



# Decolonization of AI: a Crucial Blind Spot

John W. Murphy<sup>1</sup> · Carlos Largacha-Martínez<sup>2</sup>

Received: 21 March 2022 / Accepted: 7 October 2022 / Published online: 28 November 2022

© The Author(s), under exclusive licence to Springer Nature B.V. 2022

## Abstract

Critics are calling for the decolonization of AI (artificial intelligence). The problem is that this technology is marginalizing other modes of knowledge with dehumanizing applications. What is needed to remedy this situation is the development of human-centric AI. However, there is a serious blind spot in this strategy that is addressed in this paper. The corrective that is usually proposed—participatory design—lacks the philosophical rigor to undercut the autonomy of AI, and thus the colonization spawned by this technology. A more radical or substantial proposal is advanced in this discussion that is known as community-based design. This alternative makes a theoretical maneuver that allows AI design to be directed by human agency, thereby introducing a safeguard that may help to prevent colonization by this technology.

**Keywords** Neocolonialism · Human-centered organizations · Technological imperative · Dualism and Cartesianism · Community-based ontologies

## 1 Introduction

Calls are coming from various sectors to decolonize AI (artificial intelligence) and establish a more human-centered version (Auernhammer, 2020; Ehn, 2017). Some critics are arguing that technology, and AI in particular, is a product of the West and has undermined local cultures around the world (Mhlambi, 2022). In this regard,

---

✉ Carlos Largacha-Martínez  
clargacha@areandina.edu.co

John W. Murphy  
j.murphy@miami.edu

<sup>1</sup> Sociology Department, University of Miami, Coral Gables, USA

<sup>2</sup> School of Management, Fundación Universitaria del Área Andina, Bogotá, Colombia

Abeba Birhane (2019) argues that “tech evangelism” has occurred, whereby worship of these devices has pushed aside other values and customs. The phrase that has been used to describe this trend is “algorithmic colonization.”<sup>1</sup>

Another side of this issue is more technical. In this case, colonization occurs because of the abstract nature of AI (Wei, 2019). These critics argue that because of the way in which AI is portrayed, as an idealized rendition of rationality, other modes of knowledge are overwhelmed. Consistent with the Western penchant for searching for ultimate foundations—Ideas, God, natural laws, etc.—algorithms have been given a seigniorial status. In more practical terms, there has been little transparency in AI, and even the engineers, let alone the public, do not often know the content of most algorithms and accept the reasoning that is operating.

In both cases, the problem with AI is that this technology appears to be autonomous, that is, cut-off from the contingencies and interpretive character of everyday life. Descriptives such as universal, natural, or abstract are often used to describe this condition. When AI is considered to be autonomous, machines can operate independently of human involvement, thereby resulting in the alienation of employees (Braverman, 1975). When autonomy results from the pursuit of universals, persons are severed from their creations, including self-regulating machines. A much bigger issue is raised when autonomy is associated with particular modes of knowledge that are granted a unique ontological position.

Therefore, simply looking to reduce the political influence of the West, and the power that is exerted, may not be sufficient to remedy the problem of colonization. The reason for this assessment is quite simple: no attention is directed to confronting the theoretical maneuver that makes this autonomy, and thus colonization by AI, possible.

When cloaked in logic and mathematics, algorithms can begin to assume an almost ethereal form and dominate less stylized modes of knowledge. The abstract, and arguably scientific appearance of algorithms, has enabled this technology to marginalize local considerations.<sup>2</sup> Most persons are thus unaware of the ongoing manipulation and loss of their agency, since algorithms are given a *sui generis* status and become “gods” (O’Neil, 2016: 3, location 196 of 4460, Kindle version). For example, the unemployment that is a product of this technology is regularly thought to be entirely rational, although discomfiting in the short run. After all, reason dictates that some persons may become redundant because of technological innovations.

Another, but related, side of this issue is that limited achievements in AI development are presented as huge successes. Although natural language processing, sentiment analysis, and affective computing have outcomes that are far from useful,

---

<sup>1</sup> Traditionally, colonization refers to the domination and exploitation of one country by another, including the inferiorization of every facet of life of those who are dominated (Memmi, 1968). In this discussion, the focus is on the domination of local knowledge by computer technology, particularly by the algorithms associated with AI.

