[Draft Only]

It's Incomprehensible: On Machine Learning and Decoloniality

Abeba Birhane and Zeerak Talat

Abstract
As technologies begin to permeate society, people, and the society itself reorganize around the technology. It is therefore imperative that academic communities, civil society, and regulatory bodies address the impacts of a technology on society and human life. With machine learning, scholars, activists, and regulators have dedicated significant efforts towards quantifying and mitigating the expressions of social biases in machine learning models. Recently, researchers have argued that machine learning reproduces colonial logics and proposed decolonialization as an avenue for future machine learning efforts. Here we examine the aims of machine learning and decolonization, and argue that their goals, i.e., to abstract away and attend to detail and histories, respectively, are inherently in tension. Their origins, statistics and phrenology on one hand, and the liberation from marginalization and abstraction on the other, point in different directions. However, this tension can be resolved by situating machine learning within communities that can fill in detail that the technologies abstract away.

Keywords (6 keywords): Decolonization; Machine Learning; Afro-feminism; Phrenology; Colonization; Artificial Intelligence

Introduction

The movement to decolonizing AI has been gaining momentum over the past decade, with various positions on the matter. Some have proposed ways to decolonize technology (Alexander, 2020), while others have applied more strict criteria, highlighting the challenges of uprooting coloniality and raising the question of whether AI can be decolonized at all (Adams, 2021). Yet others have criticized the slapdash use of decoloniality. Paballo Chauke, for example, warns that "'decolonization' in Africa risks becoming a buzzword due to haphazard use by those who want to 'seem open-minded'" (Byrne, 2022). It is therefore important that we critically examine if and how decoloniality can be applied to machine learning (ML). Whether decolonization is possible is a particularly important question, as ML is rapidly being integrated into the social sphere, with deeply harmful consequences, as these technologies perpetuate historical injustices such as white supremacy and colonialism (Birhane et al., 2021; Talat, Lulz, et al., 2021). Due to such harmful outcomes, there has been a large-scale effort towards mitigation of harms, with proposals of technical and theoretical contributions, practical principles, guidelines, and protocols (Abdilla et al., 2020) for developing and deploying ML.

In this chapter we explore the two divergent threads of ML on one hand, and decoloniality on the other. We trace the roots of ML and explore decolonial work from and for the African continent. In exploring ML, we draw a line from 19th century phrenologist work to contemporary ML methods as applied to human and social data. We argue that through its goals of abstraction, ML risks constituting a modern form of phrenology. In exploring decoloniality, we discuss what it means to decolonize and situate our understanding in decolonial thought from and of the African continent. We reiterate the arguments that decoloniality and decolonization require attention to details and histories that are disconnected from the colonial gaze. Relying on these two divergent strands, what ML does and what it means to decolonize, we highlight the inherent tensions that exist between these two forces, such as their values, methodologies, and objectives. In light of these tensions, we reiterate Adams' (2021) question: Can ML be decolonial? We argue, that despite the tensions inherent to the meeting of these two fields, that decolonial ML is improbable yet not impossible. For ML and AI to be truly decolonial, it must fulfil key conditions. For instance, AI and ML initiatives must centre the needs of indigenous peoples and relegate financial profit to a secondary concern. We conclude the chapter by discussing the Te Hiku natural language processing (NLP) project. The project, through its focus on *Te Reo Māori,* and resistance to co-option by capital interests, situates ownership within the impacted communities and serves as an example for future decolonial ML efforts.

From Phrenology to Machine Learning

Phrenology is a pseudo-scientific field that focuses on the correlation of physiological features with personality traits and social attributes. Originally, phrenologists relied on statistics to make claims on social groups as they sought physiological traits to affirm the white Western European man's superiority over all other groups (Sekula, 1986). Although statisticians have mostly abandoned phrenology, we argue that applying current ML methods to the social sphere – such as datafying, sorting, predicting, and classifying social phenomena, (e.g., human behavior or emotion) – provides a space for a resurgence of phrenology. As such, the ML field risks repeating past transgressions by enacting tools that correlate social value with demographic groups.

Contemporary ML methods, like early phrenologist work, enshrine hegemony and construct the Western white man as the normative and exemplar. Meanwhile other groups are relegated to stigmatised positions. Although there is an increasing body of work in computer vision which seeks to directly correlate personal attributes with physiological traits (see for instance Agüera y Arcas et al. (2017) for a rebuttal to such work), we focus here on language.[1] We select language technology as our focus as language is co-constitutive with identity: how we express and construct our identity is inextricably linked with how we communicate (Di Paolo et al., 2018). Further, many of the issues that are readily apparent with computer vision technologies such as facial recognition and detection have an opaque mapping to language technologies. It is this mapping that we seek to expose through our use of language technology as our driving example Moreover, language affords a lens through which phrenology has progressed beyond its visual expression to expression through optimization technologies such as ML that link social value with speech. However,

[1] The relationship between computer vision and physiognomy is particularly clear in (proposed) tasks such as criminality detection from facial images, where the governing distinction from 19th century phrenologist work appears to be the use of optimization technologies.

the issue of phrenological patterns in ML extends beyond computer vision and language to the use of ML on social data.

