using data

for and not to be able to monetize that. We would want to say things like “data protection” and “data privacy,” and we will be doing this grassroots organizing and saying things like “ni muhimu sana kulinda privacy yako”. We would start with the English word without its linguistic context, without any kind of support around it. This continuously happens in various events, but like I said, the main precipitating event was the 2019 effort to roll out this digital ID project in Kenya, which happened under a shroud of threats and intimidation. Everybody in the country was given 30 days to show up at their local chief’s office and register for this digital ID—everybody over the age of six—and if you didn’t register for the digital ID you could not get a passport, you couldn’t get a birth certificate, you couldn’t get a death certificate. Basically, your civic life was supposed to cease to function.

In the context of my advocacy about the rollout of the project, we kept running up against this wall because the vast majority of Kenyans do not speak English as a primary language. We learn English in school, and it is an official language, but it’s not the primary language of communication. And I wouldn’t even say Kiswahili is the primary language of communication for most Kenyans. Kiswahili is the language of interoperability. It is the language most of us speak when we have to work across ethnic lines, regional lines, official lines, and things like that, so it is the language of interoperability. But most people are probably more at home juggling in between all three languages. Because of the inability to translate them properly, you find that

there is usually a preference for one over the other depending on where the person finds themselves.

With that precipitating event in mind, an idea started to take shape. Why not create a project to provide the translations for some of these key concepts: surveillance, privacy, data protection, informed consent; none of these terms had translations in Kiswahili. It is important to keep in mind that Kiswahili is the most widely spoken African language in the world. It is spoken in at least 11 countries, it is an official language in two, and it’s being taught in many more countries. More importantly, because it is part of the largest language family on the continent, it is also a foundational language for people who are trying to do other work in other languages. In other words, it is much easier to translate from Kiswahili into, say, Kikuyu or Banyala, or any other Bantu language. And it’s easier to translate from English into Kikuyu, or any other language, because the grammatical structure is similar—the underlying philosophies of the language, linguistic philosophies, are similar. And then, there was this long year of grappling.

We had another digital rights catastrophe with the census and the idea that the census was mandatory, that you have to provide a Hiduma Namba, which was in this digital ID project that was being rolled out under duress, with no data protection law, but also GPS coordinates for everybody who is participating in the census, which of course is illogical because the census is supposed to function for enumeration, it’s not supposed to function for identification.

It's not supposed to pinpoint your location to very specific coordinates. But they're not really knowing how to start, because this is another piece that's really important in a digital rights context.

A lot of the discourse is dominated either by people who work in tech or people who work in law. Either one of these spaces is responsive to the questions of language. And here, I do not mean language as a tool or a technical thing that is being deployed in building better technology, but as a social thing, as a thing that people build through social interactions and political interactions. I feel that the language question was something that needed the input of social scientists, of people who work in humanities, but also people who work in the creative arts. People who are making things creatively. And so [I was] wanting to do this project for many, many months—many, many years really—and not really seeing a way of moving forward given that it did not exist within these two domains that I had already been straddling as my practitioner. That's where the Digital Civil Society Fellowship came in, because they provided the funding to be able to do this creative work and this linguistic work that I feel is really important to opening up the space into tech for thinking about digital rights, not as a technical thing, but as a social thing, as a thing that people build through their participation and through their interactions in the social world.

It has been about, as I said, a year and a half. It started off with workshops with Kiswahili professors from different universities in Kenya and in Tanzania. This is the joy of

Kiswahili. There are multiple ways of creating words in Kiswahili. In some cases, it was building on the foundation of a pre-existing word. For “surveillance,” there was no word for “surveillance” in Kiswahili. Before we did this project we used the foundational word *kudoea*, which the closest translation is “to covet” or to look at something that someone has and desires. So it is like to covet someone's information. And we used that as a baseline and created the word *udokezi*, which is now the word that we offer for “surveillance.” Sometimes there is this thing that Kiswahili does, which is called *utohozi*, and *utohozi* basically means “stealing,” but English does this all the time, it is fine. You take the word from another language and use it to kind of strategically break it in specific places and in many cases, because people would have been familiar with the word, *utohozi* makes more sense than introducing a completely new word, because, as I implied, the goal was to create a list of words that people would use. It was not meant to be the fanciest word or to have the most elaborate word, it was to make things that people would use.

For example, *algoridhimu*, which is an “algorithm”; we basically broke it strategically and changed the pronunciation in specific places so it would still sound familiar, but it would be a Kiswahili word.

And then in some cases we had nothing to start from and we had to start from scratch, or someone else had started from scratch and we basically just used it to build the word for “program,” *nyavuti* where that one had already been established in grammar in Tanzania, but had not had wide use.

So those are the three basic categories we put together and we came up with these amazing—if I do say so myself—flashcards. One side is the English word and the other is the Kiswahili word. We also made them into playing cards. This is where the joy of working with artists comes in because we wanted to make something people would use. So each card is also a playing card because there are 54 words. And we made these flashcards and have been distributing them for free, specifically targeting young people, because African populations are generally very young. Kenya is not even the youngest country in Africa. It is not even the youngest country in East Africa. Sixty percent of Kenya's population is below the age of 30, and all of these people grew up in a world where they don't know what it's like to live in a world without internet. And yet digital rights education has not kept up with that. So the other piece of the pie is obviously raising awareness and disseminating these cards as widely as possible.

The point that I wanted to leave you with is this: a lot of the digital rights aspects that affect tech deployment in Kenya are not technical issues; they are political and social issues that can be resolved by specific social innovations that target awareness, education, and inclusivity.

As a general example, right now the current Kenyan administration has made “tech first” a central pillar of their policymaking. If you want a driver's license, if you want to register for taxes or for a passport, you have to do it online. None of those websites are currently translated into Kiswahili. Not even one, even though all of those websites

are integral to living civic life! If you have to register for any of these services, you have to use these websites.

More importantly, only 12 percent of Kenyans use personal computers. So if you are a person who has to use those websites, most often you have to go to a cybercafe. Which means you have to pay for transport to the cybercafe, you have to pay a cybercafe attendant to do all of this inputting for you, and so on. Basically there is a tax on people who are not able to access these services. Poor people are paying a tax to access government services in a language that isn't their primary language. It is the language of business, but it is not the language of interoperability. So there are all of these silent rights issues that are very easily articulated as social issues and not technical issues, but as long as we keep focusing on decolonization of language as a technical thing that needs to be solved at a technical level, then we will keep leaving people behind. And it ends up being an insider conversation between people who already have power. What we want to do is democratize the space. Decolonization is an active process of taking back power from where it was stolen, from where it was disrupted. Thank you.

