Colonization by Algorithms in the Fourth Industrial Revolution

Wynand Lambrechts¹, Snr., Member, IEEE, Saurabh Sinha², Fellow, IEEE, and Sarah Mosoetsa³

¹University of Johannesburg, Faculty of Engineering and the Built Environment, Johannesburg, South Africa
²University of Johannesburg, Deputy Vice-Chancellor: Research and Internationalization, Johannesburg, South Africa
³National Institute for the Humanities and Social Sciences, Johannesburg, South Africa

Corresponding author: Wynand Lambrechts (e-mail: wynand.lambrechts@ieee.org).

ABSTRACT Data gathering and information processing have evolved to where it is almost unfathomable how much exists in digital form today. The generation thereof also no longer involves an explicit instruction from human to machine but can happen in real-time without human intervention. Artificial intelligence, machine learning, and cognitive computing are being utilized to mine data from a variety of sources. One such (profitable) source is human beings. Digital algorithms are designed to harness the power of technology to gather information. There has always been a sense of secrecy regarding some information (classified, top secret, confidential, etc.) but the Fourth Industrial Revolution has created the means to gather extremely large amounts of data, unknown to its sources. Anthropological value systems should become a fundamental foundation of digital algorithms. Such an approach could prevent software from exploiting its sources, especially minorities. Value systems together with ethics are guided by people’s culture. In ethically aligned algorithm design, value systems and digital technologies intersect and govern how algorithms are developed, the way data is engaged, and further the discipline of digital humanities.

INDEX TERMS Artificial intelligence, digital algorithms, COVID-19, digital colonialization, ethically aligned design, invisible data

I. INTRODUCTION

If data is considered a raw material, then information can be considered organized data that has been given meaning and can be interpreted by a receiver. In theory, there is little wrong with this perspective, and it gives humans a means to readily communicate and make sense of numbers and patterns. Some of the earliest forms of modern data representation (numerals) are described in [1] where, according to [1], the natural way of recording numbers (dating over 30000 years ago) was by tally marks, as shown in Fig. 1.

![FIGURE 1. A representation of tally marks used as the first form of counting and the origin of modern numerals.](image)

Tally marks, as shown in Fig. 1, were used not only to count, but to keep a record of occurrences. Historically tally marks varied in different parts of the world, with the tally marks in Fig. 1 most common in Europe, Southern Africa, Australia, New Zealand, and North America. [1] gives an extensive history on the various tally marks globally. In their essence, however, tally marks (in whichever form) are simple marks, or scores, used for counting and are among the first forms of visually representing data. Fast-forwarding past the Roman numerals (I, V, X, L, C, D, M with their first usage between 900 B.C. and 800 B.C.) and the Hindu-Arabic numerals (1, 2, 3, 4, 5, 6, 7, 8, 9, 0 with their first usage between the 6th and 7th century), the binary system (1’s and 0’s and dating back to the 17th century [2]) is how most modern data and information are generated and stored.

Again, as yet nothing seems out of the ordinary. Data storage plays a key role in modern society and presenting it as information can be accomplished in various ways, including

- lean information or captcha (electronically processed and reformatted data to give it meaning),
• rich information (data that is processed by a human and visually presented to an audience),
• appropriated information (a human acquiring data or information and mentally processing it by means of existing knowledge),
• tacit information (demonstrating information through developing and practising a skill, such as operating a vehicle), and
• explicit information (expressing information through language or art).

All these forms of presenting data and information are, for the most part, deliberate, and hence the processing, presentation, consumption and interpretation of the information is a conscious decision. However, through technological advances in the Fourth Industrial Revolution (4IR), data is being contributed by users, technology, and cyber-physical interfaces, often unknowingly as forms of cognitive technologies. For example, a modern 64-bit central processing unit can hold any of 2^64 different values (and therefore some representation of data or information) at any given time. Furthermore, modern commercial processors can process this information at well over 100-million instructions per second. Considering the number of processors globally, it is evident that big data is being generated and analysed each second.

[3], writing in 1962, already hinted that statistics would contribute significantly to data analysis in the future as more data is gathered. In 1962, however, data gathering was still a manual process and often obtained through stochastic processes that potentially contain bias. Modern researchers in the social sciences are debating the concept of drawing generalized assumptions for entire populations based on subsets of data obtained from historical information, since this may contain bias. [4] illustrates which simplified patterns in a hypothetically selected sample can be applied to demonstrate the existence of sample-selection bias in a source. Importantly, [4] acknowledges that sample-selection bias exists in historical data and that individuals’ values and observations (the human factor) can lead to biased information.

In modern society, big data can be generated through cognitive technologies and from users where the data and the algorithms find a way to collude, often through artificial intelligence (AI), bringing about unintended consequences and even bias depending on the sources used to produce the data. Some of these unintended consequences have been specifically correlated with invisible data. [5] describe visibility processes in social media data and what data remains invisible to the users. Essentially, if data is generated, analysed, and used by ‘someone’ but invisible to us (the source), this invokes a sinister feeling.

