Revista Jurídica unicuritiba

Submetido em: 15/12/2024
Aprovado em: 23/12/2024
Avaliação: Double Blind Reviewe
ISSN: 2316-753X

ALGOCRACY, ALGORITHMIC INSTITUTIONALISM, DIGITAL RATIONALITY AND RISKS TO DEMOCRACY

ALGOCRACIA, INSTITUCIONALISMO ALGORÍTMICO, RACIONALIDADE DIGITAL E RISCOS À DEMOCRACIA

PAOLA CANTARINI
Advogada, professora universitária, bacharel em direito, mestre e doutora em direito, em filosofia (PUCSP) e em filosofia do direito (Univ. Salento/Unisalento-It.), pós-doutora (EGS/Suíça, Univ. Coimbra, PUC SP – TIDD, USP /IEA – Cátedra OS); pesquisadora visitante (SNS - Scuola Normale Superiore - Pisa/It, Florida International University, European University Institute, Law Department, e CIJIC (Centro de Investigação Jurídica no Ciberespaço) – Univ. de Lisboa). Pesquisadora do C4AI - Centro de Inteligência Artificial (USP). Presidente e Pesquisadora da EthikAI – "ethics as a service" (ethikai.com.br). Membro da Comissão da Criança e do Adolescente, da Comissão de Proteção de Dados da Ordem dos Advogados de São Paulo e da Comissão de IA (OABSP- Santo Amaro) e ex membro da Comissão de IA da OAB/MG. ORCID: orcid.org/0000-0002-9610-8440. Pós-doutoranda USP com bolsa Fapesp. Academia EDU:https://independent.academia.edu/PaolaCantarini Lattes: CV: http://lattes.cnpq.br/5057397566570034

ABSTRACT:
This article aims to provide a critical perspective through an interdisciplinary, inclusive, and decolonial analysis of the theory of algorithmic institutionalism, addressing problematic issues related to artificial intelligence—especially how AI could be made democratic. It also brings forward critical reflections on the concepts of algocracy, algorithmic institutionalism, digital rationality, and democracy, based on the work *"Algorithmic Institutionalism: The Changing Rules of Social and Political Life"* by Ricardo Mendonça, Virgilio Almeida, and Fernando Filgueiras. The article contributes to the debate on the interaction between the development of artificial intelligence and the democratic system.

Objectives: The article aims to offer a critical perspective through an interdisciplinary, inclusive, and decolonial analysis of problematic issues surrounding artificial intelligence, avoiding utopian or dystopian approaches and promoting critical theories. It also seeks to expand Brazil's participation in the scientific and academic discussion on AI within the humanities, particularly from the Global South, which is one of the essential pillars for addressing epistemic justice. Furthermore, it seeks to engage with the authors of the book *"Algorithmic Institutionalism: The Changing Rules of Social and Political Life"*, highlighting both the strengths and weaknesses of the proposed theory in order to contribute to the scientific debate.

Methodology: An analysis of the state of the art on central themes of AI, especially issues regarding the lack of consent, authorization, and legitimacy in AI applications, the relationship between AI and democratic perspectives, and the main arguments of algorithmic institutionalism theory.

Submetido em: 15/12/2024
Aprovado em: 23/12/2024
Avaliação: Double Blind Reviewe
ISSN: 2316-753X

Results: The article presents important reflections aimed at rethinking the possible paths toward the democratization of AI. It analyzes the proposal of multistakeholder governance and questions whether algorithms can truly be considered institutions. It also reflects on issues of consent, authorization, and legitimacy, as well as on power and the potential for resistance in the face of such power, drawing on the proposals of Foucault, Antoinette Rouvroy, and Thomas Berns, particularly their concept of algorithmic governmentality.

Contributions: The article offers significant insights for rethinking potential pathways toward democratizing AI. It critically analyzes the proposal of multistakeholder governance and questions whether algorithms could be considered institutions—similar to social institutions such as marriage, the Church, or even the law. Given the opacity and lack of transparency in algorithmic systems, and the frequent absence of awareness that one is subjected to an algorithm (along with the lack of control and accountability), it would not be accurate to claim that individuals are consciously learning and accepting the values imposed unilaterally. Some preliminary conclusions suggest that it may be more appropriate to speak of a *process* of algorithmic institutionalization, rather than asserting that algorithms already constitute institutions or are fully institutionalized—depending, of course, on the definition of "institution." After all, if institutions are those structures that ensure social coexistence and therefore deserve to be protected, how can we regard as such something that still poses significant threats, especially given numerous cases where AI applications have caused harm to fundamental rights?

KEYWORDS: Algocracy, algorithmic institutionalism, digital rationality, and democracy.

RESUMO

O presente artigo visa fornecer uma perspectiva crítica em uma análise interdisciplinar, inclusiva e decolonial acerca da teoria do institucionalismo algorítmico envolvendo questões problemáticas da inteligência artificial em especial, como se poderia ter uma IA democrática, trazendo, outrossim, problematizações acerca dos conceitos de algocracia, institucionalismo algorítmico, racionalidade digital e democracia, a partir da obra “Algorithmic Institutionalism: The Changing Rules of Social and Political Life” de Ricardo Mendonça, Virgilio Almeida e Fernando Filgueiras, contribuindo para o debate em torno da interação entre o desenvolvimento da inteligência artificial e o sistema democrático.

Objetivos: O artigo tem como objetivo fornecer uma perspectiva crítica em uma análise interdisciplinar, inclusiva e decolonial acerca de questões problemáticas em torno da inteligência artificial, evitando abordagens utópicas ou distópicas e promovendo teorias críticas, ampliando, outrossim, a participação do Brasil na discussão científica e acadêmica envolvendo a temática da IA na área das humanidades. E, pois, do sul Global, o que seria um dos pilares essenciais para se falar em justiça epistêmica. Outrossim, busca dialogar com os autores do livro “Algorithmic Institutionalism: The Changing Rules of Social and Political Life”, apontando os pontos fortes bem como os pontos de maior fragilidade da teoria proposta, de forma a contribuir para o debate científico.

Metodologia: análise do estado da arte acerca de temas centrais da IA, em especial, problemáticas em torno da ausência de consentimento, autorização e legitimidade de

aplicações de IA, relação da IA com a perspectiva democrática, bem como acerca dos principais argumentos da teoria do institucionalismo algorítmico.