From Whence We Came

Phrenology and physiognomy arose during the expansive colonial and empire-building projects of western European nations (Belden-Adams, 2020; Challis, 2014). The express goal of phrenology was to create schemes of classification through which the myth of the (biological) superiority of the Western European white man could be maintained (Belden-Adams, 2020). The creation of racial hierarchies and the physicality of purported inferiority were necessary to this end. Distinguishing between populations relied on the notion of “the average man” proposed by Adolphe Quetelet (Sekula, 1986). By seeking the statistical average of demographic groups, phrenologists such as Francis Galton and Alphonse Bertillon sought to use demographic divergence as keys to access or explain “unwanted” behaviors and characteristics. In attempting to lighten the burden of understanding, phrenologists linked physiological features with undesirable characteristics such as criminality to demographics. For instance, Bertillon's criminal archive project sought to develop a system that would allow for the re-identification of criminals by classifying individuals based on their features' relationship to the average. Although Galton used images in a similar fashion, he instead sought to dispel individual differences through composite photographs that were constructed from images of multiple individuals. Galton created and later overlaid the composite photograph by modulating the exposure that each image received as a fraction of the number of images comprising the composite photograph (Sekula, 1986). In this way, Galton sought to highlight the *average* intra-demographic features and the inter-demographic disagreements. Such agreements and disagreements were subsequently assigned to stereotypes about different demographic groups. By highlighting averages through composite photography, Galton specifically de-emphasized the intra-demographic patterns that would contradict the eugenicist positions that he held, i.e., the patterns that demonstrated inter-demographic agreement.

The Phrenology of Machine Learning

Although statistics has sought to distance itself from its phrenological pasts, we argue that ML for and on social and behavioral data constitutes a step towards such pasts by creating models that emphasize and de-emphasize similarities and differences between constructs in data. Thus, creating schemes which link demographic belonging to different classes. ML models are optimized by applying particular views over data. For instance, Convolutional Neural Networks (CNNs), a common algorithm for computer vision (e.g. Voulodimos et al., 2018) and NLP (e.g., Gambäck & Sikdar, 2017), provides a clear comparison with Galton's composite image.[2] CNNs identify patterns from data by passing input data through a convolutional layer that uses a window to compute a feature mapping. The feature map is subsequently condensed using pooling functions. As the model is optimized, the feature mapping and the weights of the model are refined on the basis of disagreements between

[2] We focus here on CNNs, but our argument extends to other ML algorithms that rely on the distributional hypothesis, by considering how a given model operates on data.

model predictions and the ‘ground truth’ (Goodfellow, et al., 2016).[3] In constructing convolutional layers as exposure of digital data, their similarities with Galton’s composite imaging become readily apparent.

Although the similarities between phrenologist and contemporary ML methods are hazardous, these methods are not always a cause for concern. For instance, in applying ML methods to astronomical data, data aggregations methods, such as convolutional feature mapping may be appropriate. It is the application of ML methods to social and human data that raises concern and comes to resemble the methods of early phrenologists, through constructing patterns in data as supposedly inherent attributes of objects or humans (Birhane & Guest, 2020; Spanton & Guest, 2022). For instance, identity attributes such as gender and race have been found to be recoverable in machine learning models and methods for language data without explicit signals of identity (e.g., Bolukbasi et al., 2016; Davidson et al., 2019).

In the case of classification, the weights and features of neural networks are optimized to minimize classification error (Talat, Blix, et al., 2022). Through the associations of features, weights, and classes, ML models create links between the classes and the social demographic features latent in the data. Often, the rationalization for ML is to lighten the burden of in-depth understanding in favor of shallowly casting attributes onto objects. For human data, the quest for linking attributes and data creates a deeply troubling likeness between phrenology and ML.

Much like the colonial hierarchical knowledge structures, in ML, a handful of elite, privileged, Western white men define and operationalize concepts such as “performance”, decide what needs to be optimized, what “acceptable performance” is, and how technologies are embedded in society (Birhane et al., 2021). Like those of the 19th century phrenologists, these decisions are made such that they reflect the interests, objectives, and perspectives of their creators.

Understanding Decoloniality

European powers executed colonialism for centuries using various strategies across the globe. What is referred to as traditional colonialism, or the control of indigenous resources, land and their people through physical and military coercion has largely ended for many nations decades ago. Yet, the remnants as well as its indirect influences -- often termed as coloniality -- remain intact and permeate day-to-day life (Ndlovu-Gatsheni, 2012; Tamale, 2020).

The impacts of colonialism and coloniality have drastically altered the course of history for most non-Western societies and indigenous populations, to the extent that it is challenging to conceptualize the histories of the colonized independent of colonization. Subsequently, there have been various movements towards *decolonizing* education (Barongo-Muweke, 2016; Battiste, 2013), legal systems (Currier, 2011; Kuwali, 2014, Oyěwùmí, 1997), languages

[3] The idea of a single ground truth which objects can be compared against relies on an illusion that an “objective”, context-less, history-less, perspective-less conception of the world can be constructed and reflected in data (Gitelman, 2013; Raji et al., 2021, Talat, Lulz, et al. 2021).