Data collection is also leading to digital colonization where minorities and lower socio-economic communities are being data-mined by large corporations which have the means to establish and maintain dominance in the digital domain relatively quickly. The effects can be positive or negative, either as a form of digital inclusion for minorities, or as a system which biases digital technology through invisible data. In August 2021, the World Economic Forum [6] concluded that, based on 2 million participants, modern young people have more faith in algorithms than they have in politicians. For numerous reasons this is an important discovery. The intent of digital or automated data collection should fundamentally be based on value systems that recognise the culture of a community and are ethical. To develop digital algorithms that respect and assimilate the value systems of a people, an understanding of their culture is essential. Furthermore, for digital technologies to be introduced into the developing world, specifically to the youth on the African continent, an adherence to local culture and value systems through value-sensitive algorithm design (VSAD) [7] can help integrate the underrepresented. Foreign concepts often deter participation and interest, whereas known, respected, and acknowledged practices could encourage upskilling, economic stimulus, and socio-economic participation and growth in local communities. VSAD is therefore aimed at maximizing the contribution to human dignity and minimizing its potential for misuse [8]. To take advantage of big data and prevent it from victimizing minorities, it should be founded on value systems in digital algorithms.

III. VALUE SYSTEMS IN DIGITAL ALGORITHMS

Anthropology is the study of culture, the set of traits of a people that are distinctive from other societies, communities, and nationalities [9]. These differentiating traits include language, traditional dress code, music, work, arts, religion, and other [9]. Deeply rooted in culture are value systems, the set of social, moral, political, religious, aesthetic, and economic values that govern how a group of people live, work, and worship.

For VSAD to succeed, anthropology as the scientific study of humanity in both the present and the past must be combined with machine learning, AI, and data collection. VSAD can be used to increase human influence on computerized algorithmic procedures by engaging relevant participants in the early stages of creating algorithms. Tacit values, knowledge, and insights into abstract and analytical machine learning are incorporated in a VSAD approach [8] which aims to engage human experience in mitigating digital bias by integrating ethically aligned design (EAD). [8] proposes incorporating values in terms of what has been shown to be objectively good for the community by researchers in the humanities and social sciences through an evaluation of their current environment. The goal is to develop VSAD that discerns how algorithms promote justice and security from the standpoint of value sensitive design (VSD). [8] identifies key values for the design,
implementation, and deployment of VSD-algorithms as: being accurate, autonomous, private, equal, true to ownership, accountable, and transparent. According to the research presented by [7], transforming a traditional VSD approach into the context of algorithm design can be achieved in six steps:

1. Have a clear and accurate understanding of the value systems, goals, culture, and motivations of the people who will use the algorithms as well as the people who will be impacted.
2. In terms of details as to who will be affected and how, algorithm identification and prototyping should ensue and be shared with the community along the development phases.
3. The third step entails finding fitting ways to implement the prototypes and, importantly, engage with contributors within the community for detailed feedback, criticism, and suggestions.
4. The deployment step follows on community feedback on prototypes; however, this step should importantly remain interactive and iterative to enhance the performance and results of the algorithms.
5. According to [7], the algorithms should undergo a form of acceptance in terms of their accuracy and whether they impact positively on the community. Again, community involvement is critical at this stage and their observations of efficacy should be prioritized over those of the designers.
6. Finally, continuous revision and adjustments to the algorithms will allow for evolution over time that can improve their effectiveness.

In [7], these six steps are also extended to an automated, machine-learning, AI-based algorithm approach where feedback should be focused not only on the effectiveness but also on the (ideally, lack of) digital bias when extracting big data. Such an approach would further involve clearly defining the envisaged objective, using accurate and non-biased historical data, and applying the models to newly generated data for further validation.

[7] presents an in-depth case study of the open-source project, Wikipedia, based on the principles outlined above. Wikipedia is an example of an online encyclopaedia that is written, edited, and policed by a global group of volunteers. The information on Wikipedia is free for anyone to access and is used as a significant source of data and information by digital algorithms. The data and information that are available on Wikipedia, although reviewed and authenticated by the community, consist of peer-reviewed published material, non-biased scientific information, and individual contributions that are not necessarily verified by specialists. As a result, data mining from Wikipedia by means of AI and machine learning is likely to include both non-biased and biased facts [10], potentially with a disregard for value systems. Governing information on Wikipedia in terms of value systems would be a big task considering the amount of information that is available and the number of contributors to the website. However, it is possible to innovate and create novel digital resources that are fundamentally based on a set of values.