KATHLEEN SIMINYU:

Today I'll be speaking about building for Kiswahili on [Mozilla's Common Voice](#). During this presentation I'll be giving you an overview of my team's work and how we're thinking about different aspects of our work.

I am currently a Machine Learning Fellow at the Mozilla Foundation. Broadly, I wear two hats: the first is community organizer and the second is NLP practitioner. Let me tell you more about the Common Voice Project and my involvement in it.

What is Common Voice? The project launched in June 2017. It is Mozilla's project to build open and publicly available data sets of labeled audio that anyone can use to train voice-enabled applications. It is part of Mozilla's efforts to help teach machines how real people speak. It's a bet that we're making about the future of human-machine interaction, and it's a project that is making voice recognition open and accessible to everyone.

Common Voice is about collecting voice data in a safe and ethical way that can be used to train speech recognition algorithms. With respect to Kiswahili, the Mozilla Foundation embarked on a three-year funded project with a focus to build out language data corpora in East Africa. The core work is aimed at building skills and capabilities, both within the Foundation and within the region, to democratize voice tech through community engagement and contribution. This work is also supporting use-case development within agriculture and finance.

For Kiswahili in the latest release of the data sets, we already have 655 hours of data that is available. Our target is to get this data to between 1000-2000 hours of data in the coming year. Cognizant of the fact that voice recognition systems have been biased with poor performance on women, as well as older populations, we are

working to ensure that these demographics, which are likely to be under-represented in the corpus, are intentionally included, especially at the points of data collection.

There is a slight disparity between men and women, so we're working to build a diverse and inclusive voice community that is accessible to all, taking into account structural issues which impact gender inequality. This may be location, sexual orientation, education, region, and accent. Additionally, the two areas in which we chose to focus on are agriculture and finance, which also heavily impact women, seeing as the majority of farmers in local communities are women. Leaving them out would be leaving out the people we aim to impact with this work.

We have a Gender Action Plan that guides how we foresee inclusion across the gender spectrum. The plan takes into account the challenges that hinder inclusion, what measures and actions we can take to address them, and also how we can track through indicators and targets if we are attaining gender parity. This is work that is led by my colleague Rebecca Ryakitimbo, and it embodies a participatory approach from ideation to data collection to use-case development all the way to model creation and development of end-user applications.

Additionally, we have worked with linguists and language experts, and you may be asking yourself, why do we need linguists? Or you may not, as Nanjala was a good segway to my part of the work. It's to make voice technology sound more like us. Speech recognition systems have been found to be biased or to fall short

in scenarios where the source data is not balanced. And balance implies having equal quantities of data in different categories which are represented in the data. So, again, research has shown that speech recognition is more accurate for men than it is for women and more accurate for individuals younger than 30 years of age than those older than 30 years of age. Some other characteristics of data that may present bias in downstream applications, if not balanced, are accents, dialects, and variations of languages. To avoid a situation where one application developed for one locale performs considerably worse in another locale (which may actually be only a couple of hundred kilometers away), we need language experts to help us identify what these nuanced differences are and to help us ensure we are intentional about representation in the overall Kiswahili data sets.

What we know today as standardized Kiswahili has a controversial history. You'd be surprised to learn that among the Swahili people there is some level of dissociation. This is because locals were not consulted or involved in the process of selecting what dialect to standardize.

What we know as Kiswahili today originated from a dialect known as Kiunguja. And there are 23 known Kiswahili dialects, 13 of which are more widely used than others. So when it came to the decision about which dialects should be standardized, the decision was made exclusively by missionaries. It came down to a debate with different missionary groups arguing in favor of the dialects which were spoken in the locales in which

they were stationed. So it was based on what the missionaries were most familiar with and had little to do with the people themselves or the languages.

Then with the decision to standardize Kiunguja came the realization that it was not linguistically rich enough and needed supplementary vocabulary. And once again at this stage the Swahili people were not involved, which resulted in a growing distance between Kiunguja, the dialect selected for standardization, and the Kiswahili which emerged after the process of standardization.

Then with standard Kiswahili came efforts to drive its propagation by ensuring it was the language taught and used in schools. This gave rise to linguistic insecurity and contributed to the continued massacre of Kiswahili and of dialects related to Kiswahili.

Linguistic insecurity is the negative self image of a speaker regarding his or her own speech variety or language. It may happen if a speaker compares his or her phonetic and syntactic characteristics of speech with those characteristics of what is perceived to be correct, and unfortunately this is exactly what the native Swahili people are subjected to in school and in other formal settings where they interact with those of us who learn and use the now standard Kiswahili. They are repeatedly corrected, made to believe what they speak at home is incorrect, but the truth is much more complex than that.

Unfortunately, the result is that there naturally occurs a drift towards that which we are made to believe is correct, and away

from the diversity that is present among the native Swahili people. There are presently Kiswahili dialects on the brink of extinction and others that are falling away from use every day.

I am going to conclude by saying that from linguists and language experts, we learn that building in isolation as technologists, as developers, or even as NLP researchers is not the right thing to do. We learn that even if our main intention is to build a Kiswahili data set on Common Voice and to make it a publicly available resource that can benefit Kiswahili speaking populations, this will not inherently happen if we don't take the time to understand aspects of the history of the language and of the people themselves. We learn that if we proceed without conscious consideration, we risk alienating some of the populations that this research should benefit.

Our time with the linguists involved working to identify the dominant dialects or variants of Kiswahili that are most widely used at present. We then worked with them and had them do some fieldwork which serves to develop texts that are reflective of the variants and dialects which we identified as prevalent.

In comparison to the work that is being done for the wider Kiswahili data sets, which will largely be based on standard Kiswahili, these subsets will be significantly smaller. But our intention is to have the text and audio collected from the respective communities be subsets of the whole. These subsets will have two main purposes: the first is to help us quantitatively evaluate how our models and downstream applications

perform on related dialects and variants. We would like to work towards models with equal performance across various variants and dialect speakers, not forgetting their gender and age aspects as well. Our first step will be figuring out if there is indeed degraded performance for the different groups. The second purpose is in the event that the performance is degraded for the different demographics, we would like to make resources available for developers so that depending on the particular local context that they are building applications for, they will be able to fine tune, so as to improve performance if necessary. Thank you very much.