Resultados: O artigo traz importantes reflexões a fim de se repensar quais os possíveis caminhos em prol da democratização da IA, analisando a proposta de governança multisetorial, questionando se de fato poderíamos pensar nos algoritmos como uma instituição, refletindo, outrossim, sob a questão do consentimento, autorização e legitimidade, bem como acerca do poder e da possibilidade de resistência em face de tal poder, abordando as propostas de Foucault, de Antoinette Rouvroy e Thomas Berns, ao cunharem a noção de governamentalidade algorítmica.

Contribuições: O artigo traz importantes reflexões a fim de se repensar quais os possíveis caminhos em prol da democratização da IA, bem como analisando a proposta de governança multisetorial, questionando se de fato poderíamos pensar nos algoritmos como uma instituição, a exemplo de instituições sociais, como o casamento, a Igreja ou até mesmo a lei, já que diante da opacidade, da falta de transparência dos algoritmos e da total ausência em diversos casos de consciência de que se está submetido a um algoritmo, ausência de controle e de responsabilização, não seria possível falar em indivíduos estarem aprendendo e aceitando os valores impostos de forma unilateral. Algumas das conclusões provisórias seriam no sentido de ser mais adequado se falar antes de um processo de institucionalização dos algoritmos do que propriamente afirmar que sejam já instituições e estejam institucionalizados - a depender do conceito de instituição, claro, pois se elas forem o que nos garante a convivência social e, por isso, merecem ser defendidas, como considerar assim o que ainda nos ameaça, diante de vários casos de danos a direitos fundamentais em razão de aplicações de IA?

PALAVRAS-CHAVE: Algocracia, institucionalismo algorítmico, racionalidade digital e democracia

1 INTRODUCTION

Although AI is generally associated with neutrality and objectivity, often perceived as apolitical and ahistorical, numerous studies and reports highlight the existence of algorithmic bias, racism, and discrimination. These biases perpetuate existing inequalities through an intersectional lens, encompassing race, gender, and class. They also introduce inherent challenges related to contestability and control over personal data—particularly in the context of profiling and algorithmic decision-making. These challenges are exacerbated by the opacity, inaccessibility, and lack of transparency in algorithmic processes, as well as the fragility of legal frameworks authorizing the processing of personal data.

Several authors have pointed to the threat that AI poses to democracy, its principles, values, and foundational norms (Mark Coeckelbergh, 2024), due to its potential to intensify polarization, intolerance, and hate speech through the spread of

Submetido em: 15/12/2024
Aprovado em: 23/12/2024
Avaliação: Double Blind Reviewe
ISSN: 2316-753X

misinformation and fake news—conditions that could enable the rise of totalitarian regimes.

There is a robust and ongoing debate in the literature concerning AI's impact on democracy, typically framed by two opposing perspectives. One perspective argues that AI can contribute positively by supporting democratic values through the improvement of public service delivery, fostering citizen trust, and expanding economic opportunities across industries and markets. The opposing view asserts that AI fundamentally threatens democratic life (SIMONS, J., 2023).

Some scholars have emphasized the need to focus attention on the Global South, reviving the notion of ethnocentrism and associating it with concepts such as "data colonialism," "technological imperialism," and "algorithmic coloniality" (ROCHE et al., 2023, p. 1099). These terms critique the tendency of AI systems to interpret other cultures through the lens of dominant Western paradigms, thereby reinforcing the existing inequalities between North and South. This dynamic contributes to the imposition of Northern ideologies and philosophies upon the South in a vertical relationship, undermining global democratization processes and perpetuating colonial power structures.

There appears to be an intrinsic relationship between state fragility and democratic resilience. Unstable regimes with weak institutions and political turbulence are more vulnerable to the risks posed by AI. Part of the literature argues that AI has emerged from an epistemic echo chamber, where bias is a structural feature—one rooted in white supremacy and developed through a layered network of interconnected systems.

In light of this, questions arise regarding the recognition and protection of Indigenous peoples' right to self-determination in an AI-dominated environment. Indigenous communities are affected by both the development and deployment of AI, with some scholars referring to AI as a new form of colonization. This mirrors broader critiques of developments in biotechnology, which often depend on Indigenous resources and have been labeled as "biocolonialism."

AI, therefore, reflects the specific values and ideals of the Western scientific worldview. It is shaped by the ideology and assumptions of programmers trained within that paradigm, which is grounded in a reductionist data ontology and an artificial epistemology of algorithms—algorithms that prioritize the efficiency of task execution over the morality of those tasks.

This concern is echoed by the authors of the aforementioned book, as illustrated in the following excerpts:

> "Indeed, social media algorithms can influence people's preferences in various contexts, from elections to romantic encounters. A study of social media platforms reveals that some algorithmic systems directed users toward radical and extremist content, increasing their political bias and contributing to social polarization."
> (...)
> "For example, when platforms like Facebook or YouTube recommend content based on user behavior, they not only shape individual preferences and behaviors but also have broader social implications. These include potentially reinforcing echo chambers, polarizing viewpoints, or even spreading misinformation." (op. cit., p. 22 ff.)

In this regard, the authors argue that regulation alone may be insufficient to prevent hate speech and the dissemination of fake news. They propose a path toward democratization through multistakeholder governance:

> "For instance, hate speech, disinformation, or fake news continue to be disseminated on social media, even when regulatory changes create additional layers of control. The corporations controlling these platforms are not fully capable of altering the foundational business model decisions that underpin them, as hate speech and misinformation are embedded in the way these algorithms operate. Therefore, content regulation on such platforms is inherently and continuously path-dependent." (op. cit., p. 82 ff.)

2 POWER RELATIONS AND ALGORITHMIC GOVERNMENTALITY

When it comes to power relations—a central element in the theoretical framework of Algorithmic Institutionalism—the authors examine a range of definitions of power, with an emphasis on the views of Thomas Hobbes, Max Weber, Bachrach, Baratz, Lukes, John Danaher, Michel Foucault, among others. They argue that, "from a Hobbesian perspective, the power of algorithms lies in a more coercive structure, leading to something akin to algocracy," a term coined by A. Aneesh.