In the southern half of Africa, the culture of ubuntu exhibits a shared ethical commitment whereby individuals, communities, and nationalities are agreed on a set of core bioethical values which allow them to strive towards a quality life engrained in a spirit of family [11]. The Nguni word “ubuntu” is expressed by different terms across the range of Bantu languages, but these share a common philosophy of caring, sharing, respect, and compassion [12]. [13] defines ubuntu as “A collection of values and practices that Black people of Africa or of African origin view as making people authentic human beings”. In 2004, Mark Shuttleworth launched a free and open-source operating system named after the philosophy of ubuntu. In open-source software development, ubuntu is advanced by bringing together the power of various computer programmers – to use and leverage the power of collaboration towards a common goal of improving, learning, and innovating a resource that benefits a specific community.

The data and information gathering, collaborating, and presentation processes from the above two examples are largely visible to their users and contributors. Their intent, whether driven by value systems or not, is clear and aims to achieve a specific objective, either to enlighten people or to serve as a predictable tool for consuming information. On the other hand, modern technology and big data allow for invisible data to be generated and used in circumstances that could be undetectable to users.

IV. INVISIBLE DATA

To introduce invisible data through a practical example, adapted from [14], consider the following scenario: Two business colleagues meet up for a work-related discussion in a coffee shop. They order a Cappuccino and a Caffe Americano respectively. On concluding the discussion, a photo of the event and a reference to the fruitful discussion is posted to social media to encourage followers. Afterwards, on the same social media platform, one colleague notices several advertisements for Cappuccino’s and Caffe Americano’s, where to purchase these, how to make them, and the best coffee shops to enjoy them. Even though this data was not completely invisible (photos were uploaded and the order placed audibly), it can be argued that no consent was given to use this data for the purposes that it was subsequently used for. AI algorithms from social media and other online services do affect how individuals and groups are exposed to specific advertisements [15].

[5] argues that a variety of social phenomena are being studied through social media since these leave digital traces
offering insights into human interaction. For example, Vision AI or Computer Vision (CV) (which trains computers to replicate human vision) benefits greatly from digital traces in social media. Dr. Murphy Choy, Executive Director at MC EduTech, presented on “Intercepting youth at risk using computer vision on social media platforms” at the virtual conference on Computer Vision DevCon 2020 (cvdc.adasci.org) where he addressed increased domestic violence [16] during the Coronavirus Disease 2019 (COVID-19) lockdowns. Dr. Choy attributed success in detecting and identifying people in distress, and especially women and children, to good CV algorithms which could identify human emotions behind social media posts. In [5], specific focus is placed on the data in social media research that remains invisible. It is concluded and recommended by [5], that users distance themselves from any assumptions that all social media data are visible and rather focus on the invisible (meta)data to enhance their understanding of modern digital information. Through a better understanding of digital data, society could become more aware of how not to be exploited by digital records, colonized by algorithms, or how to avoid individualized data being used against users. Digital humanities (DH) study the structured use of digital resources and analyze its applications for impact on cultural heritage and digital culture. However, DH will be required to progress at a far quicker pace, given hyper-advancing digital technology. In [17] DH have evolved into a discipline, and rightly so considering the big data and algorithms to be governed. Furthermore [17] proposes that “the difficulty in defining DH arises from the disciplinary diversity and a set of rather heterogeneous scholarly practices within the field”. Research in [18] has also suggested that it is nearly impossible to govern algorithms, especially already-implemented ones, but that communities should be a part of the emerging and ongoing discussion around identifying best practices. Among these are to

- be skeptical of digital resources,
- be reflective of methodologies,
- consider if DH can be engaged to modify technologies (such as algorithms), and
- embrace the methodological and statistical aspects that serve as the foundation for DH.

[19] reported on some ways in which invisible data can be used against the user, using a few examples that presumably do not implement DH. These include financial exploitation, profiling in political campaigns, allegedly “free” Wi-Fi networks gaining access to the privacy and security details of a user where the user ends up paying for the service with their data as the “currency,” discrimination (in various forms), and profiling potential employees. As a result, big data as a complex set of political, cultural, economic, and scientific practices which make sense of or generate information on society [20] has become a means for some companies and institutions to achieve digital colonization, primarily through adept digital algorithms. An ethical approach to algorithms is not always evident or followed and can lead to disruptive, unforeseen, or unintended consequences. It could therefore be argued that deliberately ignoring and omitting the ethical alignment of algorithms could be a conscious attempt towards digital colonization and marginalization. The analogy of digital colonization through algorithms is appropriate considering that corporate agendas [21] are “mining” people for data, a phrase which equates people with raw material from whom miners have the potential of generating wealth. There are similarities between classic and digital colonization where, most notably, developed entities aim to enforce power and control over developing minorities.