SABELO MHLAMBI:

Thank you so much Toussaint, Kathleen, and Nanjala for your comments. I am glad to join you all here.

The project that I have been working on, especially since being at Stanford, is this idea of decolonizing language from an ethical, decolonial, and technical perspective. Language is how we access the world; language contains worldviews. What happens when we no longer have the power to be in charge of our own languages?

I like to start with this story from around the 1800s. When the missionaries first came to southern Africa, they introduced an alphabet based on the Latin script. Their primary purpose for this alphabet was to make it easier and cheaper to proselytize Christianity to the African

masses. As it probably was cheaper to print more bibles and religious literature than to send missionaries over by sea, it made sense. But now when actually studying these alphabets for the Nguni languages, for example, in South Africa, we find that it takes you almost three times as long to read an isiZulu sentence or paragraph than it does an English paragraph. Why? Because these alphabets were not designed to meet needs. They were designed to spread one objective, which was “How do we quickly convert people, the fastest way possible, and the cheapest way possible?”

Bantu languages usually have a high morphology, where morphemes (prefixes, suffixes etc..) are compounded to create a word. There has been research where they have tracked the eyeball movements of when people are reading text to see how long it takes to read a Xhosa paragraph versus an English paragraph, and it takes more than twice as long to read the same paragraph in Xhosa than in English due to the inefficiencies of the Latin script when applied to Bantu languages. So the Latin script is not even practical to communicate in Bantu languages

The written Zulu script, for example, is impractical, and people often use shorthand on social media, in order to quickly write the language..

Similarly to the colonial era, we are now relying on big foreign tech companies, for example, to be in charge of shaping the mechanisms and ways of how we communicate. Foreign companies are the ones who are going to be creating our technology for using our African languages,

and this is the same process again where we are divesting this power, allowing others to create how we view and see the world. I find that really difficult.

Here is a quick example. We have here a Zulu word or phrase: *umuntu wabantu*, which means “a human being” or “a person of the people.” But the Google translation says, “a man of the people.” Wait, wait—nobody said anything about “a man”! We said “a person of the people.” Why does it automatically assume that it is “a man of the people”? That is not what we are saying. This is just one example of how when we have systems designed by others, they can shape our worldviews.

But why can't we be the ones to create our own technologies and have access to this? Another threat we have with these systems is surveillance. Nanjala spoke about surveillance. The more insight that we give about our languages, the better that these companies are able to surveil us or misuse our information.

And so there is another ongoing debate on how do we use our language as a way to tactically protect ourselves against surveillance or even from foreign worldviews that might be in conflict with how we think about and see the world? I began to think about this a lot, and in the early days of my fellowship I ended up thinking more about how most Africans access the internet or technology. It is through a cell phone, but that is not quite specific enough. It is actually through a keyboard. It is the entryway to the digital world for many people, and even in that domain, we do not have good keyboards

that support African languages. And it's simple tools such as spell correction. You cannot even find this for the Zulu language, the second largest language in South Africa and the second biggest economy in the African continent. If such a language cannot even have simple, basic resources, when will the thousands of other languages that exist in the African continent have a strong economy supporting them?

So it's quite a difficult thing. If people have to change who they are in order to access technology, in order to access the digital world and to learn another language, then how can they further participate in this digital world that we live in?

I began to also then look at technical interventions to try and address what is considered a decolonial issue. One of the things that we did is we made a keyboard, an African keyboard, where you are able to use GIFs. We have the largest collection of African GIFs and thousands of users are using the app. We have word predictions, correct spelling, and whatever you want to give people a way to naturally communicate in their own culture. In addition to the keyboard you have a collection of GIFs. Many of them are hand-crafted, selected, and created to give true authentic voices across digital platforms.

For example, if you went to any of these Giphy websites and you type in "Africa," you immediately start seeing things like a savanna or animals, because you cannot have "Africa" without any animals! This is the general thinking. Again, how can we take advantage or possession of our visual languages and how we want to express

ourselves? This is just one way we can think about doing it differently, so I have been curating this list, where we can take advantage of that and create our own systems.

I would like to end with this warning, yet again, that if we do not take heed of the lessons of the past, if we do not take advantage of our own languages and are able to preserve them in order to sustain a piece of our worldview or our values, and if we are not able to transform how we think about even writing on a piece of paper or a keyboard, we are more likely to remain behind a digital gap that is constantly expanding, especially as language increasingly becomes how we access technology.

This is something that has not been widely studied and looked at, but I am thankful to be able to pursue this work through my fellowship at Stanford and the DCSL.

CLAUDIA MAGALLANES BLANCO:

I am very happy to be here and to share these ideas with you. I want to start with a few key concepts like: coloniality of power, which refers to the interrelations of modern forms of exploitation and domination; coloniality of knowledge, which is the role of epistemology and knowledge production tasks in reproducing colonial thought regimes; and coloniality of being, which is a lived experience of colonization and its impact on language as the mechanism where knowledge is inscribed.

The lived experience of racialized subjects is deeply marked by language and how it expresses the world as a battlefield where the colonized subjects are defeated. The knowledge that counts is the Western Eurocentric, Christian, White, masculine, and speaking the language of the metropolis. Knowledge cannot be expressed in any language other than that of power, because the “others” do not think, therefore, they are not. And how do we know that they do not think? Because they could not write. They could not delve into the writing world.

So the prevailing language is always in line with what power determines as such, and in conformity with the dominant linguistic policies. Language is therefore the support of representations and can become the subject of struggles for its influence.

I want to introduce you to Yásnaya Aguilar, who is an Ayuujk Indigenous woman and a linguist. The phrase here says, “Our languages do not die alone, our languages are being killed.” She considers that language does not need a space to occur, rather it makes everything happen. There is a very strong interdependence between language as a cognitive territory and the defense of the geographic, symbolic, material, and spiritual territory. This reflects a relationship between autonomy, territory, and language. So the linguist permeates everything. It is not only an instrument of communication.