(Ngũgĩ wa Thiongʼo, 1986; Wa Thiong'o, 2018), technologies (Leslie, 2012; Raval, 2019), and so on. No *single* decolonial project exists, rather decolonial efforts are marked by a multiplicity and plurality of objectives, aims, and focuses. In this chapter, we ground our approach to (de)coloniality primarily within the context of the *African continent*, informed by works from the continent. With 54 independent, heterogeneous, diverse, and dynamic nations, each with diverging cultures and values, the African continent is anything but homogeneous. For the purpose of this chapter, however, we focus on the similarities found in decolonial work across the continent. Thus, our perspective is informed by feminist and decolonial theories, practices, and experiences emerging from the African continent.

Colonialism has profoundly altered the African continent and its relationship with the rest of the world. It has impacted the way legal systems are structured, educational institutions are established, knowledge is constructed, and even the languages we speak. For example, the concept of race served as a tool that was used to legitimize colonialism by portraying the Western white man as the "human par excellence" while portraying Black and indigenous populations as inferior and merely human (Saini, 2019).

An important tactic of colonialism is to *erase* historical, cultural, and intellectual contributions of the entire continent in efforts to construct it as devoid of history. This has contributed to difficulties in identifying histories of the African continent that are independent of colonialism. Such erasure of the continent's history creates a vacuum that can be filled with Western narratives and ideologies about the continent and its people (Tamale, 2020). Acknowledging this, some decolonial efforts focus on *restoring* African heritage, by exploring its rich culture and intellectual contributions. Such efforts highlight and expose these intentionally erased contributions and render a more truthful narrative and image of the continent. This might include projects that, for example, restore languages that have come close to extinction.

The erasure of indigenous knowledge systems, which are replaced by Western perspectives, has a far more insidious consequence: the *colonization of the mind* (Fanon, 1986). The colonization of the mind refers to the process of erasing and replacing indigenous knowledge systems with Western structures and norms such that Western knowledge systems are perceived as "natural", even by the colonized. Recognizing the potency of this process, Biko emphasized that, "the most powerful weapon in the hands of the oppressor is the mind of the oppressed." (Biko, 2013). Colonization of the mind is a gradual and inconspicuous process, that relies on indigenous communities embodying the colonizer's version of African history.

Breaking free from colonization of the mind requires understanding this history, peeling underneath the white man's version of our history, recovering our own history, and perceiving this recovered history as "natural." In short, this means replacing the whitewashed history of Africa with accurate historical records. Decolonizing the mind therefore requires reviving indigenous histories and knowledge systems and raising critical consciousness to undo the internalized conceptions, categories, and knowledge systems in order to counter racist hegemonies (Biko, 2013; Fanon, 1986; Freire, 2018). This is no small challenge as the white man's knowledge system permeates everything from our classifications of gender and sexuality to our understanding of race, to how legal institutions and education systems operate. An in-depth *understanding*, therefore, is the first and foremost requirement, rather than prediction, classification, clustering, or abstraction.

We cannot escape the colonial mentality as long as the theories, concepts, and questions that inform our research, study, and practice are generated from Western experiences and tradition. Currently, coloniality pervades our day-to-day life. It provides the foundation for social, cultural, institutional, and scientific endeavors. We witness coloniality through the normalization and internalization of arbitrarily constructed conceptions and racial categories. Subsequently, the very objectives of AI, its methodologies, and cultural ecology – and therefore its conceptions of classification and categorization – are inherently colonialist.

Furthermore, the colonialist project is built on individualistic, “rational”, and positivist science that stems from the enlightenment's obsession to sort, classify, and categorize in a hierarchical and dualistic manner (Tamale, 2020). In contrast, decolonizing research comprise methods that value, reclaim, and foreground Indigenous voices and ways of knowing. That is, relational, contextual, and historical approaches; specifically, the African philosophy of Ubuntu (Birhane, 2021; Mhlambi, 2020) which seeks to decentre Western hegemonic knowledge systems and ideologies.

Machine Learning and Abstraction

Modern ML is a field that develops upon the statistical sciences with the aim of uncovering patterns from data through optimization processes (Bishop, 2006). ML and statistics share the assumption that meaning can be made of complex human behavior and experiences through processes of simplification and abstraction, in data that is “frozen in time” (Talat, Lulz, et al., 2021). ML distinguishes itself from statistics in its preoccupation with prediction: Where statistics has, to a large degree, moved on from seeking to predict *future* behavior from past data,[4] ML is concerned with predicting the future from the past. We can therefore understand ML as a discipline that overarchingly focuses on *future-making* based on salience, that is hegemony, in past data (Talat, Lulz, et al., 2021). Further, ML often assumes that data and models exist outside of context, i.e., that seizing data from the World Wide Web – language data, for example – and processing it using ML affords models that are devoid of the many contexts from which data and model are wrought. Here we examine how ML abstracts from contextual, dynamic, contested, and unjust pasts to devise hegemonic futures.