V. DIGITAL COLONIZATION OF THE DEVELOPING WORLD

Colonization is an amalgamation of territorial, juridical, cultural, linguistic, political, mental, and/or economic dominance of one group of people over another [22]. For example, European colonization involves the territorial control by European powers of non-European colonies. Historically, colonization is often associated with developed countries or groups dominating developing countries or groupings. Colonization in the digital domain targets three main issues of modern times, high levels of income disparity, digital divides, and shortage of skills [15]. There are broad tendencies shared between colonialisms, however domination, law, appropriation, and containment are distinct and dynamic over time for each colonial territory [22]. Similar trends in adaptive, biased, and tailored digital colonization are being witnessed.

Digital colonization is ingrained in technology infrastructure and its main purpose is profit. It monopolizes digital sectors such as social media, electronic commerce, transportation, and politics. Furthermore, [21] argues that technology and algorithms developed with Western capitalistic viewpoints in mind are typically unfit for African needs and concurrently diminish the development of regional products by asserting a dependence on Western software.

Companies such as Facebook, Google, Apple, Amazon, and Samsung have clearly engineered dominance in the technology sector, more so in data collection, analysis, and manipulation [21]. In doing so, these companies can monitor, track, and perform constant surveillance of their users, and in developing countries, the effects of this power can be amplified for several reasons. Firstly, emerging economies often do not have the financial means to regulate or take on technology giants and are at the mercy of these companies. Furthermore, emerging economies often do not have the resources to develop competing software or services and are forced to rely on the major technology companies for ownership of smart devices that can access services such as
connecting with friends and family, storing data, purchasing online goods or services. Frequently, large technology corporations engineer complete ecosystems that can ‘trap’ users, making it difficult or sometimes impossible to remove themselves from it so that they are forced to continually purchase these branded products or services. In so doing, these companies are then able to manipulate behavior or nudge users into action and towards lucrative consequences for the firm, or to advocate other forms of personal benefit. Digital colonization can also be found within a country where discrimination against minorities who are already on the borders of civilization is increased, though not by the doing of another entity. Such trends often include a form of digital divide and can be considered digital colonization between people with and without access to the internet. Such effects have been emphasized during COVID-19 and are especially notable on the African continent. COVID-19 is globally associated with local government-imposed lockdowns based on current rates of COVID-19 infections and deaths. Many businesses are forced to close their doors during a lockdown or can only operate during specific hours and following strict safety protocols. Digital technology has facilitated the persistence of work, entertainment, education, and interaction and in many cases, ensured that economies continue to grow. It has also changed the way many employees will in future conduct their work and business-related tasks, with some companies adopting permanent work from home policies or hybrid approaches. In a country with high levels of income disparity, digital colonization as a result from, for example, COVID-19, will transpire almost automatically. [23] argues that improved vaccination education can result in better vaccine uptake, an important consideration especially for developing countries where slow vaccine rollout will have an even larger long-term effect on economic recovery.

Large disparities between regions and communities have resulted in differing impact of COVID-19. According to the A4AI Affordability Report 2020 [24], the digital divide combined with COVID-19 has created challenges for professionals in Latin America, teachers in Southeast Asia, and learners from across Africa to South Asia. [25] reported that only 28% of the African population are considered to form part of the online population. Significant growth in online access of over 6% was reported in [26], however, the number remains low. The cost of connectivity is one of the unsurmountable obstacles that keeps most of the African population offline [26].

Rural inhabitants with no means to access the internet and unable to travel due to lockdown restrictions are more likely to forfeit their income or lose their jobs. Wealthier communities are more likely to have broadband internet access and therefore some means to continue working, or at least to perform some tasks, even if their remuneration is reduced based on their outputs. The poor, with already very little of their basic needs met, will end up with even less, whereas the wealthier community will perhaps have to cut down on luxury spending, often only for a limited time. The possibility for digital colonization thus increases as the digital divide and socio-economic disparities widen. This creates opportunities to enforce digital colonization in these regions, often under the umbrella of digital inclusion, that are likely to be maintained for the foreseeable future. Internet access for empowerment of minorities and marginalized users should be distinguished from attempts at digital colonization [27]; not all attempts are to be considered malicious, and they could be used to educate people on the benefits of vaccination, as [23] suggest. However, if the aim is for personal benefit only, providing internet access to the marginalized is only one phase of digital colonization, while data collection and generating information through tacit collusion of algorithms driven by invisible data is another phase.

V. COLONIZATION BY ALGORITHMS

[15] explicitly reports that certain digital products and services have built-in biases designed for and constantly learning through AI algorithms. A 2020 survey within the United Kingdom revealed that nearly half the local councils were using algorithms to assist in decisions based on benefit claims, social housing, and other issues [15]. Furthermore, very little research and analysis as to the reliability of these algorithms was being conducted, yet the outcomes were still implemented based on the suggestions from the algorithms. Concerning public service issues such as overrepresentation also emanate from algorithm bias. As an example from [15], people from lower socio-economic backgrounds will often seek aid for mental illness, drug abuse, or co-morbidities at public clinics, whereas wealthier individuals will have access to private services. As a result, more data is available on the lower socio-economical groups, which could result in biased decision-making when, for example, custody of minors is under scrutiny. As technology further advances together with the rapid growth and evolution of social media, models, theories, and approaches that mitigate the divide between Western approaches and African decolonization are required to encourage indigenization.