They identify the main features of algorithmic power as its "coercive nature, its capacity to influence individual decisions by narrowing the available alternatives, its effects on public agenda setting, its consequences on individual preferences, desires, and self-conceptions, and its broader social implications, shaping social relations and the rationalities that guide social life."

Submetido em: 15/12/2024
Aprovado em: 23/12/2024
Avaliação: Double Blind Reviewe
ISSN: 2316-753X

Algorithmic power should be understood as a set of relational emergences, distributed—often tacitly—throughout the social structure. Moreover, algorithmic power must be seen as both enabling and constraining, insofar as it inhibits certain behaviors and social outcomes while promoting others.

It is worth delving further into the analyses offered by Foucault, and by Antoinette Rouvroy and Thomas Berns, who coined the notion of algorithmic governmentality.

According to the authors, Michel Foucault conceptualizes power as pervasive in nature, shaping the structures that reproduce the patterns giving meaning to action, with an emphasis on the generative forces of power. They explain that “power is not merely a form of constraint; it must be seen as a productive network that runs through the entire social body.” Power is not episodic: it transcends and embodies the social body in discourse, knowledge, and regimes of truth.

As for the Foucauldian concept of governmentality, they interpret it as a rationality based on institutions, procedures, strategies, techniques, and practices used to manage populations—something Foucault viewed as the core of contemporary institutions. Algorithms function as instruments of population governance, expanding the possibilities of governmentality. Rouvroy and Berns’ notion of algorithmic governmentality is based on a rationality that extracts knowledge from individual behaviors, accumulating it to manage populations not through statistical averages, but through profiling based on the relationships between data points.

According to Rouvroy and Berns, the advancement of algorithmic governmentality fuels the expectation of overcoming subjectivity, proposing a model for understanding individuals through their relations—which become the primary target of surveillance and categorization by self-learning algorithms. The focus on relationality in these models echoes the philosophical proposals of thinkers like Simondon, Deleuze, and Guattari, who saw in relational configurations a path toward transindividual and rhizomatic formations with emancipatory potential. While there are points of convergence, key differences remain, particularly regarding the role of “difference,” disparity, and “failure” as fundamental elements in the creation of new relational worlds (Rouvroy & Berns, 2015).

Still, the authors could have expanded their analysis by engaging more deeply with Foucault’s work on the panopticon and its developments concerning the 21st-century digital panopticon. In this digital iteration, surveillance is no longer exerted by

Submetido em: 15/12/2024
Aprovado em: 23/12/2024
Avaliação: Double Blind Reviewe
ISSN: 2316-753X

a central gaze or despotic omnipotence, as described by Foucault, but rather, everyone monitors everyone else. Byung-Chul Han reflects on this transformation by expanding the concept of the "homo sacer," arguing that we are now transparent clients—the new inmates of the digital panopticon (Han, 2017, pp. 29–30), as our lives are permanently exposed to monitoring.

Our behavior is constantly observed and exposed within a sort of glass house in this "transparent society," raising serious concerns about our freedom. The inmate of the digital panopticon is simultaneously the aggressor and the victim—this is where the dialectic of freedom, which manifests as control, becomes evident.

Gilles Deleuze, in turn, introduced the concept of the "society of control," emerging in the second half of the twentieth century—a society built upon a digital language of control, composed of codes that grant or deny access to information (Deleuze, 1992; 2013). He wrote: "There is no need for science fiction to imagine a mechanism of control that provides the position of any element in an open space at any given moment: an animal in a reserve, a human in a company (with an electronic collar). Félix Guattari envisioned a city where each person could leave their apartment, street, or neighborhood using an electronic card that would remove any barrier; however, that same card could also be rejected on certain days or during certain hours. What matters is not the barrier, but the computer that identifies one's location—licit or illicit—and operates a universal modulation."

This concept is deeply connected to the crisis of the welfare state and the transformation of capitalism, particularly after the 1974 oil crisis, which prompted an exponential increase in market control through the creation of new surveillance and defense technologies, ultimately altering how we relate to technology and how new subjectivities are formed.

The book also analyzes algorithms from the perspective of discursive institutionalism, asserting that "algorithms can be interpreted through the lens of discursive institutionalism in at least four ways, as algorithmic experiences are always discursive in some form."

However, there is a certain fragility to this approach if considered in light of Jürgen Habermas's theory of communicative action, which underscores the importance of dialogue and public deliberation as essential to the legitimation of power. For Habermas, democracy must be grounded in a deliberative process through which citizens, via communication and debate, reach legitimate political decisions.

Submetido em: 15/12/2024
Aprovado em: 23/12/2024
Avaliação: Double Blind Reviewe
ISSN: 2316-753X

Byung-Chul Han further contends that contemporary society suffers from an absence of rituals and, therefore, from a loss of memory—something that, for Foucault, is an essential value of resistance. “Verbis”: “Since memory is an important factor of struggle (indeed, it is within a kind of conscious historical dynamic that struggles unfold), if the people’s memory is preserved, their dynamism is preserved. And so is their experience, their knowledge of previous struggles.” In the era of dataism, living memories are transformed into dead memories—machine-like—and the possibilities for resistance are thus weakened or even erased.

We live in the age of “infomania” (Byung-Chul Han). In such a phase, the excess of information reduces our capacity for reflection and knowledge, both of which require time, while we are imprisoned by the frantic and exponential speed of the present.

According to Deleuze, control societies operate “through third-type machines—information machines and computers—whose passive danger is interference and whose active danger is piracy and the insertion of viruses.” These dynamics are characteristic of a transformation in capitalism. Byung-Chul Han offers a critique of transparency as an ideology—fetishized total transparency—which results in a society of transparency and control through indiscriminate and excessive surveillance. This transforms everyone into potential suspects, legitimizing itself through a culture of voluntary self-exposure as an economic imperative (Deleuze, pp. 43 ff.).