<sup>2</sup> In this discussion, local refers to how persons, in workplaces or communities, interpret their situations and relationships, and act on the stock of knowledge that is accumulated.

advocates continue to claim that big achievements are at hand. The downside is that AI cannot grasp human values, thereby revealing the alignment problem (Christian, 2020). Where dualism comes into play is that success at programming driverless cars and winning at Go is very different from understanding the *lifeworlds* of persons and how they give meaning to everyday actions. Again, the issue is that this technology seems to have the autonomy to restrict applications and frame criteria for success.<sup>3</sup>

In his discussion of colonization, Enrique Dussel (2012) goes to what he believes to be at the core of cultural autonomy. He argues convincingly that colonizers were helped in their conquest by Descartes and the dualism fostered by Cartesianism. Armed with the claim of advancing objective, universal knowledge, the colonizers could control indigenous opposition with a seemingly sound rationale. In fact, the negation of local culture as legitimate is how colonization has been constructed. What is going to be presented later on in this discussion is that the same autonomy is exhibited when developing AI tools, since most employees' knowledge and expertise, their local knowledge, are not treated as valid to include in algorithms.

In this presentation, colonialism refers to the domination of local knowledge by algorithms, assisted by Cartesianism (Cruz, 2021). Of course, big tech companies and Western capitalism benefit from this process, but attacking these enterprises will not necessarily solve the problem of knowledge autonomy and discrediting local knowledge. Although power and politics are important in this regard, the point is that ignoring philosophy in this decolonization effort is a problem. Certainly, people and companies are promoting these systems. Nonetheless, the legitimacy of this technology is not necessarily derived from something political or technological (Heidegger, 1987).

Additionally, while the biases that are often built into algorithms and discriminate against minority groups—for example, in terms of facial recognition or employment selection—are not the focus of this exposition, these prejudices are protected by the autonomy attributed to these devices (Cave & Dihal, 2020). These biases are imagined to be beyond critique because of a particular philosophy.

The key omission in recent discussions of colonization by AI is that no attention has been directed to the underlying philosophy that supports the autonomy of this technology. What is going to be presented later on in this discussion is that this philosophy is operative when developing AI tools, since most employees' knowledge and expertise, their local knowledge, are not treated as relevant to constructing algorithms. The legacy of Cartesianism treats their contributions to be replete with error and unproductive.

Current assessments of the colonizing ability of AI could benefit from a critique of dualism, so that knowledge supremacy can be adequately addressed and

---

<sup>3</sup> Note should be taken of recent challenges to the autonomy of AI. Critics such as Safiya Umoja Noble (2018) have pointed out the biases that have been built regularly into algorithms. The racial bias in algorithms is his under scrutiny (Cave and Dihal 2020). The problem, however, is that a lot of the correctives proposed rely on improved technology to reduce discrimination, instead of incorporating local knowledge into the creation of algorithms (Mitchell 2019).

colonization rolled back. In this way, decolonization can occur, whereby the knowledge associated with algorithms is brought down to earth and viewed as one possible rendition among others. As Jean Luc-Nancy (2016) argues in another context, when all knowledge is understood to be mediated thoroughly by human agency, every position is vulnerable, subject to critique, and modification, and, if deemed unworthy, rejection. Through this mediation, the special status reserved for alleged absolutes is undercut, along with their ability to bedazzle persons.

## 2 Cartesianism and AI

Shakir Mohamed (2018) is perhaps the first author to argue for the decolonization of AI. This problem, he contends, stems from AI researchers having a scientific, idealized view of knowledge. And according to Appadurai (2001; 2013), this abstract, almost unreal position, strips everyday persons of self-confidence and ownership of knowledge. As he points out, AI appears to be perfect, whereas humans are flawed; humans are emotional and algorithms are entirely rational. To correct this situation, and nullify these images of colonization, they must be able to create and use their own knowledge (Mohamed et al., 2020). Persons must retrieve their ability to act with authority and confidence, and reclaim their agency.