Social network websites and other forms of invisible data affect, often adversely, the balance between private, commercial, and public space [28]. Facebook allegedly used user data, including information about friends, relationships, and photos as leverage over companies it partnered with. Facebook rewarded favored companies by giving them access to the data of its users and denying rival companies user data [29]. Similarly, in 2020, the US Federal Trade Commission (FTC) investigated Twitter for potentially misusing people’s personal information to serve advertisers [30]. Numerous other examples of social networking companies being investigated for some sort of misconduct on sharing user data through optimized algorithms can be found.
on online sources. It is therefore essential specifically for the Global South, which relies on Western software, to be aware of how algorithms are being used to digitally colonize users.

According to [15], the bearing of AI on inequalities between countries will, to an extent, rely on the nature of the input data. The three primary forms of input data are through digital platforms (social media, electronic commerce, and others), harnessing the internet-of-things (IoT), or by human-machine AI interaction. Countries such as the USA and China will be more likely to generate big data through competing digital platforms, whereas countries such as Japan, the Republic of Korea, and to an extent the European Union, generate large quantities of big data through the IoT [15]. Human-machine AI interactions are still limited; however, it is likely that developed countries will increase their dominance in big data over developing countries as this technology matures. As AI algorithms improve and the gap between developing and developed countries widens, ethically aligned designs and integration become crucial to protect vulnerable users from being digitally colonized. An ethical approach is critical; however, standardization and accountability are needed to enforce and regulate such ideals.

V. ETHICALLY ALIGNED ALGORITHM DESIGN

Algorithms have widespread potential to modernize procedures, decrease human bigotry, and reduce expenditure, however, algorithms are not inherently neutral. In some cases, discrimination can be an unintended side effect of machine learning with algorithmic bias. Machine learning is fundamentally based on aggregation and analysis of historical information, and any past bias that is rooted in the data can be learned and imitated. Historically, many minorities have fallen victim to colonization and these events are captured in history books. Allowing AI and machine learning to mine these resources without any ethical foundation or value systems will likely result in further marginalization of these minorities.

For example, in 2016, Microsoft set to release Tay, a chatbot designed to understand conversation and chat with real human beings. Through the juncture of machine learning, natural language processing, and social networks, Tay was designed to discover more about language through AI, as opposed to being pre-programmed with standard responses from interpreting responses [31]. Within 16 hours of Tay’s being introduced into the public domain, the chatbot generated over 95000 Tweets. A significant percentage of the Tweets were offensive, racist, fascist, or abusive. Consequently, Tay was quickly removed from the experimental website and later replaced with Zo, a version of Tay tailored to avoid any forms of hate speech. Microsoft and many other entities learned through this experience that algorithmic bias will find its way into machine learning, regardless of what is intended. But if Zo was therefore programmed to respond more in a more politically correct manner, does this not lead to another form of pre-disposed bias? This experiment showed in clear terms that historical information as well as people’s behavior is innately biased and machine learning is likely to expose this if no EADs are incorporated.

Academics, non-government organizations, governments, and industry have realized the growing concerns about ethical principles in AI and are collaborating on initiatives to ensure fair, transparent, accountable, and inclusive processes. The Institute of Electrical and Electronics Engineers (IEEE) has endeavored to advance the principle of EAD in the machine-to-machine and person-to-machine interfaces that are likely to form an integral role in 4IR. EAD borrows principles from the humanities and social sciences, uses value systems and leverages the power of communities through open-source software to map how algorithms think, utilize data, and cooperate. The University of Johannesburg, South Africa, offers online modules that teach African Insights to create awareness of African value systems, and incorporates these value systems in generating algorithms that follow the EAD approach. The goal is to form collective intelligence that is useful, ethically, and perhaps emotionally intelligent. Through EAD, algorithms could represent a critical step forward in correcting flawed or biased feedback loops and technology could be used to assist in addressing social issues, as opposed to relying on it to solve these issues through pure automation.