Tools of Abstraction

Common ML paradigms include supervised (e.g., classification), unsupervised (e.g., clustering), and reinforcement learning (RL, e.g., game-playing). These paradigms assume that meaning can be made through abstraction. For instance, when DeepMind developed a RL system to play Go, named AlphaGo, the system optimized the decision-making process for selecting which move to play (Silver et al., 2017). AlphaGo optimized for patterns that would lead to victory by playing thousands of games against “itself”, and ultimately beat several of the world’s best players. In comparison to human behavior, Go is a relatively simple game that can be captured with large volumes of data. For instance, while the game

[4] While statistics is still used to forecast future events, e.g. election results, the emphasis lies within explaining why a specific trend appears to be salient and uncertainty is at the forefront of communication..

of Go is played according to the same rules across the globe, the norms of interaction vary greatly for human behavior and expectations.[5] Norms, practices and cultures from one society do not necessarily map onto another (Talat et al., 2022). Thus, while the game of Go lends itself to pattern recognition, cultural practices are deeply contextual. For example, Geertz (1973, p. 6) details the distinction between a wink and an involuntary twitch:

> *[T]he difference, however unphotographable, between a twitch and a wink is vast; as anyone unfortunate enough to have had the first taken for the second knows. The winker is communicating, and indeed communicating in a quite precise and special way*

While there exists no photographic distinction between a wink and a twitch, detailed and in-depth queries make it clear that the two are distinct. Only by attending to the details of the phenomena, can we obtain the in-depth understanding required to distinguish the two.

Language is another realm that does not readily lend itself to abstraction. Despite impressive advances for classification and text generation, NLP models still fundamentally fail to understand text (Bender & Koller, 2020). This discrepancy between performance and understanding is a result of the distributional hypothesis that models rely on. Under this hypothesis, meaning can be made from (sub-)words by observing the frequencies of their (co-)occurrence (Harris, 1954). That is, by enumerating word interaction patterns to create vector spaces, NLP technologies can supposedly understand the meanings of words. Famously, this has afforded the recognition that “king” and “queen” often occur in similar contexts Pennington et al., 2014).[6] However, even within this success, NLP technologies fail to attune to crucial details, namely the gendered power imbalances. This has resulted in a host of methods to “debias” these through an abstractive view of social marginalization (Bolukbasi et al., 2016; Zhao et al., 2017).[7] Rather than seeking to understand which processes give rise to discriminatory outcomes, the field has primarily attended to devising abstractive methods that treat concrete harms as abstract subjects, framing harms as “bias” (Blodgett et al., 2020). This framing, in a sense, minimizes the ‘unpleasantness’ and constructs the topic as one that can be studied abstractly from afar. Under this construction, ML models can only be biased, and therefore be de-biased, by abstracting away from individual instances of discrimination, in favor of considering the discrepancies through collective metrics for fairness. Moreover, only harms wrought directly by the system itself are considered, any external discrimination, e.g., over-policing, is not considered relevant for determining whether a system is biased and therefore an eligible candidate for “de-biasing”. Thus, when vector spaces tie the nursing profession more strongly to women than other genders, the dispute is neither the economic or social marginalization, nor is that a genderless profession is more strongly tied to one gender. The dispute arises from the hyper-fixation on the abstraction of several words being more closely associated to one gender than others. The dispute, then, disregards the socio-economic systems that give rise to the association and instead attends to the abstract expression of marginalization. Addressing the dispute therefore fixates on the abstraction of marginalization, rather than the source or

[5] We note that there exist 5 different scoring systems systems for Go.

[6] “Context” wrt. Word embeddings refers to a sliding window within a sentence, rather than a wider context.

[7] Within NLP, which biases are addressed remain heavily skewed towards gender bias, at the expense of other directions such as racialized biases (Field et al., 2021).

direct expression. Thus, although ML models come to reproduce factors of identity and marginalization, they do not fully capture these issues. That is, machine learning models capture proxies of identity, rather than features of identity itself. Crucially, even after "de-biasing" vector spaces, the biases they were treated for can still be reconstructed in the vector space (Gonen & Goldberg, 2019).

Social Infrastructures of Machine Learning

Beyond the ML tools that seek to abstract or generalize, the infrastructures within which ML operates create the foundations upon which ML models function. A handful of for-profit companies, that serve financial rather than scientific purposes, provide sponsorship for conferences and direct funding for research. This grants them access to researchers and allows them to influence the direction of public research. As a result, these corporations hold an outsized influence on the research conducted through direct means (i.e., dual positions and research internships) and indirect means (i.e., through priority setting and research funding) (Abdalla & Abdalla, 2021; Ahmed & Wahed, 2020; Birhane et al., 2021). Although many such companies claim altruistic purposes, their primary interest is to profit, rather than provide utility. The astronomical profits of these companies require a deep commitment to the continued oppression of marginalized communities. For instance, there would be no Alphabet, at the scale that we see today, without the continued abuses of post-colonial subjects e.g., in the physical mining for resources. Moreover, as Meta, Palantir, and Cambridge Analytica have shown, the people that were colonized, continue to face exploitation for labor by Western companies and are externalized for territorial expansion or technological exploitation (Perrigo, 2022).

When ML initiatives arise by and for colonial subjects, they are quickly approached by large technology companies in attempts to co-opt their resources. One may hope that the development of high-quality technologies for under-served languages or communities, e.g. Amharic or Amhara people, would provide benefits for the population. However, the benefits for a community must be weighed against the additional marginalization that the technology affords. Take for example Google's freely available translation system. On one hand, having access to a high-quality translation system could afford greater ease of communication, which may be in some cases critical, e.g., for refugees fleeing war.[8] On the other hand, such systems afford an easy way to increase the surveillance of speakers of languages that are largely disregarded by research efforts. The interests of corporate entities such as Google would then be financial, for example, by providing targeted advertisements, rather than humanitarian. While providing communities with relevant advertisements may seem benign, targeted advertisements come with high costs, e.g., privacy (Ullah et al., 2021) and risk of destabilizing democracies (Dubois et al., 2022). Thus, rather than providing a benefit to marginalized communities, the corporate (use of) ML tools and resources would result in significant risks and exploitation of such groups by capitalist and colonial interests.