This retrieval of agency is not a simple task, particularly in the aftermath of Cartesianism. Throughout most of the Western intellectual tradition, persons are taught that their intentions and passions are a problem in the search for and the discovery of true knowledge and sound ethical principles (Bordo, 1987). In short, the turmoil of the human condition must be overcome if truth or reason is ever going to be achieved. In many ways, this trend took a radical turn with Descartes.

Instead of making speculative claims about universal foundations, Descartes made a fairly straightforward argument that set in motion a widespread dualist agenda.<sup>4</sup> That is, he argued that subjective (mind) and objective (body) elements could be separated. Following this insight, scientists argued later that facts and values occupy different spheres. Their point is not only that this distinction is possible but necessary to acquire valid knowledge. With subjectivity sequestered, through the efforts of a strict and neutral methodology, reliable, objective knowledge is accessible. The negative influence of values, beliefs, or opinions can thus be minimized.

But how did this seemingly innocuous epistemological maneuver support colonization? According to Dussel (2012), this dualism fostered the opportunity for claims to be made about superior forms of knowledge and, eventually, proposals about dominant cultures. The prospect was available that some cultures could be treated as progressive, and even universal, and others as primitive and backward. As Dussel

---

<sup>4</sup> Although Descartes (1596–1650) is often associated with mind–body dualism, his influence extends beyond this division (Bordo 1987). Although certainly an offshoot of Descartes' distinction, dualism has appeared in a variety of disciplines, such as physics (Bernstein 1983) and medicine (Aho, 2008). In these cases, the so-called objective element is treated as autonomous and, thus, a source of valid knowledge, while the subjective is marginalized. Indeed, overcoming the subjective is essential to discovering valid knowledge.

(*Ibid.*) points out, the project of colonization rests on the inferiorization of indigenous cultures, so that these societies can be controlled and exploited by outsiders. Because of the dirempt character of these cultures, their survival, along with any progress, depends on the interventions of the colonizers. In a manner of speaking, workers who are not tech specialists are being treated worldwide as an indigenous population in the process of developing AI tools at their workplaces.

Note should be taken that not all workers are affected at the same time and in the same ways by AI (Walsh, 2018). At this moment, some jobs appear to be immune to its influence. Nonetheless, this technology is encroaching in most workplaces in ways that make employees nervous and feel replaceable. There is an abiding fear that their worth and labor could easily be diminished by this technology (Zhang & Dafoe, 2019).

The supremacy of AI is reinforced by associating particular knowledge and commitments with science and reason. Colonizers, for example, do not merely base their claims of superiority or inferiority on opinion but on science. Judgements about social standing are thus predicated on calculation and the rational separation of persons, since “ideas only deserved the appellation of knowledge if they could be put in to the test of empirical evidence” (Hughes & Sharrock, 1997: 29). Particularly noteworthy is that eugenics became a key weapon in the colonizers’ arsenal. In some other cases, declarations about divine preference performed this function. Although ostensibly different, both rely on abstractions, divorced from the contingencies of everyday life, to support the dominance of one racial or ethnic group over others. In legitimating these distinctions, the abstract references that are used are practically inviolable.

Critics of AI contend that this technology performs a similar role in the colonization of knowledge. Linked to logic and science, the knowledge and outcomes associated with AI are a sign of advanced intelligence (Bostrom, 2014). Algorithms, in fact, are treated as idealized formulas that operate in ways that supersede human capabilities. Attributes are given to these devices, such as advanced thinking and learning, that provide them with a special status. Correspondingly, algorithms are often portrayed in ways that render humans obsolete and vulnerable.

A current example of this colonization is taking place at Amazon fulfillment centers. In these locations, workers find, sort, and prepare the goods that customers select for delivery. The pace and regulation of these tasks are controlled by algorithms that are insensitive to the grueling nature of this work. These employees complain regularly about the unrelenting pace of work but to no avail; they feel as if they are treated as robots (Guendelsberber, 2019). With algorithms scheduling practically every movement of these persons, based on rationality and efficiency, their humanity seems to be drained.