V. MORALITY IN MACHINE LEARNING

The introduction of value systems in algorithms for developing countries that are often at the mercy of global digital monopolies could encourage sustainable socio-economic growth through their participation in digital innovation. In [6], one recommendation by modern youth explicitly calls on large technical corporations to be “more transparent about misinformation and its spread on their online platforms...be more inclusive of stakeholders, including users and governments, in developing and designing underlying technologies, such as algorithms for content moderation, enduser [sic] policies and community guidelines”. It can be deduced from [6] that the modern youth understand the reach of algorithms and are aware that they are often being kept in the dark. Digital colonization can only be achieved if the target group is not educated as to the dangers and impact thereof. It is therefore imperative to educate, upskill, and inform not only the youth in developing countries, but all people about the implications of transparency in algorithms. Global conglomerates operate on profit, and digital colonization is a convenient first step towards generating and exploiting invisible data. Some institutions often overlook ethically aligned algorithms or are not aware of the cultural values or concerns of users. Digital expansion and ethically aligned algorithms should be treated similarly to how traditional values, for example on the African continent, are identified, taught, and encouraged. In [32] a course of action is presented to identify traditional
values and spread their importance through education, and this can be adapted for value systems in algorithms. From [32], and adapted for global relevance, this course of action is to:

- Establish transdisciplinary reflection committees to classify conventional and contemporary national values and to define shared values among age groups and societies within these local communities.
- Use the national education systems, media, and information sessions as foundations to integrate frameworks for the sharing of customs and values.
- Form an infrastructure of champions to distribute and promote the importance of values and culture specifically among the local youth.
- Further encourage, promote, and preserve cultural values through traditional and generational ceremonies, socializing, and education.
- Formally classify and execute national policies on fundamental values and cultural heritage and assess the long-term impacts of early contact with traditional media, social media, and the rational and moral development of the youth.

Although this course of action seems straightforward (a short, bulleted list that can be adapted per culture) indigenous people interpret ontologies in the context of culture to be “philosophical, semantic, conceptual, formal, informal, representation, logical-theoretic, property-driven, purpose-driven, vocabulary, specification, and/or multi-levelled” [18]. Teams of researchers, indigenous people, and algorithm development engineers would be required to integrate true morality in machine learning.

Considering the traditional identification of values as in [32] and referring to the five steps to apply VSD within the context of algorithm design in [7], certain similarities can be drawn:

- In both cases, in identifying traditional value systems as well as developing algorithms that are value-sensitive, any actions should be preceded by an in-depth understanding of local value systems, goals, and culture. Algorithms will be used by and have an impact on local individuals and for the algorithms to have a positive impact, the needs of the people should be reflected.
- After the identifying phase, in both cases (traditional value systems and VSD algorithms), there should be a feedback stage. Initial research, as in-depth and accurate as it claims to be, can nevertheless have bias or inconsistencies. Machine learning bias can be corrected by active involvement from local contributors.

- Acceptance, formalizing, and possibly developing policies and standards that govern algorithm design will ensure its longevity and adoption/acceptance by local communities for ideally positive and sustainable results.

The fundamentals of machine learning should be rooted in morality, especially since the information and knowledge contained in the algorithms are continuous with new data and information constantly being generated. Lacklustre initial design and implementation have a relatively high likelihood of mutating into biased or prejudiced information that can obscure matters regarding assimilating culture into a basis of sustainable development.

V. CONCLUSION

Data and information are important building blocks of society. Data gathering leading to information generation has changed drastically over time: initially a manual process, it has evolved through digitalization into autonomous systems capable of mining big data at a tremendous pace. In 4IR, data is contributed by users, technology, and cyber-physical interfaces, often unknowingly as a form of cognitive technologies. At the heart of data gathering and processing are algorithms, originally designed by humans and capable of operating and evolving on their own. The intent of these algorithms is therefore guided by human involvement and can take on unanticipated forms. Another building block of society are the cultures and value systems that have been studied in anthropology. In this work, it is argued that, as a core principle, anthropologic objectives should be included in digital algorithms to facilitate outcomes that benefit people. Value sensitive algorithm design is a relatively new term that addresses such aspects when digital algorithms are to be used to autonomously generate and analyse large amounts of data and information. This principle is especially important when invisible data that is unknowingly mined from users could be used as a form of digital colonization. The modern age of technology hardware has to an extent removed any limitations on the amount of data that can be collected and stored. Its use, however, is still largely determined by its intent, with private institutions, governments, and research facilities able to customize information being fed to users, especially in developing countries where security and privacy policies are at the mercy of conglomerates. Social media, electronic commerce, transport, and politics are sectors that thrive on big data and often neglect ethical value systems at their core. As it is already clear that modern youth have more trust in digital algorithms than in politicians, it becomes ever more important to ensure that digital algorithms and AI have solid foundations in value systems, morality, and ethics.

REFERENCES

[1] C. E. Woodruff, “The Evolution of Modern Numerals from Ancient Tally Marks,” The American Mathematical Monthly, vol. 16, no. 8/9, pp. 125–133, 1909, doi: 10.2307/2970818.
2. F. G. Heath, “Origins of the Binary Code,” Scientific American, vol. 227, no. 2, pp. 76–83, 1972. [Online]. Available: http://www.jstor.org/stable/24927408

3. J. W. Tukey, “The Future of Data Analysis,” The Annals of Mathematical Statistics, vol. 33, no. 1, pp. 1–67, Mar. 1962. doi: 10.1214/aoms/1177704711

4. A. Zimran, “Recognizing Sample-Selection Bias in Historical Data,” Social Science History, vol. 44, no. 3, Jul. 2020, doi: 10.1017/shh.2020.11.