A society without the Other is a society without “Eros”—one of the consequences of the digital transparency society, where everything and everyone is constantly displayed in a liquid state of happiness, dominated by the absolutization and totalization of visibility. In such a society, the pornographic prevails over the erotic, since the latter thrives on secrecy, silence, and the unrevealed. Eroticism emerges from an excess of signifiers that circulate without being exhausted by their meanings. Today, we witness the erasure of rituals—symbolic acts grounded in symbolic perception—whose purpose is to transmit values and orders that serve as the foundation of a community. Rituals serve recognition, which is a specific form of repetition. Recognition is a lasting symbolic perception. However, with the digitalization of the world, this permanence is lost, as data and information lack symbolic force and cannot sustain recognition. The experience of duration is diminished—everything becomes instantaneous and consumable. Rituals no longer stabilize our lives; everything is disposable, designed not to last, because this feeds production and consumption.

Submetido em: 15/12/2024
Aprovado em: 23/12/2024
Avaliação: Double Blind Reviewe
ISSN: 2316-753X

Seduction is bound to rituals—it is a continuous ritual exchange requiring duration—and thus it has no place in a society where rituals and “Eros” are dying. Seduction is the essence of eroticism, which demands chance—like a dice game—rather than certainty or calculation. We can no longer hear the voice of the Other amidst the digital noise of the same; within the digital echo chamber, one hears only oneself, and the voice of the Other is increasingly silenced. Relationship requires sight, voice, the Other—it is a sensory and corporeal experience. “Eros” is being erased in our society of performance and positivity, along with rituals and memory, and we have become self-centered beings, with no negativity of the Other to contrast ourselves against. For Byung-Chul Han, we are living in a post-immunological era—a surplus of positivity and sameness that annihilates negativity and difference. The Other, who would allow our bodies to produce antibodies, is absent. This may also partly explain why we have become so vulnerable to pandemics. The immunological system has been replaced by the neuronal system, and the defining illnesses of our time are depression, attention disorders, and burnout.

As Paul Celan (2011) states, the voice that comes from the Other is also constitutive of literature, which begins where language becomes “vocal.” The world is becoming voiceless, sightless, and without literature. The digital represents an excess of meaning with no corresponding signifiers, whereas seduction represents an excess of signifiers that cannot be reduced to meaning.

With the exclusion of the Other comes the death of difference, and as thought turns toward the calculation of probabilities, everything immeasurable, unclassifiable, incalculable is excluded—reducing the value of difference to negativity. Paul Virilio argues that the computer ends up replacing the human eye with the optical—with the lens and with technology—causing the elimination of perception, which requires sight and engagement. With that elimination, reality is also erased (Lotringer & Virilio, 2005).

Language and our very essence are tied to the Other and to difference. Today, everyone is under suspicion as digital transformations reshape language and our subjectivities.

With the total gamification of the metaverse—as imagined in Stanislaw Lem’s futuristic congress—we are on the verge of achieving the “perfect crime”: the death of death itself. This is the scenario Jean Baudrillard predicted: the rise of the “homo deus,” perfect and immortal, even at the cost of life itself. In this vision, we fundamentally transform our being, which, according to Heidegger, is characterized as being-toward-

death and open-ended existence. In capitalism's struggle for a life without death, we create a necropolis (Byung-Chul Han)—an antiseptic space of death, cleansed of human sounds and scents. The processes of life are transformed into mechanical procedures, fully adapting human existence to mere functionality.

3 TOWARD THE DEMOCRATIZATION OF ALGORITHMIC INSTITUTIONS

The authors argue that the path toward the democratization of algorithmic institutions must begin by first conceptualizing algorithms as institutions. Only from this point can a movement toward their democratization begin—one grounded in critical analysis emerging from public scrutiny of their political consequences. They propose two steps for this transformation:

> **1. A debate on the legitimacy of algorithmic decision-making.** It is also essential to reflect on the core values necessary for fostering more democratic institutions.

Thomas Hobbes, in his seminal work *Leviathan* (1651), argues that the legitimate power of the sovereign derives from a social contract in which individuals agree to transfer their natural rights to a central authority in exchange for security and order. In situations of escalating insecurity, this social contract is broken. Shoshana Zuboff, in *The Age of Surveillance Capitalism* (2019), highlights this rupture, asserting that Google's data collection practices broke the social contract. The use of personal data now disregards the principle of data minimization, violates the principle of transparency, and results in a behavioral surplus—a stark imbalance between data collection and processing and the declared purposes disclosed to data subjects. Zuboff also points to the Cambridge Analytica scandal and the events following the 2017 elections in both the United States and the United Kingdom, where fake news became a focal point. The company responded with public apologies, expressions of contrition, attempts at appeasement, and appearances before the U.S. Congress and the European Parliament, yet failed to implement any real operational changes to reduce economic incentives for misinformation or to establish systems that would warn users of probable information corruption.

Submetido em: 15/12/2024
Aprovado em: 23/12/2024
Avaliação: Double Blind Reviewe
ISSN: 2316-753X

Reflecting on the social contract from the perspective of John Locke's *Two Treatises of Government* (1689), legitimate power must be based on the consent of the governed and on the protection of natural rights—namely life, liberty, and property. In the case of algorithms, both of these requirements are lacking, and thus, it cannot be said that algorithmic power is legitimate.

The 2012 Facebook Emotional Contagion Experiment serves as a powerful example of this concern. Researchers from Facebook and Cornell University conducted a large-scale experiment on Facebook users, manipulating the emotional content of News Feed stories to determine whether emotionally biased content would influence users' own emotional expressions in subsequent posts. Given the scale of the study, there were no mechanisms in place to exclude vulnerable individuals (e.g., those with depression or psychological disorders), nor to measure individual impacts and mitigate potential harm. Furthermore, as the study imposed no age restrictions, minors may have been included without any knowledge or informed consent. Most Facebook users were—and remain—unaware that such filtering is even taking place. Even among those who know that their news feeds are curated algorithmically, the lack of transparency regarding what data is collected and how it determines relevance makes it nearly impossible to reason why certain stories appear and others do not.

> **2. Algorithms must be aligned with democratic principles**, integrating them into political dynamics oriented by values such as participation, equality, pluralism, accountability, transparency, public debate, and freedom.

Although the respect for the above democratic values is essential, the authors make no mention of social justice, respect for human dignity, or fundamental rights.