Safiya Umoja Noble (2018) calls this maneuver as the creation of “algorithms of oppression” to highlight the technological redlining produced by these devices. As she argues, “part of the challenge of understanding algorithmic oppression is to understand that mathematical formulations to drive automated decision are made by human beings. While we often think of terms such as “big data” and “algorithms” as being benign, neutral, or objective, they are anything but (Noble, 2018:27, 18 left in chapter 0). Her point is that AI developers introduce numerous biases into AI

that are concealed by the illusion of disinterestedness associated with dualism and assumptions about objectivity.

Nonetheless, this technology is the product of human effort and has the related shortcomings. Because of dualism, however, this association is easily obscured or forgotten, and the knowledge linked to AI is able to eclipse other modalities. As a result, AI is able to colonize workplaces or other institutions where this technology is operative. The decisions that are made or the practices engendered by supervisors or administrators are recognized to be ultimately rational and effective. And because of the status of AI-generated decisions, all other options or challenges pale in comparison.

Because AI appears to be objective and divorced from local concerns, colonization occurs. In this context, merely striving for inclusion is no guarantee that any additional perspectives will not be marginalized. Reversing colonization will not occur this easily. What is required is that Cartesianism be challenged to the extent that knowledge supremacy is curtailed. Only in this way will the colonization brought about by AI be seriously confronted and possibly ended.

Why is inclusion insufficient to end colonization? Due to the influence of dualism, certain social or ethnic positions can be treated as universal and the centerpiece of a society. An asymmetry is thus established between these select persons or groups and anyone else who is introduced into this framework. Simply introducing more perspectives into this arrangement will not necessarily dislodge the center, but may only reinforce the inferiority of those who are newly introduced (Smith, 1988). For a more commodious situation to be established, one where dominance is curtailed, the exclusivity of any center must be undermined; a so-called space that can be invoked to protect claims of universality must be eliminated.

At fault is not just the bad or irrelevant logic of AI algorithms but the overall narrow view of the complexity of human endeavors, where algorithms “can’t be trusted with anything that hasn’t been precisely anticipated by their programmers” (Marcus & Davis, 2019:35). Examples of the resulting reductionism include Facebook’s M, IBM’s Watson at the MD Anderson Cancer Center, Google’s Duplex, Microsoft’s Tay, Amazon’s recruitment system, and Google “Bombs” and “Gorillas” (Marcus & Davis, 2019). Hence, besides colonialism, dualism, and biasing, the structures behind the development of algorithms have some AI-traps—coined by Marcus and Davis as the “AI-chasm”—that ought to be systematically challenged if human-centric AI tools are the expected outcome.

Marcus and Davis (2019) found three deep problems with the current approaches to artificial intelligence: gullibility, illusory progress, and robustness. The *gullibility* trap is when human characteristics are attributed to artificial intelligence because “the behavior of machines is often superficially similar to the behavior of humans” (Marcus & Davis, 2019: 40). The second or the *illusory* trap is “mistaking progress in AI on easy problems for progress on hard problems” (Marcus & Davis, 2019:43). The overpromising of AI always has been part of tech development. The third trap related to *robustness* is poignantly argued by Marcus and Davis (2019:47) when they write that “nobody will buy a home robot that carries their grandfather safely into bed four times out of five.” Their point is that almost everyone seems to be carelessly compliant about the success in some areas of AI development that have

demonstrated minimal effectiveness when dealing with humans, due to the reasoning power that is presumed to be available.

But why would not these traps be expected? Through this technology, persons almost expect to be carried to another dimension. In this dualistically sustained realm, reason prevails in every activity. Due to the effectiveness of this technology, persons learn to think better and analyze events and behavior more effectively. The autonomy of AI, for example, offers the promise of judgements that are not compromised by emotion or other distractions. In short, this technology does not have bad days and thus can elevate the human condition. In this context, even minimal gains are thought to be a sign of future greatness.

### 3 Participatory Design and Beyond

Decolonization efforts are centered on the idea of making AI more human-centric (Mhlambi, 2022). A key suggestion is that a range of persons should be consulted regularly in the process of designing algorithms and other facets of technology (Ehn, 2017). The ongoing mantra is that engineers and other professional should listen to stakeholders and follow their suggestions. The assumption is that the involvement of those who use or are affected by this technology will make AI more relevant and curb abuses.