5. C. Neumayer, L. Rossi, and D. M. Struthers, “Invisible Data: A Framework for Understanding Visibility Processes in Social Media Data,” Social Media and Society, vol. 7, no. 1, 2021. doi: 10.1177/2056305120984472.

6. WEF. “Youth Recovery Plan,” Geneva, Aug. 2021. Accessed: Aug. 13, 2021. [Online]. Available: http://www3.weforum.org/docs/WEF_Davos_Lab_Youth_Recovery_Plan_2021.pdf

7. H. Zhu, B. Yu, A. Halfaker, and L. Terveen, “Value-Sensitive Algorithm Design,” Proceedings of the ACM on Human-Computer Interaction, vol. 2, no. CSCW, Nov. 2018, doi: 10.1145/3274463.

8. P. Hayes, I. van de Poel, and M. Stein, “Algorithms and values in justice and security,” AI & SOCIETY, vol. 35, no. 3, pp. 533–555, 2020, doi: 10.1007/s00146-019-09932-9.

9. G. E. Idang, “African culture and values,” Phronesis, vol. 16, pp. 97–111, 2015.

10. A. Zhavoronkov, “Capturing Biases In The Age Of AI - The Interview With The Founder Of Wikipedia Founder Jimmy Wales,” Forbes, Apr. 14, 2021. https://www.forbes.com/sites/alexzhavoronkov/2021/04/14/capturing-biases-in-the-age-of-ai-the-interview-with-the-founder-of-wikipedia-founder-jimmy-wales/?s=23130a5f5996 (accessed Aug. 07, 2021).

11. L. T. Chwasa, African Indigenous Ethics in Global Bioethics, vol. 1. Dordrecht: Springer Netherlands, 2014. doi: 10.1007/978-94-017-8625-6.

12. J. Broodryk, Ubuntu: Life Lessons from Africa, 1st ed. Ubuntu School of Philosophy, 2002.

13. J. Mugumbate and A. Chereni, “Editorial: Now, the theory of Ubuntu has its space in social work,” 2020.

14. S. Sinha and S. Mosoeta, “Colonisation by algorithms: The Fourth Industrial Revolution and the ethical use of predictive analytics,” Business Maverick, Jul. 21, 2021. https://www.dailymaverick.co.za/article/2021-07-21-colonisation-by-algorithms-the-fourth-industrial-revolution-and-the-ethical-use-of-predictive-analytics/ (accessed Jul. 21, 2021).

15. UNCTAD, “Technology and innovation report 2021: catching technological waves, innovation with equity,” Geneva, 2021. Accessed: Jul. 29, 2021. [Online]. Available: https://digitallibrary.un.org/recording/3926808/files/tir2020_en.pdf

16. M. L. Evans, M. Lindauer, and M. E. Farrell, “A Pandemic within a Pandemic — Intimate Partner Violence during Covid-19,” New England Journal of Medicine, vol. 383, no. 24, Dec. 2020, doi: 10.1056/NEJMSp2004046.

17. J. Luhmann and M. Burghardt, “Digital humanities—A discipline in its own right? An analysis of the role and position of digital humanities in the academic landscape,” Journal of the Association for Information Science and Technology, Jun. 2021, doi: 10.1002/asi.24533.

18. M. J. Hill, “Invisible interpretations: reflections on the digital humanities and intellectual history,” Global Intellectual History, vol. 1, no. 2. May 2016, doi: 10.1080/23801883.2017.1304162.

19. “Invisible Manipulation: 10 ways our data is being used against us,” Privacy International, Jan. 16, 2018. https://privacyinternational.org/long-read/1064/invisible-manipulation-10-ways-our-data-is-being-used-against-us (accessed Jul. 14, 2021).

20. A. Mattoni and E. Pavan, “Politics, Participation and Big Data. Introductory Reflections on the Ontological, Epistemological, and Methodological Aspects of a Complex Relationship,” Partecipazione e Conflitto, vol. 11, no. 2, pp. 313–331, Jul. 2018.

21. A. Bhirane, “Algorithmic Colonization of Africa,” SCRIPT-ed, vol. 17, no. 2, pp. 389–409, Aug. 2020, doi: 10.2966/scrp.170220.389.

22. A. Murray, “Colonialism,” in International Encyclopedia of Human Geography, A. Kobayashi, Ed. Elsevier, 2020. doi: 10.1016/B978-0-08-102295-5.10804-2.