Machine Learning and Decoloniality: An Inherent Tension?

[8] Prior audits have found that biases exist in the ways in which the system makes translation (Hovy et al., 2020).

In our account of characteristics of decoloniality and ML, it becomes apparent that the two schools of thought stand on opposite sides of a spectrum in their aims, objectives, methodologies, interests, motivations, and practices. Due to the substantial influence from Western individualistic thinking, ML reflects Western principles such as the emphasis on 'objectivity', rationality, and individuality, which tend to approach subjects of study (including human behavior and society) as ideas devoid of context and history. Furthermore, ML can be used to advance neo-colonial and pseudo-scientific ideologies, through its reliance on methodologies that are similar to those used by 19th century phrenologists.

Through narratives and dichotomies such as 'near vs far', 'us and others', and 'west vs east', colonial powers justified dehumanizing indigenous populations (Willinsky, 1998). Conceptualizations of the colonized as the 'other' from 'afar' have enabled the colonizers to treat indigenous communities, their cultures, and languages as abstract 'subjects' that can be studied, manipulated, controlled, and molded at will. Similarly, through abstraction, ML treats human beings (and their feelings, desires, wishes, and hopes) as abstract data points. By dealing with 'data', people become distant statistics.

At its core, ML aims to detect patterns across large datasets. This requires abstracting away from the details, idiosyncrasies, and particularities of individual situations. This is often motivated by capitalist wealth extraction, which operates under the Western values of individualism and white supremacy. The core values of decolonizing, conversely, include correcting historical records (that have been intentionally erased and manipulated), illuminating indigenous knowledge systems (e.g., grounded in Ubuntu philosophy, which is fundamentally relational at its core), and raising critical consciousness against internalized coloniality. Most importantly, decolonizing is about undoing past harm and injustice and mapping alternative just futures.

In contrast to decoloniality, ML is a field that is noted for its lack of critical awareness where it "produces thoughtlessness, the inability to critique instructions, the lack of reflection on consequences, and a commitment to the belief that a correct ordering is being carried out" (McQuillan, 2019). Currently, much of current ML is embedded in institutions and organizations with oppressive histories. ML, in turn has endowed such institutions and organizations with a seemingly scientific justification for colonialism and white supremacy. Due to a lack of reflexivity and critical examination of the past, current ML research and practice functions as a tool that extends such unjust histories.

Conditions for Decolonial AI

Lewontin (2003) notes the difference between defeatism and scepticism: defeatism leads to passivity and scepticism to action. Similarly, we contend that imagining a decolonial future – difficult and challenging as it may seem – is a necessary step towards making such a future a reality, despite the apparent incompatibility between decoloniality and ML. With this in mind, we list the core conditions that need to be present for a decolonial ML to materialize. This list is by no means exhaustive but instead can serve as a starting point.

A necessary step towards this future, is that AI systems must serve the needs of indigenous peoples, in a manner that is informed by and grounded in indigenous epistemologies, experiences and needs. Therefore, such systems must be built, controlled, and owned by indigenous peoples, and the primary beneficiaries must be people. Profit or capital interest must be a secondary concern.

The idea of a general(izable) AI is both vacuous and serves the interests and needs of the status quo. Decolonial AI systems must therefore aim to serve a small group of people, rather than “all of humanity”, who do not require context to be provided by the machine. ML, at its core, is a process that abstracts away, obfuscates, and standardizes. Therefore, people must be able to read the details themselves, which is only possible when the abstractions of the machine are tasks that can be performed by humans but are more efficiently performed by the machine (i.e., the machine can save human time spent). Additionally, procedures for the right to contest ML models, processes, underlying assumptions, outputs, and training regimes must be put into place. In extension, the objectives of a decolonial AI or ML system must challenge racist, colonialist, white-supremacist, patriarchal, and other unjust and marginalizing ideologies.

It is also important that the system is divorced from ideas of eugenics, phrenology, and similar racist and white-supremacist ideologies. Many of the ML applications that are being integrated into the social sphere – whether in aiding decision-making in housing or social welfare benefits – serve the purpose of excluding or filtering out (those who are deemed undeserving for housing or welfare benefits), which is inherently punitive. In contrast, decolonial AI should be constructive, restorative, and built on communal values that contribute to the current and future prosperity of marginalized communities.

Decolonial Futures

We close this chapter by highlighting the Te Hiku NLP project as an example of decolonial AI. Various factors make the Māori community data collection, management, and technology development practices stand out. For the Te Hiku NLP project, entire communities of Te Reo Māori speakers in Aotearoa, or New Zealand were mobilized in participatory efforts to develop AI systems to revitalize their rapidly disappearing language. The Māori community hold full control of the project. From the very conception, every step of the developmental pipeline has been based on Māori principles and taken in concordance Māori values, emphasizing benefits the community. Their use of AI is driven by the need for language revitalization efforts and to obtain equal rights for *Te Reo Māori*. Unlike much of the current “the bigger, the better” mantra behind language models, the Te Hiku NLP project is driven by small communities and relatively small data sizes.