The aim of this participatory design is to introduce humans and make AI design more human-centric (Durant, 1999; Hennen, 2012; Heigl et al., 2019). Indeed, Winograd (2006) about this point. The question is whether the usual approaches to participation go far enough to be human-centric. That is, recognizing local knowledge and providing the opportunity for periodic input from users is not the same as giving this pool of information the latitude to direct an AI project in conjunction with community actors. Most often, participation takes the form of periodic consultations, with projects controlled by a principal, professional director (Ezekilov, 2011).

Some writers contend (Slota, 2020; Winograd, 2006), however, that consultations are not enough to reduce the harm caused by AI. Their point is that those affected by AI are more than important, and any serious correctives to this technology will require more than periodic input from these persons. What is needed, instead, is a more community-based approach to the design of AI.

Community-based strategies go beyond participatory design, in that participation is not the focus of attention. Likewise, the aim is not to expand the culture of a workplace, so that employees feel comfortable offering their opinions. For example, a change in leadership style or organizational socialization does not necessarily lead to community-based design. Accordingly, and along these lines, a community-based approach treats inclusion as far more than a practice to obtain buy-in or support for an AI project (Lepri et al., 2021). With respect to a community-based approach, the basic idea is to take seriously local knowledge and have this collection of information guide a project. Simply put, inclusion is not sufficient.

Two principles guide community-based work. The first sounds quite simple—local knowledge is important in creating relevant AI. But with a move



away from Cartesianism, an entirely new way of thinking about knowledge is proposed. Specifically, the sub-text is that persons are understood to create knowledge based on how they interpret their situations. In this way, as phenomenologists say, they create a “stock of knowledge” that is used to make decisions, evaluate situations, and plan actions (Berger & Luckmann, 1966). The result is that persons produce their respective realities, not alone but in concert with others. A community or organization, therefore, consists of various realities that are coordinated, often for a limited time, so that certain tasks can be completed (Weick, 2009; Murphy, 2014).

What this shift from dualism means is that AI is never merely introduced into an organization. Instead, several, possibly competing, realities are present that must be successfully navigated, if an AI system is not going to colonize these knowledge bases. For this reason, employee participation is not sufficient to liberate them from this condition. What is necessary is that the input of persons be interpreted correctly, specifically in terms of the realities that they bring to bear on AI. In other words, they should not merely be consulted but engaged in dialogue.

The second principle is that local persons should control all interventions, such as the introduction of AI into an organization. In this regard, emphasis has been placed on employee participation in the design of AI systems (Bodker, 1996). This participation is usually divided into levels, that is, managers and line employees. Input from each group is valued, at least on paper. In actual practice, the reality is somewhat more complex. As might be expected, managers are invited without question to participate in design. On the other hand, the debate continues about the degree of non-professional involvement. Nonetheless, there is little discussion of local knowledge guiding AI design and non-professionals controlling this process (Bodker, 1996; Deininger et al., 2019). This kind of restricted participation goes to the heart of the need for dialogue.

Hans-Georg Gadamer (1996) remarks, as someone who has written a lot in this topic, that dialogue is more than a discussion, a transaction, inclusion, or the exchange of information that occurs with a consultation. Dialogue, instead, requires world entry, so that persons are understood in their own terms. The attempt must be made, in other words, to enter the narratives that persons have invented, for example, about themselves, others, and their situations. The idea is that those stories provide crucial insight into how they perceive AI and want to deal with this technology. Unique styles of reasoning and evaluation, for example, may be revealed, along with novel plans of action.

In this sense, persons do not simply have different opinions of a common reality but varying “lifeworlds,” or knowledge bases that contain norms, values, and commitments that are shaped by interpretation (Schutz & Luckmann, 1973). This realization leads to the second principle of community-based work that requires local persons to take the lead, or control, of all interventions. In the case of AI development, the point is to have a team of persons, and not only engineers and other experts, contributing to an AI system. Specifically important is that local knowledge and guidance are not supplementary but central to this process (Adler and Winograd, 1996). In this way, a new role is allotted to regular employees and others who may be affected by AI.