Attacking the language is an action of dispossession of the instrument of thought through the systemic violations of linguistic rights. I will come back to Yásnaya later on, but first I want to introduce the concept of coloniality of language, which

is a process of linguistic racialization that strips populations of their humanity by ascribing them to a natural communicative and mental inferiority. The languages of the colonized are not languages because they are not capable of expressing the political organization of a territory. They are not subsidiaries of the language of “truth,” which were Latin, Greek, or Hebrew. They're not written in letters and are not constituted by the relationship between writing and civility. All those imaginaries of the colonizers upon the languages of the colonized make them non-humans because they did not have that linguistic capacity. The languages and linguistic practices of colonized populations are forms incapable of expressing the ideas that the colonizers imagine as integral to be fully human. So in the colonial situation, the presupposition is that the colonized cannot naturally be interlocutors.

Now I want to move from these frameworks of the coloniality of languages to the presence of Indigenous languages online. We're facing a new form of imperialism, where there are only a few hegemonic languages and there is a pragmatic dominance of English in the online world. Exclusion in the capitalist context is not only not having access to the material goods, but also not having access to the language that is recognized as having the highest prestige. That is why the presence of Indigenous languages online is very important.

Yásnaya Aguilar says, “The future of Indigenous languages is also in cyberspace. And we have to start building it so that it becomes a more just and balanced place for the speakers of the languages of the world.”

There are currently more and more pages on websites, apps, and all kinds of digital media that are using Indigenous languages and putting Indigenous languages online, creating awareness for not only the people who speak them, but for the people who do not speak those languages and to know that they exist and that they are present.

I want to bring attention to Sabelo's point about the communicative languages for self representation, which is a notion of Alondra Barba Ramirez. Although many native Indigenous peoples had forms of graphic expression before the colonization process, these forms lost the possibility of existing in the face of Spanish writing and examination on the spheres of production and state administration.

So the grammar, the norms, consolidate a language that moves away from use, from life, to put itself at the service of commercial and political development. For example, there is a clear notion that the textiles told stories. It was a grammar on the textiles and there was reading and writing in this language that was not acknowledged, because it was not the written alphabet.

Alondra Barba Ramirez actually talks about how these cultural and community languages for self representation use different communicative discourses to make known different dimensions of Indigenous peoples' lives.

What are these communicative languages? Well, the Indigenous language, both oral and written, but also ceremonies, pilgrimages, sacred sites, dances,

traditional music, offerings, and ceremonial objects.

The territory, the sounds of the territories, the data is represented in diverse elements and beings of nature, the clothing, the arts and crafts, all of these are languages that are present online. They are present in digital social networks through pages and profiles: Facebook, YouTube, Instagram, WhatsApp, TikTok, [and] Twitter.

Indigenous peoples in Mexico are using photographs, music videos, reports, posters, audios, and memes. I have a photo here of a person who won a poetry-slam in an Indigenous language and then you can find all these videos of Indigenous languages' poetry-slam. Images like these, which are images of a sacred ceremony in a sacred site, or this Facebook page called, "Let's talk about Hikuri" [*Hablemos de Hikuri*] which is a Wixárika name for peyote, and is a sacred element of the Wixárika culture. So the Indigenous languages that are present online are not only written language forms, they are all these other forms of languages that come from the cosmology or the cosmovision of Indigenous peoples.

I want to finish with this notion, "Tequiology," which was coined by Yásnaya Elena Aguilar Gil. She considers that capitalistic technology is developed on the blood of many Indigenous and Afro-descendent peoples, which we certainly know for sure. The plundering of territories and bodies for technological development is very much a fact, in addition to recognizing that the racialized individuals whose territories and bodies are plundered

are not able to fully participate in that technological development.

It is important to think about the production of a technology that does not operate from an extractivist point of view. Taking into account the collaborative work carried out through mutual support, which is a very important part of Indigenous cultures and which represents communal work, Yásnaya considers Tequiology a form of collaborative technology.

Indigenous communities and free software both operate on a communal basis of mutual support, Yásnaya says. An example of this project is Mozilla Mexico and a browser called Mozilla Native, which allows any person speaking an Indigenous language to voluntarily take part in translating digital content. Each member can put in the hours that he or she has time to dedicate to the online project, translating different words that you can find on that server. This is a modest proposal to save the world—this is Tequiology, the common construction of a mutual internet or technology.

There are many people in the world who speak different languages, and those languages are not only in the written form or oral form, they have many, many layers. And the internet would have to reflect that. We have to aspire to net neutrality, which is complicated in the internet that we have, but we also need to have more balance and fairness on the internet in general. This is one particular proposal that I would like to put on the table.

NICK COULDRY:

It is an honor to have the chance to say something after these brilliant perspectives. Let me speak with all humility as someone whose native language is English. I cannot separate my capacity to speak to a tool from the colonial history of English's domination as a language, so I won't try.

Let me start with one small comment about the role of language in the Tierra Común network which is co-sponsoring this event, and which I co-founded with Paola Ricaurte and Ulises Ali Mejías. The goal of the network is to support perspectives on data colonialism from the Global South. From the beginning we were very aware of the issues of language. Most members of the network are from Latin America, where their first language is Spanish or Portuguese, not English. So we work in three languages. Some of our meetings are in Spanish, some of our activities are in Spanish or Portuguese, but of course this is an ongoing battle to fight against the domination of English as a language and its deep colonial legacy. It is a battle that is very hard to make progress on but that is part of what we try to do.

We have heard some fascinating perspectives on how deep patterns of language exclusion go right to the core of machine learning. Sabelo talked about how easy it is to lose power over one's language, and Claudia talked about the coloniality of language. Let me try to add one final point, which perhaps underlies all these different, very important perspectives.

As I see it, the struggles that we have heard described for us are not just with the dominance of English as a language, although of course that is central to everything. They are also struggles with the forces that structure the internet and AI in such a way that they specifically amplify all those colonial and historical inequalities, around language and many other things.

Think for a moment about why we are here today. It is because underlying many, many different and valid perspectives, there is a massive concentration of power in the world. A power over the design of the internet and this infrastructure. A power over the design and the resourcing of AI processing. A power over the resources to moderate languages, which exists in some places, and is funded much more than in others. A power over the development of the business models which drive the very possibility of the internet. A power over the economics of the platforms which enable us to be in touch with each other.

And as Nanjala brought up so beautifully, a power over the basic operating languages we use to describe what we are doing when we build the internet. All this work does something very obvious which I think it is important to specify. It reinforces the power of a primarily English-speaking elite located in the Global North.