4 DISCURSE DIMENSION, OPACITY AND LACK OF TRANSPARENCY

The authors argue that algorithms can be understood as texts, stating:

> "It is essential to recognize that algorithms can be seen as texts. They are commands that establish decision-making patterns. Therefore, algorithms are discursive artifacts that participate in broader discursive interactions occurring in the public sphere. They can be interpreted as arguments expressing positions towards other actors. When used in broader contexts, algorithms become part of a social conversation, directly or indirectly influencing how individuals perceive and engage with social issues. Discursive Processes

Submetido em: 15/12/2024
Aprovado em: 23/12/2024
Avaliação: Double Blind Reviewe
ISSN: 2316-753X

> Around Algorithms. Furthermore, attention must be paid to the discursive processes involving algorithms. As Frank Fischer observes, a proper understanding of political processes requires attention to the discursive contestation surrounding them. These debates play a crucial role in shaping social problems, influencing how actors assign meaning to social phenomena and promote change."

Conversely, for Byung-Chul Han, communicative rationality is absent. Thus, we must consider how to restore the public sphere, given its disintegration and the erosion of communicative action, as also noted by Habermas. In the context of information overload, communication among people has been impaired, as the assumptions postulated by Arendt and Habermas rarely hold in practice. This is why Byung speaks of digital rationality—a form of rationality that sustains itself without discourse, as opposed to communicative rationality. In addition to the capacity for justification, the availability of learning is constitutive of communicative rationality. The concept of justification is intertwined with that of learning. Artificial intelligence does not justify—it calculates. Instead of arguments, we now face algorithms.

In automated decisions, users are bound to the outcomes of mathematical models, without an interpretive instance for decision parameters, as typically exists in judicial systems. The usual safeguards ensuring adequacy between means and ends, proportionality, or other criteria evaluated through the natural language of legal norms vanish in algorithmic judgments. Even if the parameters of the black box and its inner workings were disclosed, the outcomes would still reside in a realm distinct from natural language.

One consistently problematic issue is opacity, lack of transparency, the "black box" nature, and informational ubiquity. Transparency has thus been recognized as a core requirement in several AI ethics codes and frameworks for achieving responsible, trustworthy, and ethical AI.

However, it is important to distinguish among the different types of opacity. Jenna Burrell (2023) outlines three primary categories:

> **1. Intentional Opacity**: This is linked to protecting the intellectual property of algorithm creators. It can be mitigated through legislation that mandates decision-makers to use open-source systems, or through balancing opposing fundamental rights, since industrial secrecy grounded in market freedom is not absolute. This type of transparency relates to the disclosure of the system's source code—the mechanism responsible for analyzing input data and producing outcomes. Yet many scholars point out the technical limitations and ineffectiveness of this approach (Joshua Kroll, Nicholas Diakopoulos), citing the code's potential for manipulation, incomprehensibility, or posing a

Submetido em: 15/12/2024
Aprovado em: 23/12/2024
Avaliação: Double Blind Reviewe
ISSN: 2316-753X

national security risk when used by public authorities. A more viable alternative would be disclosing the data used to train the algorithm and implementing detailed pre-filtering.
2. Opacity Due to Lack of Literacy: Most people lack the technical skills and critical analysis necessary to grasp even the basic principles of algorithms and machine learning models, including profiling outcomes. This can be addressed through free, high-quality education programs designed to democratize knowledge and reduce informational inequality and power imbalances.
3. Intrinsic Opacity: This arises from the nature of certain machine learning methods that are inherently difficult to interpret (e.g., deep learning models), resulting in a broader problem of interpretability. It can be mitigated by using alternative models, imposing transparency obligations in decision-making or profiling, and requiring the prior implementation of AI Impact Assessments (AIAs).

In response to criticisms of open code and the mistaken perception that industrial secrecy is an absolute right, other notable proposals have emerged.

One such contribution comes from Joshua Kroll, who emphasizes the use of technological tools and suggests four technical mechanisms for verifying procedural regularity. These are intended to address the inefficacy of transparency and audits, arguing that systems should be designed from the outset to be controllable, as post-hoc control may prove ineffective. His four technical tools include:

1. Software Verification: Using mathematical logic to demonstrate that an algorithm has not functioned correctly. This may involve certification by a separate program to confirm system characteristics, or exhaustive testing to ensure variable integrity. One example is using supervised machine learning to monitor unsupervised systems.
2. Cryptographic Commitments: These involve providing a digital key that allows verification of the algorithm's consistency in decision-making parameters. However, they do not prevent information destruction nor explain system errors.
3. Zero-Knowledge Proofs: These build upon cryptographic commitments and allow verification that a decision-making policy possesses a certain property—without revealing the property itself. This ensures that valid assertions can be verified independently.
4. Fair Random Choices: In some cases, randomness is required. Mechanisms must exist to verify the randomness employed. One such method is using a predetermined input (seed), allowing the reproducibility of random values. These systems must be designed to enable impartial third-party control from the outset.

Another alternative is auditing, which involves examining the system's inputs and outputs—i.e., the data submitted and the results generated—without scrutinizing internal mechanics (the "black box"). Though this approach helps verify procedural compliance and may expose manipulation, some scholars argue it remains insufficient or ineffective in identifying system flaws (Joshua Kroll).

Submetido em: 15/12/2024
Aprovado em: 23/12/2024
Avaliação: Double Blind Reviewe
ISSN: 2316-753X

Finally, the implementation of **algorithmic affirmative action** to combat discrimination is gaining attention. This includes the use of independent reviewers and mandating systems to be programmed in ways that prevent discrimination. The concepts of "privacy by design" and "privacy by default" are central, requiring that legal requirements be embedded into system architecture. Emerging as a complementary notion is "explanation by design" or "by default," along with the expansion of the right to review automated decisions through the notion of **"contestability by design,"** as proposed by Kars Alfrink in "Contestable AI by Design: Towards a Framework."

This model aims to ensure that various stages of an automated decision-making system's lifecycle allow for human intervention. Data subjects should be able to intervene even during system development through participatory design approaches, granting them access to challenge decisions that cause them harm.

Such a model would allow for the reconstruction of the decision-making cycle and help identify the exact point of failure. Transparency and the principle of explainability are essential here, as data subjects must be provided with comprehensible access to the arguments and factors underlying the automated decision.