## 4 FlourishingAI

The actual practice of this community-based method is part of a project, referred to as “FlourishingAI” [Organizational Flourishing and Augmented intelligence<sup>3</sup>], that is currently underway at the University of Miami (USA) and University Areandina (in Colombia). This project grew from the criticisms lodged over the years about the alienating effects of technology and is now directed at AI (Heidegger, 1987; Jonas, 1982; Livingstone, 2018; Winner, 1977). Since AI is not likely to disappear, this project is devoted to producing a more humane version.

In many companies, AI is being introduced in a variety of jobs. In fact, according to “2021” Global Adoption Index,” many businesses are currently using AI and exploring the use of this technology (Global and Adoption Index, 2022) . Examples include a host of secretarial tasks that are deemed appropriate for AI application. Many labor-intensive tasks can be regulated by this technology, such as quality control. And even managerial tasks related, for example, to scheduling, forecasting, monitoring, and decision-making can be undertaken by AI.

When a company decides to develop an AI system, several steps are proposed to generate Flourishing AI.<sup>5</sup> The first is that personnel are queried about the organization and AI by using a modified procedure developed by Chris Argyris (2010).<sup>6</sup> The purpose of this initial assessment is to understand the setting, but equally important is to prompt these persons to reflect on what they are saying. They should begin to recognize that what they say is never obvious and needs to be interpreted correctly, if dialogue is to occur.

The next step consists of furthering this reflection. The point at this juncture is to enter the worlds of these persons and thus the organization. This reflection, for example, allows for language use to be interrogated and clarification reached that is consistent with the lifeworlds that are present. The organizational translator or OT that emerges has the time and space to do “play backing,” so that deep meanings can emerge and a human-centric dialogue take place (Colaizzi, 1978).<sup>7</sup>

Once this insight is achieved, the third stage proceeds—that is, real dialogue about the creation and use of AI in the organization. This mode of dialogue, informed by the present lifeworlds, allows for bureaucratic niceties and managerial ideologies to be by-passed, so that persons can speak in their own terms and be understood as they intend. The goal is that these stages lead to discussions that culminate in the creation and deployment of a human-centric, FlourishingAI.

---

<sup>5</sup> AI refers to Augmented Intelligence, which relates to the integration of human intelligence powered and “augmented” by artificial intelligence. Through this terminology, the point is to advance an image of technology that is used by humanity and not the other way around (Russell 2019).

<sup>6</sup> Chris Argyris and Donald Schön created a social methodology in the 1980s to contrast deep beliefs with the discourse of human beings. They called this method “The Left Column.” As part of the project on “FlourishingAI,” some innovations were made to this methodology. Now called “Situational Analysis,” this strategy has four additional columns, one for individual reflection, one for situational emergence, and two more for playacting.

<sup>7</sup> Playacting is another term for “member check” (Colaizzi 1978). Both of these terms refer to an iterative process between persons designed to achieve the dialogue proposed by Gadamer.

Another key facet of the FlourishingAI project is that a training-of-trainers model is used when engaging employees in dialogue (Mormina & Pinder, 2018). In this framework, the organizational translators learn from their experiences to continue the dialogue that was started by the outside initiators. The point is to stress that dialogue is based on openness and reflection that should not be terminated. To suggest that dialogue be used simply to unearth specific insights and end would be to distort this activity (Gadamer, 1996).

What should be noticed is that this process is time-intensive and must be gradually established. For this reason, most consultants shy away from this approach. As a result, dialogue is not achieved, and after a few consultations, an AI system is unveiled that is disconnected from most employees and other users. Based on initial observations, employees find the strategy of FlourishingAI satisfying and have an optimistic view of their relationships to AI and its likely impact (Murphy and Largacha-Martínez, 2021). This insight is interesting, since many employees are wary initially of AI and how this technology will change their jobs (Walsh, 2018).

Of course, many factors and not merely reflection and dialogue may be affecting how these employees perceive and approach AI. Ever since the research that led to the founding of the Human Relations School, attributes and behaviors are known to be influenced by a myriad of factors in an organization (Gillespie, 1991; Merrett, 2006). Nonetheless, if employee participation is considered to be helpful in the production of human-centric AI, then dialogue is certainly consistent with this outlook. A community-based strategy, therefore, deserves serious consideration.<sup>8</sup>

## 5 Conclusion

Proponents of the decolonization of AI agree that this process will not be successful until persons are brought into the design of this technology, particularly the algorithms. For this end to be achieved, the autonomy of AI must be challenged.