So I want to ask whether it is possible for us to add an additional layer to the struggle which the other speakers have articulated so beautifully, which is the wider struggle only just beginning today: to build a differently configured space, a different universe of connection online, which

does not build in those inequalities on a daily basis. Where the power of control, monitoring, processing, and designing every layer of the internet stack is shifted much closer to local communities, on whatever scale we want to define them, communities who therefore will have a chance to use their own language in those acts of designing, monitoring, and so on. At the moment we are so far away from that possibility because of the profound dislocation of the space where the power over the internet exists and the space where most of the world's population happen to live.

We need a vision of what that larger struggle would look like, which would connect all these different struggles. Of course there is important work going on to decenter the internet's infrastructure, the work of the fediverse, mastodon, and various other approaches, but I think just by framing the larger struggle in this way it becomes very obvious we are a very, very long way from having that common struggle listed as a priority. Maybe that is what we also have to do, alongside bringing out the enormous violence that is recorded in the different struggles which the previous speakers have talked about. So that is one further idea I wanted to add. It is an honor to be part of this panel. Thank you.

NANJALA NYABOLA:

I wanted to build on something that Kathleen said that also came up in the conversation about our work with Kiswahili. We often forget that there is also an

element of imperialism in a lot of languages of the Global South. Amharic, for example, has a very unique history in Ethiopia. For a lot of people it is an imperial language, even though it was not imperialist in terms of coming from outside places like Europe or Italy.

It is the same for Kiswahili. Kiswahili has a very unusual political position and there is a lot of subconscious resistance to the use of Kiswahili in many contexts in Kenya, which is why we have this patois Sheng, which is actually more popular. It is the lingua franca but it is the lingua franca that is not recognized by the state or official systems.

I wanted to put that point on the table because I feel like part of the colonial energy that exists around linguistic practices is the failure to see that languages of the Global South can also be complicated. They can have complicated political and social histories.

It is important to acknowledge these social histories in the efforts of trying to make the whole process work. There were certainly a lot of interesting conversations around the choice to use Kiunguja versus Kibajuni versus any of the other dialects that exist, and we have had a lot of pushback from a number of people who hold the same perspective. They said things like, “Why are you using this? Why don't you do it in Sheng instead of doing it in Swahili? Why did you do it in this dialect instead of the other one?” There is a lot of political context that is embedded. I think giving – let me speak for African languages, but I know this is the case in Latin America as well – giving languages of the Global South

the opportunity to be complicated, to be politically and socially complicated, and not pretending that English is the only neutral language in the world, is also part of the decolonization process.

CLAUDIA MAGALLANES BLANCO:

I completely agree with Nanjala. I think we need to stop romanticizing Latin America Indigenous languages and placing them in a pristine environment. They are also complex and political, and there are power struggles within different Indigenous languages and the different dialectic forms of one particular language, as Nanjala and Kathleen already explained in the cases of Africa. It is the same in Latin America. You have Nahuatl, and you have several variants, or you have Zapoteco, or Ayuujk, with several variants, and there is a prominence of certain groups over other groups based on linguistic differences.

So it is important to promote Indigenous languages and to have them mirror the real world in the online world. But this real world is not a magical, enchanting, all peaceful world – it is very complex and contentious and difficult, and that makes it much more rich and interesting. All too often, in the online world, when you see an Indigenous language, such as in the case of Latin America, or a dialect in the case of the aboriginal in Australia, it has to be “perfect,” because you know, it's seen as “folkloric.” But we need to put the political dynamic at the core to actually have the conversation.

TOUSSAINT NOTHIAS:

Thank you, Claudia and everyone else. I am going to ask a big picture question to the whole panel, and everyone should feel free to respond. I am going to make it a bit schematic for the purpose of discussion. Many examples you shared today support advocacy for more language inclusion, such as creating more representative voice recognition systems or more localized keyboards. But I wonder if in some ways that does not also support the idea of data expansion across the world. I wonder if you have any reactions to that tension between advocating for more diversity in languages and pushing technology further in all spaces of social life.

KATHLEEN SIMINYU:

I would like to highlight the case of the Maori. This is something I was reading about a couple of weeks, or maybe months, ago. They have been building audio resources for the language in an attempt to preserve it. So aim number one is language preservation. But then the bigger that data repository grows, the more they receive interest from big tech and other companies interested in profits.

Basically, these companies are interested in exploiting that dataset and building speech recognition technologies. Given that their main intention was language preservation, I feel that in this case they were very clear about the ownership of the data being theirs, as the language is their own right.

I feel that as we encourage more languages to be made accessible, or more language data repositories to exist, we should also educate the language speakers on ownership and what that means or can mean when we start building language datasets. Ownership can mean different things. It can mean a share of the returns if/when the data is used to make a profit. It could also mean having the right to speak to which use cases the communities involved would like to be built versus which ones they are against because it may cause harm or have some other undesirable outcome or effect on them.

At the end of the day, if the language that you speak is used to build an application in finance, it means that you and that language community then have access to those resources. But then, if it is used to build something harmful, it means that you and your language community are the objects of that harm. In making language data more available, it is an opportunity to have a much wider conversation about data governance and how we can start to explain or view data as more belonging to the owners or subjects of the data and giving them the power to decide what is okay, what is not okay, and giving them the rights to opt out. Because we all do not have to live in hyper-connected, AI-powered worlds where, you know, my voice assistant does everything.

NICK COULDRY:

It is a very difficult question you have asked because there are so many layers to the power over the internet, and to what the internet is, so you are right. The idea that

we should all be connected in a certain way, and that machine learning is going to expand our knowledge of everything, is itself part of the ideology of big data, which itself is part of the very colonial systems that need to be challenged. On the other hand, the technology itself for voice recognition, the possibility that machines could speed up the translation of languages to, say, enable mutual translation, this is an open possibility that we could imagine being used in very, very different power structures and that would be genuinely positive. I don't think there is any reason politically to rule out in advance the use of particular technologies. What we need is to open up a world where we can see the differences those technologies can make under radically different power distributions, where communities actually have the ability to influence their uses and their design. That is very far from what we have at the moment, but all of the projects we have heard from are trying to find ways of doing that. So I see it as a techno-social approach, with the emphasis on the social.

SABELO MHLAMBI:

I think the notion of opening up data and languages to the whole world is pretty dangerous in a way, because language is also secret and intimate. There is no way of telling how our languages will be misused once they are in the hands of big corporations.