The Te Hiku NLP project collected 310 hours of speech-text pairs from 200,000 recordings made by 2,500 people over the course of ten days. The data came from speakers of Te Reo Māori throughout New Zealand and was annotated and cleaned by members of the Māori community. The data was then used to develop a speech recognition model which performs with 86% accuracy. The main drive to build such technology came from the push to preserve Māori culture and language. During the British colonial exploitation, the speakers of *Te Reo Māori* were prevented from speaking their language through shaming and physical beatings of Māori students (Auckland University Libraries and Learning Services, 2017). The motivation to reclaim *Te Reo Māori* and the rich culture that surrounds it led to the development of computational linguistic tools. As a way of digitizing the language and culture, elders were recorded, and the material kept in digital archives for younger generations to access and learn from. Te Hiku built a digital hosting platform maintain full control of their data and avoid influence from large technology corporations. The community subsequently established the Māori Data Sovereignty Protocols (Kukutai & Cormack, 2021; Raine et al., 2019) as a way for the Māori to hold full autonomy and control of their data, technology, and therefore, future. This has been described as a sign of "Indigenous resistance—against colonizers, against the nation-state, and now against big tech companies" (Coffey, 2021). This effort thus highlights a path for decolonial AI. Rather than aiming to develop a tool for a general public, Te Hiku sought to create a tool for language revitalization of *Te Reo Māori* language. By resisting scaling and the co-option from large technology companies while centering the benefit to the Māori communities, the Te Hiku project maintains the ability to use an abstractive technology within a decolonial context and provides a vision for the shape of future decolonial AI projects.

Conclusion

The rise of machine learning has elicited critical questions around the fundamentally oppressive nature of the technology, at least within the minor corners of Ethical AI, leading to the question of whether a truly decolonial machine learning is possible? As we illustrate throughout this chapter, machine learning stands in stark contrast to decoloniality. Machine learning reproduces colonial logics in the process of its search for abstraction, simplification, clustering, and prediction. At the core of decoloniality, on the other hand, is decentring Western hegemony, restoring erased histories, and uplifting and showcasing historical and current and intellectually contributions, driven by justice and equity. This work is performed through in-depth understanding rather than abstraction or simplification. However, despite the opposing logics, assumptions and objectives, Te Hiku's *Te Reo Māori* language project illustrates that there is space for the development of machine learning systems that decolonize.

References

Abdalla, M., & Abdalla, M. (2021). The Grey Hoodie Project: Big Tobacco, Big Tech, and the Threat on Academic Integrity. *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society*, 287–297. https://doi.org/10.1145/3461702.3462563

Abdilla, A., Arista, N., Baker, K., Benesiinaabandan, S., Brown, M., Cheung, M., Coleman, M., Cordes, A., Davison, J., Duncan, K., Garzon, S., Harrell, D. F., Jones, P.-L., Kealiikanakaoleohaililani, K., Kelleher, M., Kite, S., Lagon, O., Leigh, J., Levesque, M., … Whaanga, H. (2020). *Indigenous Protocol and Artificial Intelligence Position Paper*. https://doi.org/10.11573/SPECTRUM.LIBRARY.CONCORDIA.CA.00986506

Adams, R. (2021). Can artificial intelligence be decolonized? *Interdisciplinary Science Reviews*, *46*(1–2), 176–197. https://doi.org/10.1080/03080188.2020.1840225

Agüera y Arcas, B., Mitchell, M., & Todorov, A. (2017, May 20). Physiognomy's New Clothes. *Medium*. https://medium.com/@blaisea/physiognomys-new-clothes-f2d4b59fdd6a

Ahmed, N., & Wahed, M. (2020). *The De-democratization of AI: Deep Learning and the Compute Divide in Artificial Intelligence Research* (arXiv:2010.15581). arXiv. http://arxiv.org/abs/2010.15581

Alexander, D. (2020). Decolonizing Digital Spaces. In E. Dubois & F. Martin-Bariteau (Eds.), *Citizenship in a connected Canada: A research and policy agenda*. University of Ottawa Press.

Auckland University Libraries and Learning Services. (2017, October 6). *Ngā Kura Māori: The Native Schools System 1867-1969*. https://www.news.library.auckland.ac.nz/2017/10/06/native-schools/#.Yi_ULBDMJH0

Barongo-Muweke, N. (2016). *Decolonizing education: Towards Reconstructing a Theory of Citizenship Education for Postcolonial Africa*. Springer Berlin Heidelberg.

Battiste, M. (2013). *Decolonizing education: Nourishing the learning spirit*. Purich Publishing Limited.

Belden-Adams, K. (2020). *Eugenics, "aristogenics," photography: Picturing privilege* (First edition). Bloomsbury Visual Arts.