However, a philosophical maneuver is necessary for this change to occur that is currently overlooked in discussions about curtailing the colonization coordinated by AI. Employee knowledge creation and control are essential to this decolonization.

To put this philosophy in terms familiar to AI developers, questions have to be raised about the future of the black box that is AI. This facet of AI must be taken more seriously by the design community. Brian Christian (2020:12–13) named this issue the “*sorcerer’s apprentice*” conundrum, where “we conjure a force, autonomous but totally compliant, give it a set of instructions, then scramble like mad to stop it once we realize our instructions are imprecise or incomplete—lest we get, in some clever, horrible way, precisely what we asked for.” What Christian is raising is the alignment problem, which focuses on whether the values, beliefs, and other

---

<sup>8</sup> Whether or not this community-based strategy can be used outside of organizations is an interesting question. Readers who are concerned about the application of this strategy to communities should consult the following: Murphy et al., (2022). “Introduction: Participatory Budgeting as Community-based Work”, *American Behavioral Scientist*, doi.org/10.1177/00027642221086952.

sentiments exhibited by users are at the core of AI. This problem has been noted for some time, with little or no changes in the ways that AI is designed. Something more encompassing is needed.

While recognizing the importance of persons in the construction of AI, participatory design may not lead to the demise of colonization. Here is where the advances of a community-based AI become important. In this strategy, there is no equivocation about the meaning of participation. That is, local knowledge and control are the centerpiece of this approach to AI design. There is not simply participation or the solicitation of input but a focus on the guiding role of local knowledge and local control of the design, implementation, and evaluation of an AI system. In this way, as Christian (2020) notes, the black box can be deconstructed and rendered public.

Predicated on the basic principles of a community-based AI, the fear of colonization by this technology may abate. Indeed, persons may no longer be pawns in the practice of AI. A new and less threatening status is associated with AI that is subordinate to human action. To borrow from Albert Memmi (1967), the colonizer is now colonized. AI is no longer an adversary, but recognized to be an extension of human action without any autonomy. At least in theory, persons are liberated from dominion by this technology. Practice, however, is where challenges seem to arise.

But without good theory or philosophy, practice is doomed.

**Author Contribution** Authors John W. Murphy and Carlos Largacha-Martínez declare that they have done all the research needed for this scientific article.

**Funding** This article was supported by a Fellowship given by the Fulbright Commission (USA) and the Ministry of Science (Colombia) to Carlos Largacha-Martínez, as part of the “Visiting Scholar” Program, 2020–2021 cohort. Professor John W. Murphy acted as the representative of the host institution, the University of Miami.

### Declarations

All the ideas expressed here do not compromise any of the people that helped us, neither the networks that supported us, neither the Fundación Universitaria del Área Andina, University of Miami, nor the Fulbright Commission (US-Colombia).

**Ethics Approval, Consent to Participate, and Consent to Publish** Authors John W. Murphy and Carlos Largacha-Martínez declare that for this research, they did not interview any human being nor have any relation to animals, so no ethical approval was required, neither consent to participate, neither consent to publish.

**Conflict of Interest** The authors declare no competing interests.

## References

- Adler, P. S., & Winograd, T. (1996). The usability challenge. In P. S. Adler & T. Winograd (Eds.), *Usability: Turning Technology into Tools* (pp. 3–14). Oxford University Press.
- Aho, J. (2008). *Body Matters: A Phenomenology of Sickness, Disease, and Illness*. Lexington Books.
- Argyris, C. (2010). *Organizational Traps: Leadership, Culture, Organizational Design*. Oxford University Press.
- Appadurai, A. (2001). *Globalization*. Duke University Press.