Even if they decide to use the languages for our “own good,” to better connect us, we’re excluded from the profits that come from

creating those services—we are excluded from the final products of those services. We are not getting the finished goods themselves. Even if we do benefit, we are not receiving the bulk of the economic benefit. And so I think one view is to figure out how do we actually take charge of our data, protect it, and make sure that we benefit, as Africans, or Latin Americans, from our own cultural heritage? We created these languages, they represent our most intimate ideas and thoughts. How do we benefit from them, even when other companies want to “help us,” so they say? How do we make sure that they can assist us in a way where we retain the majority and vast control of the economic power and direction of these technologies?

TOUSSAINT NOTHIAS:

Thank you, Cathleen, Nick and Sabelo. Let’s turn to a question from the chat, which is related to this last point: *“What tech companies have been most open and genuinely interested in collaborating with you to learn more about your research on coloniality and decolonization and taking action? What companies do you see as being most influential in the space in the future?”*

NANJALA NYABOLA:

I want to make sure I answer this properly. Obviously everyone is going to say Mozilla, because I think Mozilla has done the most (especially in the Kiswahili language space)

to reach out to people. One of the things that I deliberately did with the **Kiswahili Digital Rights Project**, is make everything free. As far as it can be given, it is given, and it is given to the people who will use it and it is given with no expectation of use.

I think in African countries you often get stuck between two things: private capital that is trying to flip whatever you are doing to increase their bottom line, and this speaks to the previous question. And two: you get stuck between nonprofits that are giving you money with the intent for you to turn whatever you are doing into some kind of advocacy campaign. Between these two extremes you end up with skewed output.

To give you an example on language, I went to a Catholic high school and during that time our teacher refused to teach us the Kiswahili word for sex. They would not teach it. It was as if it never happened, it did not exist, it is not a thing that you should know. You go through school and you come out of your education thinking that Kiswahili does not have a word for sex. But, of course, it does. Language is use. Language is context. Language is power. Language is intent. So getting stuck between corporate and nonprofit impulses leaves many initiatives corrupted. We can describe development related things in great detail in many African languages—*maendeleo* [“development”], for example, and we can describe it in commercial terms, but we can't really describe it in social terms, giving people words that they will use on a day-to-day basis.

As we think about trying to identify companies, institutions, and supporters of

this work, it is important to fundamentally think about open sourcing, accessibility, and how to conveniently package things to give people in order to meet them where they are. Because ultimately, unless you break out of that paradigm, you're engendering a different power disparity. It might not be the disparity that you were addressing, but it ends up becoming a different power disparity, and then partners come in. I make it clear to partners or interested parties that no one is making any money out of this. All of this has to be as fair as possible, as free as possible, and with no expectation of use, with no expectation that you are then going to fill in a 50 page report and tell me how many words you cited in this particular instance. The developmental paradox that plagues a lot of initiatives in this part of the world—being stuck between these two extremes—tampers a lot of really important work because it stops being about organic use and organic cooperation and becomes about representations of certain ideas of progress.

CLAUDIA MAGALLANES BLANCO:

In Mexico there is the Indigenous Mobile Community Network, which is a communal Indigenous network for mobile services. And the reason why I am bringing this up in relation to the question is because one of the major debates that they are still having is not only offering the mobile service for calls or texting, but about opening up data on the mobile network to YouTube and social networks, and one big aspect of that debate was language.

What is going to happen with our language if we open it up? Teens already have mobile phones and when they go to the places where they have connectivity to the major corporate networks, they use their Facebook accounts and their YouTube accounts. But when they are back in the community, they still speak their Indigenous language. So if we are going to have these social networks inside the community, what will happen to the language? It is a very valid question and there is not an easy answer yet. But at least this Community Mobile Network that was developed by Indigenous communities is providing support and building the technology and antennas and so on.

It is not only a technical solution, it is also a social and community solution, where the community is saying, “What do we want? What kind of internet do we want to have access to in our community? What are the questions that we are concerned about in terms of language, but also in terms of surveillance and in terms of monitoring?” Many of these communities are actively fighting against extractivist projects like mining or hydroelectric. So what about security issues, if they go online? So there is a debate and that is what makes it a community, or a communal project — you have to have those questions. They can have the technology answer the first question of “How are we going to connect?” And then they can say, “Well, now we know how to connect but before we do, we want to connect in this way, and with what limitations, and taking into consideration the issues related to the everyday life of the community and its individuals.

TOUSSAINT NOTHIAS:

Thank you, Claudia. I am going to try to include two questions here from the Q&A. The first question is to Siminyu. *“How easy or difficult is it to make NLP that is accurate to various dialects of Swahili? What impact would this have on sustainability and energy consumption, i.e. in cost of more accurate NLP versus energy consumed to achieve this?”* And the second question is: *“I always ask myself, and I think it's also in Nick Couldry's book, [how] does the world handle the amount of data “pollution” that we want to put online?”*

KATHLEEN SIMINYU:

So first question, how easy is it to build speech recognition for the various Kiswahili dialects in use? I am not particularly sure. We have not gotten to the point where we actually trained the speech recognition models and started evaluating their performance. I will say, we realize that a lot of the work that we are doing now is going to impact the performance of the models that we develop. So stuff like whether the number of speakers that contribute are diverse enough. Are we only reaching out to women? Are we only reaching out to young people? It is going to matter whether we have older people represented in the data sets. It is going to matter whether we have people who have a Nairobi accent versus an accent that is likely from along the coast, or whether they have an accent

that speaks to the influence of French as a language that is also spoken by the speaker, or English, etc. So in terms of making it performant, we have faced challenges with having Kiswahili Sanifu as the dominantly represented language, we encounter many voice contributors who tell us that that is not how they actually communicate or interact in their day to day. We realize that if we want to build tools that people can speak to and use with ease, then we need to build tools that handle or generalize to more than only standardized Kiswahili.

And at the end of the day, it is really going to come down to how much diversity we are able to build into the data sets. Unfortunately, I cannot answer quantitatively, but I can say that we are trying to build as diverse a data set as possible with the resources that we do have at the moment.

In terms of energy efficiency of the model training or creation process—I hope I am getting that question right, unfortunately the trend in NLP, at least in the state of the art currently, is that everyone is building bigger and bigger models, and they want them to be much more multilingual. So you are throwing in language data without even looking at the quality of the data, because it has been shown that even messy data in copious amounts yields significant results. So unfortunately I am pretty sure that the work that we do contributes a lot to climate change, it is very energy intensive.