Bender, E. M., & Koller, A. (2020). Climbing towards NLU: On Meaning, Form, and Understanding in the Age of Data. *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 5185–5198. https://doi.org/10.18653/v1/2020.acl-main.463

Biko, S. (2013). White Racism and Black Consciousness. In C. Crais & T. V. McClendon (Eds.), *The South Africa Reader* (pp. 361–370). Duke University Press. https://doi.org/10.1215/9780822377450-064

Birhane, A. (2021). Algorithmic injustice: A relational ethics approach. *Patterns*, *2*(2), 100205. https://doi.org/10.1016/j.patter.2021.100205

Birhane, A., & Guest, O. (2020). *Towards decolonising computational sciences* (arXiv:2009.14258). arXiv. http://arxiv.org/abs/2009.14258

Birhane, A., Kalluri, P., Card, D., Agnew, W., Dotan, R., & Bao, M. (2021). The Values Encoded in Machine Learning Research. *ArXiv:2106.15590 [Cs]*. http://arxiv.org/abs/2106.15590

Bishop, C. M. (2006). *Pattern recognition and machine learning*. Springer.

Blodgett, S. L., Barocas, S., Daumé III, H., & Wallach, H. (2020). Language (Technology) is Power: A Critical Survey of "Bias" in NLP. *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 5454–5476. https://doi.org/10.18653/v1/2020.acl-main.485

Bolukbasi, T., Chang, K.-W., Zou, J. Y., Saligrama, V., & Kalai, A. T. (2016). Man is to Computer Programmer as Woman is to Homemaker? Debiasing Word Embeddings. In D. Lee, M. Sugiyama, U. Luxburg, I. Guyon, & R. Garnett (Eds.), *Advances in Neural Information Processing Systems* (Vol. 29). Curran Associates, Inc. https://proceedings.neurips.cc/paper/2016/file/a486cd07e4ac3d270571622f4f316ec5-Paper.pdf

Byrne, D. (2022). Science in Africa: Is 'decolonization' losing all meaning? *Nature Africa*, d44148-022-00064–1. https://doi.org/10.1038/d44148-022-00064-1

Challis, D. (2014). *The archaeology of race: The eugenic ideas of Francis Galton and Flinders Petrie* (Paperback edition). Bloomsbury.

Coffey, D. (2021). Māori are trying to save their language from Big Tech. *Wired UK*. https://www.wired.co.uk/article/maori-language-tech

Currier, A. (2011). Decolonizing the law: LGBT organizing in Namibia and South Africa. In A. Sarat (Ed.), *Special issue, social movements/legal possibilities*. Emerald.

Davidson, T. Bhattacharaya, D., Weber, I. (2019). Racial bias in hate speech and abusive language detection datasets. Proceedings of the Third Workshop on Abusive Language Online, 25-35. https://doi.org/10.18653/v1/W19-3504

Di Paolo, E. A., Cuffari, E. C., & De Jaegher, H. (2018). *Linguistic bodies: The continuity between life and language*. The MIT Press.

Dubois, P. R., Arteau-Leclerc, C., & Giasson, T. (2022). Micro-Targeting, Social Media, and Third Party Advertising: Why the Facebook Ad Library Cannot Prevent Threats to Canadian Democracy. In H. A. Garnett & M. Pal (Eds.), *Cyber-threats to Canadian democracy*.

Fanon, F. (1986). *Black skin, white masks* (Repr.). Pluto Press.

Field, A., Blodgett, S. L., Talat, Z., & Tsvetkov, Y. (2021). A Survey of Race, Racism, and Anti-Racism in NLP. *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, 1905–1925. https://doi.org/10.18653/v1/2021.acl-long.149

Freire, P. (2018). *Pedagogy of the oppressed* (M. B. Ramos, Trans.; 50th anniversary edition). Bloomsbury Academic.

Gambäck, B., & Sikdar, U. K. (2017). Using Convolutional Neural Networks to Classify Hate-Speech. *Proceedings of the First Workshop on Abusive Language Online*, 85–90. https://doi.org/10.18653/v1/W17-3013

Geertz, C. (1973). *The interpretation of cultures*. Basic Books.

Gitelman, L. (Ed.). (2013). *“Raw data” is an oxymoron*. The MIT Press.

Gonen, H., & Goldberg, Y. (2019). Lipstick on a Pig: Debiasing Methods Cover up Systematic Gender Biases in Word Embeddings But do not Remove Them. *Proceedings of the 2019 Conference of the North*, 609–614. https://doi.org/10.18653/v1/N19-1061

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press.

Harris, Z. S. (1954). Distributional Structure. *WORD*, *10*(2–3), 146–162. https://doi.org/10.1080/00437956.1954.11659520

Hovy, D., Bianchi, F., & Fornaciari, T. (2020). “You Sound Just Like Your Father” Commercial Machine Translation Systems Include Stylistic Biases. *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 1686–1690. https://doi.org/10.18653/v1/2020.acl-main.154

Iseke-Barnes, J. M. (2008). Pedagogies for decolonizing. *Canadian Journal of Native Education*, *31*(1), 123–148.

Kukutai, T., & Cormack, D. (2021). “Pushing the space”: Data sovereignty and self-determination in Aotearoa NZ. In M. Walter (Ed.), *Indigenous data sovereignty and policy*. Routledge.

Kuwali, D. (2014). Decoding Afrocentrism: Decolonizing Legal Theory. In O. Onazi (Ed.), *African Legal Theory and Contemporary Problems* (Vol. 29, pp. 71–92). Springer Netherlands. https://doi.org/10.1007/978-94-007-7537-4_4

Leslie, C. (2012). Decolonizing the internet. *Global Media and Communication*, *8*(1), 81–88. https://doi.org/10.1177/1742766512439806

Lewontin, R. (2003). *Biology as ideology: The doctrine of DNA*. Anansi.