- Marcus, G., & Davis, E. (2019). *Rebooting AI: Building Artificial Intelligence We Can Trust*. Pantheon Books, Apple Books.
- Memmi, A. (1967). *The Colonizer and the Colonized*. Beacon Press.
- Memmi, A. (1968). *Dominated Man*. Orion Press.
- Merrett, F. (2006). Reflections on the Hawthorne effect. *Educational Psychology*, 26(1), 143–146.
- Mhlambi, S. (2022). *HAI seminar: Decolonizing AI*. Stanford University, January 26, 2022. [https://Stanford.Zoom.us/webinar/register/wnt7BevycyQ\\_aD9\\_gdrEF9Hg](https://Stanford.Zoom.us/webinar/register/wnt7BevycyQ_aD9_gdrEF9Hg)
- Mitchell, M. (2019). *Artificial Intelligence: A Guide for Thinking Humans*. Farrar, Straus, and Giroux.
- Mohamed, S., Png, M.-T., & Isaac, W. (2020). Decolonial AI: Decolonial theory as socialtechnical foresight in artificial intelligence. *Philosophy and Technology*. <https://doi.org/10.1007/s13347-20-00405>
- Mohamed, S. (2018). Decolonial artificial intelligence. *The Spectator*, October. [blog.shakirm.com/2018/10/decolonizing-artificial-intelligence](http://blog.shakirm.com/2018/10/decolonizing-artificial-intelligence). Accessed Jan 2022.
- Mormina, M., & Pinder, S. (2018). A conceptual framework for training of trainers (ToT) interventions in global health. *Global Health*, 14, 100. <https://doi.org/10.1186/s12992-018-0420-38>
- Murphy, J.M. (2014). *Community-based interventions: Philosophy and action*. Springer.
- Murphy, J. W., Evans, S. D., Minutti-Meza, M. A. (2022). Introduction: Participatory budgeting as community-based work. *American Behavioral Scientist*. <https://doi.org/10.1177/00027642221086952>
- Murphy, J. W., & Largacha-Martínez, C. (2021). “Is it possible to create a responsible AI technology to be used and understood within workplaces and unblocked CEOs’ mindsets. *AI and Culture*. <https://doi.org/10.1007/s00146-021-01316-8/152403>
- Nancy, J.-L. (2016). *The Disavowed Community*. Fordham University Press.
- Noble, S. U. (2018). *Algorithms of Oppression*. NY University Press. NY.
- O’Neil, C. (2016). *Weapons of Math Destruction*. NY Crown.
- Russell, S. J. (2019). *Human Compatible: Artificial Intelligence and the Problem of Control*. Penguin.
- Schutz, A., & Luckmann, T. (1973). *The Structures of the Lifeworld*. Northwestern University Press.
- Slota, S. C. (2020). *Designing across distributed agency: Values, participatory design, and building socially responsible AI*. <https://repositories.lib.utexas.edu/bitstream/handle/2152/82848/c42d8459-27b5-4d3b-b9d9-a227536b3999%20%281%29.pdf?sequence=2&isAllowed=y>. Accessed Jan 2022.
- Smith, B. H. (1988). *Contingencies of Value*. Harvard University Press.
- Walsh, T. (2018). Expert and non-expert opinion about technological unemployment. *International Journal of Automation and Computing*, 15(5), 637–642.
- Wei, X. U. (2019). Toward human-centered AI: A perspective from human-computer interaction. *Interactions* July-August. [Interactions.amc.org/archive/view/July-August-2019/toward-human-centered-ai](http://Interactions.amc.org/archive/view/July-August-2019/toward-human-centered-ai). Accessed Jan 2022.
- Weick, K. E. (2009). *The Impermanent Organization*. Wiley.
- Winner, L. (1977). *Autonomous Technology: Technics-Out-of-Control as a Theme in Political Thought*. MIT Press.
- Winograd, T. (2006). Shifting viewpoints: Artificial intelligence and human–computer interaction. *Artificial Intelligence*, 170(18), 1256–1258. <https://doi.org/10.1016/j.artint.2006.10.011>
- Zhang, B. & Dafoe, A. (2019). *Artificial Intelligence: American Attitude and Trends*. Oxford University: Center for Governance of AI/Future of Humanity Institute. <https://ssrn.com/abstract=3312874>. Accessed Jan 2022.

**Publisher’s Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.