I do wish that we were thinking more about different ways of building NLP. Does the state of the art have to be that we are throwing in more data because that is definitely going to increase efficiency or

can we think of other ways to do it? That is something that repeatedly keeps coming up in conversations around this work, whether data sets from other languages that are closely related to Kiswahili can be of benefit. We have seen the use of pre-trained models and the use of transfer-learning – can we start to think about taking data from Kinyarwanda, which is also very well represented on the Common Voice data set, and see if pre-training that gives us an advantage in Kiswahili, versus, say, pre-training in English to Kiswahili which has also been shown to give great results? So we will see.

We are trying to think of different ways to do things. We do not have to continue the current culture of using bigger models and then expecting better results.

TOUSSAINT NOTHIAS:

We had a question early on: *“I love the framing of decolonization as re-intermediating power relationships, rather than technical issues. Given that many low income countries are committed to income growth and economic development, are there any other powerful ideas or people you found when thinking through decolonizing global capitalism?”* Big question! We are here for that.

NANJALA NYABOLA:

One of the things that I am really fascinated by in the decolonization process is thinking beyond money, and this ties again to the

previous response. Until we separate money and commercial interest from what we are trying to do we will keep getting stuck in this particular cycle.

What money does is create an incentive for competition and not incentives for collaboration. Everyone is trying to be the first, the best, the most, and usually with the intent of trying to get some money on the back end. So thinking creatively about why are we doing this, and taking money and commercial viability out of the equation, is one really important thing that we can do. Because that is really one of the powerful impulses that comes from the colonial process— making all of these worth determinations over people's lives and people's day-to-day experiences, primarily hinged on money.

One of my favorite Kenyan political thinkers, philosophers, and practitioners, Reverend Timothy Njoya, says, *“Kenya was colonized to be a market for raw materials and has never transitioned out of that process,”* which is why our decolonization is stalled. Many African countries' relations to the world are as a place of extraction where raw materials can be taken, packaged, processed somewhere else, and then eventually sold back to us.

I think we are heading in that direction when it comes to data and tech, as well, where we see countries of the Global South as primarily places of extraction and not primarily places where value exists on its own, even if it cannot necessarily be quantified in terms of money. And this ties into the question that is being discussed

in the chat about picking which dialect to preserve. Do you preserve the one that has commercial viability? Do you pick the one that is most likely to disappear in the next 50 years? The logic for me would be to preserve the one that is most likely to disappear. Even if it only has four speakers, protect the one that is not going to be here and then work backwards in relation to the one that needs the help the most. But that is not what we are doing. We are starting with the one that is spoken the most and then eventually we might get to the one that is spoken by four people.

And so it is really about the commercial interest—finding a way of removing that money as an incentive from the way we think about the work that needs to be done can help us think through what the shape of the decolonial work will be.

TOUSSAINT NOTHIAS:

Thank you, Nanjala. We are going to take one last question that covers a few of the others: *“How can one decolonize language or preserve a worldview without taking the colonialists' stance of being an arbiter or curator. How do you know when what you're doing is actually something that is desired by the community? And what happens when the community is large and contains conflicting views?”*

SABELO MHLAMBI:

Because I am in a position where I am deciding on behalf of some users, because I am creating knowledge – my biases are there. One way that I tried to confront that is to make sure that the work itself is aligned with a guiding philosophy, which in this case is the *ubuntu* framing, the *ubuntu* philosophy. If it is a well accepted way of thinking about the world or how things should be, I try to align myself with that as the first starting point and then I work with other groups as well. I share my research aligned with organizations on the continent to make sure that I am aligned with the work that others are doing. It is a way of collaborating with others. But I also think the most important thing is to understand the guiding philosophy itself and the guiding values that people treasure in that society, making sure you are aligned with that and then checking your biases each time you collaborate.

NICK COULDRY:

I would like to add something to that, because I think this audience member has asked a really good question. We are talking about fundamental power issues here and power normally gets reproduced, so how do we avoid that?

I think the key principle, building on what Sabelo just said, is that the decisions of design and what to design and why, and how to monitor it, need to be taken much closer to actual communities rather than in the abstractions that are big corporations.

I think a previous question asked about what theorists inspired us – I just want to mention one who has been very influential in Latin America who we discussed in a Tierra Comun reading group a few months ago, which was Ivan Illich, an Austrian who went to Mexico. He was writing 50 years ago and he was mainly attacking capitalism. He was not explicitly framing his critique in terms of colonialism, but his book *Tools for Conviviality* is still an amazing way of framing the question about how power and decision-making have to be close to actual social groups of people. As soon as you take it far, far away from groups of people, it becomes abstracted, it becomes violent, it divides, and it polarizes.

And it is such a brilliant way of capturing what has gone wrong with the version of the internet we have and which we have been trying to think about ways of correcting today. If we go back to any theories, I think *Tools of Conviviality* still has a lot to teach us, and it is available in many languages, including a beautiful early edition in Spanish.

CLAUDIA MAGALLANES BLANCO:

To bring some other theoretical inspiration to the table, I would also add the word of Brazilian Paulo Freire, who says you cannot decolonize something from the outside, it can only be decolonized from the inside. So it is not our work that decolonizes the languages, it is the people whose languages are colonized who are decolonizing the languages themselves.

And it is not something that has just started to happen. In the case of Latin America it has been happening for the past 500 years. The languages were colonized, but they were still in constant resistance and in constant challenging of the language of the colonizer. Otherwise we would not still have Indigenous languages. We would only have Spanish or Portuguese. The starting point, following Freire, is to raise awareness and consciousness about the mechanisms and the structure of colonization and how it operates on language so then through that you can find a way to actually move towards the decolonization of the language.

What I would say is that you cannot decolonize from the outside. All the efforts and all the things that I think we have been discussing here have a community basis. They spring from the peoples whose languages are being put online, or into a database, or being translated into technical language from their actual living language. You have to have it as a lived experience; something that you live and breathe every day. That is where you find the resistance point and that is how you move towards decolonization.

In many cases this has been happening, although people would probably not actually say, “Oh, I am decolonizing language.” But they are actually doing it, and then maybe we come along with a different lens and say “this is decolonization.” But, you know, it is not “decolonization” just when we say it is, but when it is actually happening.

TOUSSAINT NOTHIAS:

Thank you, Claudia, that is a great way to end! Thank you all for joining this webinar, and thank you to our speakers for sharing ideas, readings, prototypes, tools, and communities - and, most importantly, for their energy to reimagine our current digital ecosystem.

