

TOWARDS AN INSTAGRAMMICAL ARCHAEOGENEALOGY: SHADOWBANNING AS A TECHNOLOGY OF ALGORITHMIC GOVERNMENTALITY

POR UMA ARQUEOGENEALOGIA INSTAGRÂMICA: O *SHADOWBANNING* COMO
TECNOLOGIA DE GOVERNAMENTALIDADE ALGORÍTMICA

POR UNA ARQUEOGENEALOGÍA INSTAGRÁMICA: EL *SHADOWBANNING* COMO
TECNOLOGÍA DE GUBERNAMENTALIDAD ALGORÍTMICA

Laís Di Lauro^{*}

Regina Baracuh^{**}

Universidade Federal da Paraíba

RESUMO: No ano em que se comemora o centenário de Michel Foucault, este artigo mostra a produtividade da teoria para discutir a sociedade algorítmica do século XXI. Objetiva-se cartografar o fenômeno do *shadowbanning* no Instagram, evidenciando seus efeitos sobre a circulação de discursos acerca de corpos e vozes historicamente marginalizadas. A partir de uma abordagem foucaultiana, este estudo compreende o *shadowbanning* como uma forma de governar o sensível ou, nos termos de Foucault (2014), como uma microfísica da visibilidade: um poder que opera não pela interdição direta, mas pela modulação da atenção, pela rarefação da circulação e pela administração diferencial do campo do visível e do dizível. Ancorada em uma perspectiva arqueogenológica, a pesquisa mobiliza conceitos como governamentalidade, dispositivo e regimes de verdade para analisar como sistemas algorítmicos

^{*} PhD in Linguistics from the Federal University of Paraíba (PROLING/UFPB), specializing in Discourse Studies, with a doctoral research period at NOVA University Lisbon (UNL/PT). She holds a Master's degree in Media Studies from the Federal University of Rio Grande do Norte (PPGEM/UFRN), specializing in Media Studies and Meaning-Making. She is a member of the CORPOLÍTICA (UFRN) and CIDADI (UFPB) research groups. E-mail: laisdilauro@gmail.com.

^{**} Professor at the Federal University of Paraíba (UFPB), where she supervises research at the undergraduate research, Master's, and doctoral levels. She is affiliated with the Department of Portuguese Language and Linguistics (DLPL) and is a permanent faculty member of the Graduate Program in Linguistics (PROLING/UFPB). She completed a postdoctoral fellowship in Linguistics at the Federal University of Uberlândia (UFU) in 2023. She is a member of the Foucaultian Discourse Studies Working Group (GT) of the National Association of Graduate Studies and Research in Letters and Linguistics (ANPOLL) and coordinates the Interinstitutional Research Group CIDADI – Discourse Analysis Discussion Circle (UFPB/CNPq), which brings together faculty researchers from several Brazilian universities across different regions of the country. E-mail: mrbl@academico.ufpb.br.

participam da produção contemporânea da invisibilidade instagrâmica. O percurso metodológico consiste em fazer uma cartografia discursiva de um *corpus* composto por publicações no perfil do Instagram da criadora de conteúdo Tricia Wise (@safeslut), cuja experiência recorrente de redução de alcance permite examinar o *shadowbanning* não apenas como um mecanismo técnico de exclusão, mas como um acontecimento discursivo no qual se condensam tensões entre poder, regimes de verdade e práticas de resistência na sociedade algorítmica. Os resultados indicam que a invisibilização de perfis do Instagram é uma tecnologia de governo, articulada à modulação algorítmica da circulação, por meio da qual se regulam a atenção e as condições de possibilidade da enunciação na plataforma. Conclui-se que a governamentalidade algorítmica distribui diferencialmente atenção, legitimidade e possibilidades de enunciação, (re)inscrevendo desigualdades históricas sob a aparência de neutralidade técnica.

PALAVRAS-CHAVE: Estudos Discursivos Foucaultianos. Algoritmos. Shadowbanning. Instagram. Corpos Marginalizados.

RESUMEN: En el año en que se conmemora el centenario de Michel Foucault, este artículo muestra la productividad de su teoría para discutir la sociedad algorítmica del siglo XXI. El objetivo es cartografiar el fenómeno del *shadowbanning* en Instagram, poniendo de relieve sus efectos sobre la circulación de discursos acerca de cuerpos y voces históricamente marginados. Desde un enfoque foucaultiano, este estudio comprende el *shadowbanning* como una forma de gobernar lo sensible o, en términos de Foucault (2014), como una microfísica de la visibilidad: un poder que opera no mediante la prohibición directa, sino a través de la modulación de la atención, la rarefacción de la circulación y la administración diferencial del campo de lo visible y lo decible. Anclada en una perspectiva arqueogenealógica, la investigación moviliza conceptos como gubernamentalidad, dispositivo y regímenes de verdad para analizar cómo los sistemas algorítmicos participan en la producción contemporánea de la invisibilidad instagrâmica. El recorrido metodológico consiste en realizar una cartografía discursiva de un *corpus* compuesto por publicaciones del perfil de Instagram de la creadora de contenido Tricia Wise (@safeslut), cuya experiencia recurrente de reducción del alcance permite examinar el *shadowbanning* no solo como un mecanismo técnico de exclusión, sino como un acontecimiento discursivo en el que se condensan tensiones entre poder, regímenes de verdad y prácticas de resistencia en la sociedad algorítmica. Los resultados indican que la invisibilización de perfiles de Instagram constituye una tecnología de gobierno, articulada a la modulación algorítmica de la circulación, mediante la cual se regulan la atención y las condiciones de posibilidad de la enunciación en la plataforma. Se concluye que la gubernamentalidad algorítmica distribuye diferencialmente la atención, la legitimidad y las posibilidades de enunciación, (re)inscribiendo desigualdades históricas bajo la apariencia de neutralidad técnica.