5 DEMOCRATIC VALUES, LEGITIMACY AND AUTHORIZATION

Regarding the issue of legitimation, the authors argue that algorithmic systems lack the two main pillars of legitimacy central to democratic systems: authorization and accountability. They contend that "decisions made through algorithmic systems usually lack clear mechanisms of authorization and accountability."

While algorithmic systems often lack clear mechanisms for authorization and accountability, as the authors claim, this does not prevent the State from mandating certain best practices through binding legislation—such as the drafting of compliance documents like the AIIA (Algorithmic Impact and Integrity Assessment)—in order to anticipate and mitigate risks. It is unreasonable to expect that companies will voluntarily implement robust accountability frameworks, since they typically prioritize profit and economic perspectives over sustainable medium- and long-term approaches.

Submetido em: 15/12/2024
Aprovado em: 23/12/2024
Avaliação: Double Blind Reviewe
ISSN: 2316-753X

The authors propose a path toward democratic governance based on the democratization of algorithms and the principle of accountability, emphasizing the need for responsible AI. They argue that core democratic values must form part of the governance framework for democratic accountability of algorithms. In this regard, they deem it essential to establish a network of diverse institutions for democratic oversight, regulation, and control, since relying solely on corporate self-regulation is not a viable solution. Nor is it sufficient to limit governance to purely economic or bureaucratic dimensions. What is needed, they argue, is a collaborative governance model as a response to the complexities of algorithmic society and the limitations of state-centered administration, including inflexibility, centralization, lack of specialized knowledge, and overly generic regulations. “The aim is to incorporate non-governmental perspectives without reducing governance to mere corporate self-regulation.”

Moreover, democratizing algorithmic governance would require transnational participation, involving not only experts and governments but also a broader range of citizens affected by the omnipresence of high-risk algorithms. David G. Robinson, in his work Voices of the Code, argues that inputs from non-experts could contribute to the development of more ethical algorithms.

Although the authors provide robust theoretical foundations to support the notion that algorithms can be understood as institutions—based on traditional institutional characteristics such as relatively durable sets of organized rules and practices embedded in structures of meaning and resources, which are resistant to the turnover of individuals and to idiosyncratic preferences or external changes (March and Olson, 2006:3)—one must question whether this thesis remains viable, especially considering that AI is evolving based on machine learning and human feedback, resulting in certain instability and variation of outputs.

The authors assume that the rules of the social contract remain valid during this current phase of AI evolution—often referred to as a new industrial revolution—but do not engage in any deeper reflection on whether a new form of social contract might be emerging or whether the previous one is being ruptured. They merely note a general crisis of trust in legal and political institutions and the threat of social contract breakdown.

In the context of Big Tech monopolies and increasing informational asymmetry—intensified by the convergence of artificial intelligence, the Internet of Things, and big data—there is a growing capacity for data controllers to exercise

dominance, supported by an ever-expanding informational advantage. Some scholars speak of a new form of sovereignty: data sovereignty. This raises a pressing question—can traditional democratic values alone sufficiently mitigate the risks posed by algorithmic systems? These values have already been critiqued as inadequate or fragile within democracies even before the advent of AI, which has only magnified such issues.

In general terms, this is an original and multidisciplinary proposal, building on previous theories such as Philip Napoli's perception of algorithms as institutions—arguing for the usefulness of institutionalism to analyze changes in media production and consumption in the age of social media—as well as Lawrence Lessig's assertion that "code is law." The book seeks to explore the extent to which institutional theory can help understand algorithms as new institutions in society, assembling various contemporary debates and conceptions into a comprehensive framework.

Nonetheless, several important limitations deserve attention:

> - **Lack of a practical perspective** on how democratic values could be monitored, enforced, or reaffirmed;
> - **Omission of core democratic values** such as social justice, human dignity, and fundamental rights—values essential to a democratic state governed by the rule of law. Democratization and social justice must also be integral components of algorithmic governance (Paavola, 2007);
> - **Overreliance on the notion that conceptualizing AI as institutions is sufficient** to trigger critical reflection and democratization, a claim that oversimplifies historical institutional transformations.

In this regard, the authors assert that "the historical lens of Algorithmic Institutionalism allows us not only to understand algorithms in their current configuration but also to uncover how and why certain biases or limitations may have been embedded in their structures from the outset and over time." However, many studies and reports have identified these same issues—biases arising from the use of already-skewed data, the extreme homogeneity of technical teams (often Western, heterosexual, and male), and contextual shifts—without relying on the institutionalist lens.

Is it truly sufficient to bet on "democratization" in a classical sense, especially when the authors themselves acknowledge the broader democratic crisis even before AI emerged as a factor?

They also briefly allude to a possible rupture of the social contract but do not elaborate. Their proposal includes governance beyond companies or states, involving

Submetido em: 15/12/2024
Aprovado em: 23/12/2024
Avaliação: Double Blind Reviewe
ISSN: 2316-753X

global, multi-stakeholder participation—including non-experts and citizens affected by high-risk algorithms.

The book focuses primarily on severe risks, although it does not offer a systematic classification of what constitutes "high risk." It also somewhat simplistically equates platforms like China's Social Credit System with algorithmic platforms used for financial services or hiring decisions—despite substantial contextual and normative differences.

They refer to "responsible AI," yet omit analysis of key international frameworks that define its principles. They assume democratization alone would mitigate risks. While they rightly note the lack of two core pillars of democratic legitimacy—authorization and accountability—their claims are broad and imprecise, e.g., "algorithmic decisions usually lack clear mechanisms of authorization."

However, both the LGPD and the GDPR—governing the personal data that feeds AI systems—contain solid legal bases for authorization, despite criticism of certain bases such as consent and legitimate interest. Numerous legal scholars have developed theories to address these shortcomings.

What type of "lack of authorization" are the authors referring to? The complete absence of legal basis? Inadequacy of existing bases? Or the fictional or hypertrophied nature of consent, a frequent subject of academic critique?

Many widespread activities—like social media's prioritization of virality or the abusive data mining by data brokers—are inherently harmful and disconnected from the data subject's intent or control. These concerns have reached judicial recognition. In Brazil, for instance, the Superior Court of Justice (STJ) has ruled on the economic oppression caused by lack of control over personal data (REsp 22.337-8/RS). Similarly, in the UK, the *Lloyd v. Google* case centers on "loss of control" as a legitimate harm.