McQuillan, D. (2019). Non-fascist AI. In M. Hlavajova & W. Maas (Eds.), *Propositions for non-fascist living: Tentative and urgent*. BAK, basis voor actuele kunst ; The MIT Press.

Mhlambi, S. (2020). *From Rationality to Relationality: Ubuntu as an Ethical and Human Rights Framework for Artificial Intelligence Governance*. Carr Center Discussion Paper Series, 2020-009.

Ndlovu-Gatsheni, S. J. (2012). Coloniality of power in development studies and the impact of global imperial designs on Africa. *The Australasian Review of African Studies*, *33*(2), 48–73.

Ngũgĩ wa Thiongʾo. (1986). *Decolonising the mind: The politics of language in African literature*. J. Currey ; Heinemann.

Oyěwùmí, O. (1997). *The invention of women: Making an African sense of Western gender discourses*. University of Minnesota Press.

Pennington, J., Socher, R., & Manning, C. (2014). Glove: Global Vectors for Word Representation. *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 1532–1543. https://doi.org/10.3115/v1/D14-1162

Perrigo, B. (2022, May 11). *Meta Accused Of Human Trafficking and Union-Busting in Kenya*. Time. https://time.com/6175026/facebook-sama-kenya-lawsuit/

Raine, S. C., Kukutai, T., Walter, M., Figueroa-Rodríguez, O.-L., Walker, J., & Axelsson, P. (2019). Indigenous data sovereignty. In T. Davies, S. B. Walker, M. Rubinstein, & F. Perini (Eds.), *The State of Open Data.* African Minds, IDRC. https://www.doabooks.org/doab?func=fulltext&uiLanguage=en&rid=34137

Raji, I. D., Bender, E. M., Paullada, A., Denton, E., & Hanna, A. (2021). AI and the Everything in the Whole Wide World Benchmark. *ArXiv:2111.15366 [Cs]*. http://arxiv.org/abs/2111.15366

Raval, N. (2019). An Agenda for Decolonizing Data Science – spheres. *Spheres: Journal for Digital Cultures*, *5*, 1–6.

Saini, A. (2019). *Superior: The return of race science*. Beacon Press.

Sekula, A. (1986). The Body and the Archive. *October*, *39*, 3. https://doi.org/10.2307/778312

Silver, D., Schrittwieser, J., Simonyan, K., Antonoglou, I., Huang, A., Guez, A., Hubert, T., Baker, L., Lai, M., Bolton, A., Chen, Y., Lillicrap, T., Hui, F., Sifre, L., van den Driessche, G., Graepel, T., & Hassabis, D. (2017). Mastering the game of Go without human knowledge. *Nature*, *550*(7676), 354–359. https://doi.org/10.1038/nature24270

Spanton, R. W., & Guest, O. (2022). Measuring Trustworthiness or Automating Physiognomy? A Comment on Safra, Chevallier, Grézes, and Baumard (2020). *ArXiv:2202.08674 [Cs]*. http://arxiv.org/abs/2202.08674

Talat, Z., Blix, H., Valvoda, J., Ganesh, M. I., Cotterell, R., & Williams, A. (2022). On the Machine Learning of Ethical Judgments from Natural Language. *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 769–779. https://doi.org/10.18653/v1/2022.naacl-main.56

Talat, Z., Lulz, S., Bingel, J., & Augenstein, I. (2021). *Disembodied Machine Learning: On the Illusion of Objectivity in NLP*. http://arxiv.org/abs/2101.11974

Talat, Z., Névéol, A., Biderman, S., Clinciu, M., Dey, M., Longpre, S., Luccioni, S., Masoud, M., Mitchell, M., Radev, D., Sharma, S., Subramonian, A., Tae, J., Tan, S., Tunuguntla, D., & Van Der Wal, O. (2022). You reap what you sow: On the challenges of bias evaluation under multilingual settings. *Proceedings of BigScience Episode #5 – Workshop on Challenges & Perspectives in Creating Large Language Models*, 26–41. https://aclanthology.org/2022.bigscience-1.3

Tamale, S. (2020). *Decolonization and Afro-feminism*. Daraja Press.

Ullah, I., Boreli, R., & Kanhere, S. S. (2021). *Privacy in Targeted Advertising: A Survey* (arXiv:2009.06861). arXiv. http://arxiv.org/abs/2009.06861

Voulodimos, A., Doulamis, N., Doulamis, A., & Protopapadakis, E. (2018). Deep Learning for Computer Vision: A Brief Review. *Computational Intelligence and Neuroscience*, *2018*, 1–13. https://doi.org/10.1155/2018/7068349

Wa Thiong'o, N. (2018). On the abolition of the English Department. *Présence Africaine*, *N°197*(1), 103. https://doi.org/10.3917/presa.197.0103

Willinsky, J. (1998). *Learning to divide the world: Education at empire's end*. University of Minnesota Press.

Zhao, J., Wang, T., Yatskar, M., Ordonez, V., & Chang, K.-W. (2017). Men also like shopping: Reducing gender bias amplification using corpus-level constraints. *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, 2979–2989. https://doi.org/10.18653/v1/D17-1323