23. S. Sinha, R. Phaswana-Mafuya, and B. van Vuuren, “We must aggressively promote Covid-19 vaccination education for a better vaccine uptake,” The Daily Maverick, Aug. 04, 2021. https://www.dailymaverick.co.za/article/2021-08-04-we-must-aggressively-promote-covid-19-vaccination-education-for-a-better-vaccine-uptake/ (accessed Aug. 04, 2021).

24. A. AAI., “The Affordability Report 2020,” 2020. https://e18q3l6yv81g81h3md5g5xe-wpengine.netdna-ssl.com/wp-content/uploads/2020/12/Affordability-Report-2020.pdf (accessed Jul. 29, 2021).

25. GSMA, “State of the industry report on mobile money 2019,” Mar. 2019. Accessed: Jul. 29, 2021. [Online]. Available: https://www.gsma.com/sotir/wp-content/uploads/2020/03/GSMA-State-of-the-Industry-Report-on-Mobile-Money-2019-Full-Report.pdf

26. GSMA, “State of the industry report on mobile money 2021,” Mar. 2021. Accessed: Jul. 29, 2021. [Online]. Available: https://www.gsma.com/mobilefordevelopment/wp-content/uploads/2021/03/GSMA_State-of-the-Industry-Report-on-Mobile-Money-2021_Full-report.pdf

27. B. Mehra, C. Merkel, and A. P. Bishop, “The internet for empowerment of minority and marginalized users,” New Media & Society, vol. 6, no. 6, Dec. 2004, doi: 10.1177/1461448804047513.

28. R. Heyman and J. Pierson, “Social Media, Delinguistification and Colonization of Lifeworld: Changing Faces of Facebook,” Social Media and Society, vol. 1, no. 2, Sep. 2015, doi: 10.1177/2056305115621933.

29. K. Conger, “Twitter faces $250m fine over new data privacy issue,” Aug. 04, 2020. https://www.irishtimes.com/business/technology/twitter-faces-250m-fine-over-new-data-privacy-issue-1.4321674 (accessed Aug. 04, 2021).

30. O. Schwartz, “In 2016, Microsoft’s Racist Chatbot Revealed the Dangers of Online Conversation,” IEEE Spectrum, Nov. 25, 2019. https://spectrum.ieee.org/in-2016-microsofts-racist-chatbot-revealed-the-dangers-of-online-conversation (accessed Aug. 04, 2021).

31. WATHI, “What are the African values today?,” Sep. 2017.

JOHANNES W. LAMBRECHTS (Senior Member, IEEE) obtained his B.Eng., M.Eng., and PhD degrees in Electronic Engineering from the University of Pretoria (UP), South Africa. He achieved his M.Eng. with distinction. He has authored 2 publications in peer-reviewed journals and has presented at various local and international conferences. He is the lead author of five books; in the fields of sustainable energy and in microelectronic engineering, all published by international publishers. He has co-authored six contributing chapters in other books in the fields of green energy and technology and the fourth industrial revolution.

He is currently serving as Secretary for the IEEE South Africa Section and a Chapter Coordination Subcommittee Corresponding Member for IEEE Region 8. He previously held a position as an electronic engineer, a radio frequency (RF) engineer, and is currently appointed as a principal semiconductor engineer. He is also serving as a part-time research associate at the University of Johannesburg (UJ), South Africa.
SAURABH SINHA (Fellow, IEEE) PhD(Eng), Pr Eng, FSAIEE, FSAAE, MASSAf obtained his B. Eng (with distinction), M. Eng (with distinction), and Ph.D. degrees in Electronic Engineering from the University of Pretoria (UP). As an established researcher, rated by the National Research Foundation (NRF), he has authored or co-authored over 130 publications in peer-reviewed journals and at international conferences. He served UP for over a decade; his last service being as Director of the Carl and Emily Fuchs Institute for Microelectronics, Department of Electrical, Electronic and Computer Engineering. On 1 October 2013, Prof Sinha was appointed as Executive Dean of the Faculty of Engineering and the Built Environment (FEBE) at the University of Johannesburg (UJ). As of 1 Dec. 2017, Prof Sinha is the UJ Deputy Vice-Chancellor: Research and Internationalisation. Among other leading roles, Prof Saurabh Sinha also served the IEEE as a Board of Director and IEEE Vice-President: Educational Activities.

SARAH MOSOETSA is an Associate Professor of Sociology, at the University of the Witwatersrand, Johannesburg. She has previously worked as a senior researcher at the Human Sciences Research Council and the Development Bank of Southern Africa. Sarah completed her PhD in 2005 at the University of the Witwatersrand in the Department of Sociology. She has published in both local and international journals including Labour, Capital and Society, Transformation, and Journal of Southern African Studies. She has recently published a book entitled Eating from one Pot: Dynamics of Survival in poor South African households, Wits University Press.