Therefore, what exactly is meant by the alleged absence of authorization? Is it a lack of legal basis, improper application of consent, or failure to meet legal requirements due to algorithmic opacity?

6 THE FICTION OF CONSENT AND THEORETICAL DEVELOPMENTS

Submetido em: 15/12/2024
Aprovado em: 23/12/2024
Avaliação: Double Blind Reviewe
ISSN: 2316-753X

Given the criticisms of consent as a legal basis, many data brokers rely on another legal basis for data processing, namely legitimate interest (Art. 7, XI, of the LGPD), which requires a balancing test between the legitimate interest of the data processor and the legitimate expectations of data subjects. It is necessary to consider recent theoretical developments in this domain, which point to the need for a right to object also in relation to this legal basis, just as the LGPD provides in the case of consent. Such a distinction lacks justification, as the legitimate interest basis also presents several vulnerabilities, as discussed below. In the case of consent, the right to object is unconditional and may be exercised at any time through a free and simple procedure (Art. 8, §5 of the LGPD). The same should apply to the legitimate interest basis, through a systematic and functional interpretation of Articles 10, II and §2, and 18, §2 of the law, avoiding normative asymmetry between legal bases and safeguarding the legitimate expectations of data subjects.

An important theoretical contribution in this regard is the right to due informational process, an extension of the legal due process concept. This shifts from a purely instrumental procedural guarantee to one that ensures symmetry and proportionality, guaranteeing a form of adversarial process and broad defense. Accordingly, data processing would only be legitimate under three conditions: transparency (information), legitimate purpose, and proportionality. Based on the principle of proportionality, the processing of personal data must meet the criteria of adequacy and relevance in relation to both prior and posterior purposes.

Another essential contribution is the theory of contextualized consent developed by Solon Barocas and Helen Nissenbaum, which imposes material limits to revitalize consent as a legitimate legal tool. It seeks to shift consent from a merely formal mechanism to one embedded in real-world contexts (S. Barocas and H. Nissenbaum, “Big Data’s End Run Around Anonymity and Consent,” in Privacy, Big Data and the Public Good, Cambridge University Press, 2014, pp. 44–76):"It is time to contextualize consent by giving more attention to the background in which it operates. It is time to explore and enrich the backdrop of rights, obligations, and legitimate expectations so that consent can fulfill its appropriate role."

For Brent Mittelstadt and Luciano Floridi, privacy—conceived as the right to control data about oneself—is "an unrealistic ideal under current conditions" (predictive capabilities, mass profiling, group classification). In this reconceptualization, privacy becomes the “right to immunity” from unknown, unwanted, or unintended alterations to

Submetido em: 15/12/2024
Aprovado em: 23/12/2024
Avaliação: Double Blind Reviewe
ISSN: 2316-753X

one's identity as an informational entity, both actively and passively. Group privacy may then be legally defined as the "right to inviolable personality," which aims to protect the integrity of a group identity that is not reducible to a mere sum of its members' individual identities.

The theory of "contextual integrity" frames the right to privacy not as a right to secrecy or control, but as a right to an appropriate flow of personal information. It stems from the inadequacy of consent to support control or informational self-determination in profiling contexts, pointing to the need to assess the content of "legitimate expectations" within the specific context of data processing. Helen Nissenbaum (2011) emphasizes that profiling involves behavioral analysis of large data volumes and decision-making based on such profiles. It raises concerns about behavioral privacy economics, data confidentiality, and manipulation.

Current privacy laws are ill-suited to address privacy concerns emerging from big data analytics, especially as harms are often social and diffuse across large populations. Nissenbaum proposes a substantive model for notice and consent, emphasizing dynamic consent aligned with contextual integrity, and institutional mechanisms for normative case-by-case evaluation. The contextual integrity model asserts that all information exchanges are governed by contextual norms that define with whom and for what purposes information flows are appropriate. When contexts shift—such as in the actors or purposes of data use—privacy violations can occur unless renegotiated consent is obtained.

Dynamic consent thus requires notifying and updating data subjects about how their data are used, enabling them to provide granular, purpose-specific consent. According to the proposed framework, two critical factors must be considered:

> - **Legibility**: Data subjects are often unaware that data mining is taking place, do not know what is being collected or why, or what insights are derived. The opaque nature of data mining denies agency to data subjects, making meaningful consent impossible.
> - **Negotiability**: The context of data processing often shifts—due to changing legislation or unilateral alterations in corporate privacy policies. Data subjects must be empowered to review how their data are used as circumstances change to preserve contextual integrity. Most users lack the ability to negotiate data use terms, especially in adhesion contracts, where one either accepts all terms or foregoes service access.

For Nissenbaum, privacy protection must go beyond individual consent to include respect for the contextual integrity of data flows. A paradigmatic example is the

Submetido em: 15/12/2024
Aprovado em: 23/12/2024
Avaliação: Double Blind Reviewe
ISSN: 2316-753X

Facebook–Cambridge Analytica scandal, where the app "This Is Your Digital Life" collected data on millions, including user interests, locations, friends, and likes. Facebook users legitimately expected that such data would only be used to improve applications and products—not for microtargeting and electoral influence. This represented a clear violation of contextual integrity, requiring case-by-case evaluation.

Another key theoretical development is Sandra Wachter and Brent Mittelstadt's articulation of a new data protection right: the "right to reasonable inferences"—a "right over how one is seen." Inferences must be treated as personal data, challenging the Article 29 Working Party's position on their legal status. AI systems generate non-intuitive, unverifiable inferences about behavior, preferences, and private life, beyond the current scope of data protection laws (LGPD and GDPR), which offer individuals little control over how data are interpreted.

The "right to reasonable inferences" would require data controllers to provide an ex-ante justification establishing whether a given inference is reasonable. This would include:

> 1. Why certain data provide a normatively acceptable basis for making inferences;
> 2. Why such inferences are relevant and normatively acceptable for the processing purpose or type of automated decision; and
> 3. Whether the data and methods used are accurate and statistically reliable.

This ex-ante justification would be complemented by an ex-post mechanism allowing data subjects to contest unreasonable inferences.