ABOUT THE SPEAKERS

Claudia Magallanes Blanco is Professor in the Department of Humanities at Universidad Iberoamericana in Puebla, Mexico. She holds a PhD in Humanities from University of Western Sydney, Australia. An academic and activist for social justice, she has been working with collectives and organizations concerned with community and indigenous communication for more than 15 years, as well as with contemporary social movements in Latin America. Together with Charlotte Ryan and Alice Mattoni, she is coeditor of the book series *Media and Communication Activism: The Empowerment Practices of Social Movements* edited by Routledge. She is co-author of the article *Digital Inclusion Across the Americas and the Caribbean* (“Social Inclusion” 2020: vol 8:2) and was a collaborator in the Chapter on Media and Communications edited by Nick Couldry and Clemencia Rodriguez in the report of the *International Panel on Social Progress*. Her research and publishing interests include indigenous knowledge and technology, social movements and digital sociocultural networks, cognitive justice and data justice.

Nick Couldry is a sociologist of media and culture. He is Professor of Media Communications and Social Theory at the London School of Economics and Political Science, and from 2017 has been a Faculty Associate at Harvard’s Berkman Klein Center for Internet and Society. In fall 2018 he was also a Visiting Professor at MIT. He jointly led, with Clemencia Rodriguez, the chapter on media and communications in the 22 chapter 2018 report of the International Panel on social Progress. He is the author or editor of fifteen books including *The Mediated Construction of Reality* (with Andreas Hepp, Polity, 2016), *Media, Society, World: Social Theory and Digital Media Practice* (Polity 2012) and *Why Voice Matters* (Sage 2010). His latest books are *The Costs of Connection* (co-authored with Ulises Mejias), *Media: Why It Matters* (Polity: 2019) and *Media, Voice, Space and Power* (Routledge 2020).

Sabelo Mhlambi is the founder of Bhala, an AI startup that democratizes the advances of AI to millions of Africans through Natural Language Processing of African languages and African visual languages. He is also the founder of Bantocracy a public interest organization that focuses on ubuntu ethics and technology. He has been a fellow at the Berkman-Klein Center for Internet & Society, a fellow at Stanford’s Digital Civil Society Lab, and a 2019-2020 Technology & Human Rights Fellow at the Carr Center

for Human Rights Policy. Mhlambi's work is at the intersection of human rights, ethics, culture, and technology and emphasizes global south perspectives in AI policy. In particular, Mr. Mhlambi's research examines the human rights implications of algorithmic technology on marginalized communities and proposes a new ethical framework for governing the creation and use of Artificial Intelligence in ways that maximize social progress and harmony.

Toussaint Nothias is the Associate Director of Research at Stanford University's Digital Civil Society Lab, a Research Scholar in the Stanford Center on Philanthropy and Civil Society (PACS), and a Faculty Affiliate of the Center for African Studies. He is a communication scholar working on journalism, civil society, and digital technologies. His research explores the multifaceted legacy of colonial power relations on contemporary media representations, journalistic practices, associational life, and digital networks. He has notably published articles in the *Journal of Communication*; *Boston Review*; *International Journal of Communication*; *Public Books*, and *Media, Culture, Society*. His current work explores the intersection of digital technologies and advocacy across African contexts, with a particular focus on one of the most notorious and controversial projects to increase connectivity across the Global South: Facebook's Free Basics project. In his role as research director at the Digital Civil Society Lab, Dr. Nothias leads the Lab's research agenda and core scholarly programs, including the postdoctoral program, courses, and scholar workshops and events.

Nanjala Nyabola is an independent writer and researcher based in Nairobi, Kenya. Her work focuses on the intersection between technology, media, and society. She holds a BA in African Studies and Political Science from the University of Birmingham, an MSc in African Studies and an MSc in Forced Migration, both from the University of Oxford, as well as a JD from Harvard Law School. She has held numerous research associate positions including with the Centre for International Governance Innovation (CIGI), the Overseas Development Institute (ODI), the Oxford Internet Institute (OII), and other institutions, while also working as a research lead for several projects on human rights broadly and digital rights specifically around the world. She has been published in several academic journals including the *African Security Review* and the *Women's Studies Quarterly*, and contributed to numerous edited collections. Nanjala also writes commentary for publications like The Nation, Al Jazeera, The Boston Review and others. She is the author of *Digital Democracy, Analogue Politics: How the Internet Era is Transforming Politics in Kenya* (Zed Books, 2018) and *Travelling While Black: Essays Inspired by a Life on the Move* (Hurst Books, 2020).

Kathleen Siminyu is an AI Researcher who has focused on Natural Language Processing for African Languages. She works at Mozilla Foundation as a Machine Learning Fellow to support the development of a Kiswahili Common Voice dataset and to build speech transcription models for end use cases in the agricultural and financial domains. In this role, she is keen to ensure the diversity of Kiswahili speakers, in terms of age, gender, accent and language variant/dialect, is catered for in the dataset and models created. In her NLP research, Kathleen has previously worked on speech transcription for Luhya languages and contributed to machine translation for Kenyan languages as part of Masakhane. Before joining Mozilla, Kathleen was Regional Coordinator of AI4D Africa, where she worked with ML and AI communities in Africa to run various programs. She has vast experience as a community organizer having co-organized the Nairobi Women in Machine Learning and Data Science community for three years and continues to organize as part of the committees of the Deep Learning Indaba and the Masakhane Research Foundation.

ABOUT THE DIGITAL CIVIL SOCIETY LAB

Digital technologies are transforming civil society and democracy. Our dependencies on digital systems require new insights into how these technologies work and how civil society can engage them safely, equitably, purposefully, and in support of human dignity. The Digital Civil Society Lab (DCSL) at Stanford University aims to understand and inform civil society in a digitally dependent world. Through fellowships, research, events, and teaching, we nurture an emerging generation of scholars, community advocates, technologists, and policymakers shaping the future of digital civil society.

DCSL is a research initiative of the Stanford's Center on Philanthropy and Civil Society (PACS).

ABOUT TIERRA COMÚN

Tierra Común brings together activists, citizens and scholars who want data to be decolonized. Our specific focus is Latin America, but our horizon is the Global South, and everyone anywhere who rejects data colonialism as the latest manifestation in modernity of the Global North's desire for domination. Imagination is our greatest tool. Let's imagine a future where the terrain of human life does not involve extraction of data that discriminates between us and separates us from our own lives.