7 FINAL CONSIDERATIONS

The book examines several key case studies, including algorithmic systems used in security (Chapter 3 – surveillance and risk assessment), in policy-making (Chapter 4 – resource allocation and the shaping of public policy decisions), and in recommendation systems (Chapter 5 – influencing user choices and preferences).

In summary, the authors argue that global digital platforms pose a threat to democratic systems due to the disconnection between democratic processes and technical control within a public sphere increasingly governed by code. However, as

with other complex institutions throughout history, algorithmic systems can be democratized to mitigate the risks that some of them pose to modern societies.

By regarding algorithms not merely as lines of code embedded in systems, but as inherently political institutions—political artifacts or objects in themselves—capable of shaping and establishing social norms and behaviors, and of reinforcing or undermining democratic processes and values, the theory of algorithmic institutionalism offers a new framework. This framework enables us to move beyond the debate surrounding the agency of AI—that is, its capacity for autonomous decision-making, often seen as a prerequisite for attributing moral status to artificial intelligence.

Algorithms would thus be considered institutions akin to social institutions such as marriage, the Church, or even law, norms, or discourse. The central idea here is that such institutions influence individual behavior because individuals have learned and internalized their values.

However, if we reflect on the opacity, the lack of transparency of algorithms, and the widespread absence of awareness that one is being subjected to an algorithmic system, alongside the lack of control and accountability mechanisms, can we still speak of individuals learning and internalizing values that are unilaterally imposed?

Perhaps it would be more accurate to speak of a process of institutionalization of algorithms, rather than claiming that they are already fully established institutions—this, of course, depends on the concept of institution employed. If institutions are understood as those structures that guarantee social coexistence and are therefore worth defending, how can we extend this status to systems that still pose threats to fundamental rights, given the various documented harms arising from AI applications?

REFERENCES

ALMEIDA, Virgílio, MENDONÇA, Ricardo Fabrino, FILGUEIRAS, Fernando, “Algorithmic Institutionalism: The Changing Rules of Social and Political Life, Oxford University Press, 2014.

BURRELL, Jenna, “How the machine ‘thinks’: Understanding opacity in machine learning algorithms”, https://journals.sagepub.com/doi/full/10.1177/2053951715622512#:~:text=In%20this%20article%2C%20I%20draw,required%20to%20apply%20them%20usefully

CELAN. Paul. The Meridian: Final Version-Drafts-Materials, ed. Bernhard Böschenstein and Heine Schmull, transl. Pierre Joris, Stanford University Press, 2011.
ČERKA, Paulius; GRIGIENĖ, Jurgita; SIRBIKYTĖ, Gintarė. "Liability for damages caused by Artificial Intelligence", Computer Law & Security Review, Elsevier, v. 31, n. 3, p. 376-389, 2015.

CHEHOUDI, Rafaa. "Artificial intelligence and democracy: pathway to progress or decline", JOURNAL OF INFORMATION TECHNOLOGY & POLITICS, 2025.

CHUL HAN, Byung, "A sociedade da transparência", Petrópolis: Vozes, 2017.

COECKELBERGH, Mark. "Why AI Undermines Democracy and What to Do About It", Polity Press , 2024.

DELEUZE, G. "Conversações", tradução de Peter Pál Pelbart, São Paulo: Editora 34; 3a edição, 2013.

____. "Post-Scriptum sobre as Sociedades de Controle", Conversações: 1972-1990. Tradução de Peter Pál Pelbart, Rio de Janeiro: Ed. 34, 1992.

DWIVEDI, Yogesh, et all. "Artificial intelligence (AI):Multidisciplinary perspectives on emerging challenges,opportunities, and agenda for research, practice and policy", International Journal of Information Management,57, 1–47. https://doi.org/10.1016/j.ijinfomgt.2019.08.002.

EUBANKS, Virginia. "Automating inequality: How high-tech tools profile, police, and punish the poor", St. Martin's Press, 2018.

KROLL, Joshua, "Accountable Algorithms", 2015, https://www.jkroll.com/papers/dissertation.pdf.

FLORIDI, L. "Open Data, Data Protection, and Group Privacy", Philos. Technol. 27, 1–3 , 2014.

LOTRINGER, Sylvère, e VIRILIO, Paul. The Accident of Art, New York, Semiotext(es), 2005.

MITTELSTADT, Brent; WACHTER,Sandra. "A right to reasonable inferences: re-thinking data protection law in the age of big data and AI, Columbia Business Law Review, v. 2019.

MULHOLLAND, Caitlin (coords.). "Inteligência Artificial e Direito: ética, regulação e responsabilidade". São Paulo: Thomson Reuters Brasil, 2019.

NISSEMBAUM, Helen. "Privacy, Big Data, and the Public Good";. A Contextual Approach to Privacy Online. In Daedalus, v. 14, n. 4, 2011, <https://www.amacad.org/publications/daedalus/11_fall_nissenbaum.pdf>.

ROCHE, C., LEWIS, D., & WALL, P. J. "Artificial intelligenceethics: An inclusive global discourse?", Cornell University Library, 2021.

Submetido em: 15/12/2024
Aprovado em: 23/12/2024
Avaliação: Double Blind Reviewe
ISSN: 2316-753X

_____Roche, C., Wall, P. J., & Lewis, D. “Ethics and diversityin artifcial intelligence policies, strategies and initiatives”, AIand Ethics, 3, 1095–1115, 2023.

RODRIGUES, R. “Legal and human rights issues of AI: Gaps, challenges and vulnerabilities”, Journal of ResponsibleTechnology, 4, 4, 2020.

ROSS, M. L. “The political economy of the resourcecurse”. World Politics, 51(2), 297–322, 1999.

ROUVROY, Antoinette e BERNS, Thomas, “Governamentalidade algorítmica e perspectivas de emancipação: o díspar como condição de individuação pela relação?”, Revista ecopos, 18, v. 2, 2015.

SOUZA, Eduardo Nunes de.“Dilemas atuais do conceito jurídico de personalidade: uma crítica às propostas de subjetivação de animais e de mecanismos de inteligência artificial”, Revista civilistica, 9. n. 2, 2020, https://civilistica.emnuvens.com.br/redc/article/view/562/417.

ZUBOFF, Shoshana. “The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power”, PublicAffairs; 1st edition, 2019.