PALABRAS CLAVE: Estudios Discursivos Foucaultianos. Algoritmos. Shadowbanning. Instagram. Cuerpos Marginalizados.

ABSTRACT: In the year that marks the centenary of Michel Foucault, this article demonstrates the productivity of his theoretical framework for examining the algorithmic society of the twenty-first century. We aim to cartography the phenomenon of shadowbanning on Instagram, highlighting its effects on the circulation of discourses concerning historically marginalized bodies and voices. Drawing on a Foucauldian approach, this study conceptualizes shadowbanning as a way of governing the sensible or, in Foucault's terms (2014), as a microphysics of visibility: a form of power that operates not through direct prohibition, but through the modulation of attention, the rarefaction of circulation, and the differential administration of the field of the visible and the sayable. Anchored in an archaeogenealogical perspective, the research mobilizes concepts such as governmentality, *dispositif* (apparatus), and regimes of truth to analyze how algorithmic systems participate in the contemporary production of Instagram invisibility. The methodological approach consisted of a discursive mapping of a corpus composed of publications on the Instagram profile of content creator Tricia Wise (@safeslut), whose recurring experience of reduced reach allows us to examine shadowbanning not only as a technical mechanism of exclusion, but as a discursive event in which tensions between power, regimes of truth, and practices of resistance in algorithmic society are condensed. The findings indicate that the invisibilization of Instagram profiles constitutes a technology of government articulated with the algorithmic modulation of circulation, through which attention and the conditions of possibility of enunciation on the platform are regulated. It is concluded that algorithmic governmentality differentially distributes attention, legitimacy, and possibilities of enunciation, thereby (re)inscribing historical inequalities under the guise of technical neutrality.

KEYWORDS: Foucauldian discourse studies. Algorithms. Shadowbanning. Instagram. Marginalized bodies.

1 CARTOGRAPHING THE PRESENT: FOUCAULT AND THE ALGORITHMIC SOCIETY

*I am a pyrotechnician. I manufacture something that
ultimately serves for a siege, a war, a destruction.
I am not in favor of destruction, but I am in favor of
being able to pass, of being able to advance,
of being able to make walls fall.
– Michel Foucault*

In the year celebrating the centenary of Michel Foucault, his thought remains indispensable for understanding contemporary modes of truth production and the ways in which discourses are intertwined with power relations, processes of subject constitution, and practices of care. The “toolbox” left by the author offers a particularly fertile horizon for interrogating the various discursive regimes and power relations that span the contemporary world.

To think about digital media today, it is necessary to draw on tools capable of apprehending forms of power that are exercised not only through explicit prohibition, but through subtle modulations, probabilistic calculations, and invisible architectures that organize the field of possibility. In this sense, Foucault’s work presents itself as a fruitful field for understanding various current situations.

It is within this framework that this article is situated. By mobilizing Foucauldian analytics to examine contemporary practices of algorithmic modulation, we have selected the phenomenon of shadowbanning¹, understood as a discursive practice of governmentality, in which the exclusion of content does not operate through direct interdiction, but through rarefaction, opacity, and the reconfiguration of the circulation and reach of Instagram posts understood as discursive events.

Our interest here is to understand the way in which governmentality acts in the constitution of an algorithmic economy within the digital regime of visibility: a system that not only organizes what is seen and what is hidden, but which modulates forms of enunciation and subjectivation, in accordance with the imperatives of predictability, stability, and normativity required by digital platforms.

Analyzing the effects of this algorithmic rationality on the production of discourses and subjectivities requires the activation of a theoretical framework capable of tensioning the links between knowledge, power, language, media, and technology. Accordingly, this research is anchored in the thought of Michel Foucault, mobilizing especially the concepts of discourse, governmentality, *dispositif* (apparatus), and regime of truth, which are analytic operators that allow for the questioning of the ways in which subjects are constituted and knowledge is regulated in technology-mediated contexts.

The transformations marking the beginning of the twenty-first century reveal the consolidation of a new power-knowledge regime, sustained by the integration of digital technologies, data production, and algorithms. The innovative character of this research lies precisely in displacing Foucauldian Discursive Studies toward the field of digital technologies, proposing a reading of algorithmic governmentality as a new mode of exercising power. By articulating the concepts of apparatus, governmentality, and visibility (Foucault, 2008; 2014) with contemporary debates on the digital (Rouvroy and Berns, 2013; Bruno, 2021; Noble, 2018; Benjamin, 2019), the epistemological scope of discourse analysis is expanded, highlighting its potential to interrogate the mechanisms of control and subjectivation inherent to the hyper-technological society.

This constitutes a theoretical and methodological contribution that enables a dialogue between Foucauldian Discursive Studies and Critical Algorithm Studies, by proposing that algorithms constitute a new layer of discursive materiality. This study advances by

¹ In this article, the terms shadowban and shadowbanning are used interchangeably to designate the (assumed or effective) practice of silently reducing the visibility of content or profiles by digital platforms, without explicit notification to the affected individuals, varying only according to whether they are used as nouns or to emphasize the procedural nature of the algorithmic modulation of circulation.

demonstrating that governmentality, in contemporary times, is exercised not only over populations, but over data, profiles, and digital performances, revealing a biopolitical dimension reconfigured by informational logic.

From an ethical and social standpoint, this research responds to an urgency that transcends academic boundaries: the impact of artificial intelligence on the ways of life, expression, and recognition of the contemporary subject. By investigating how algorithms reproduce and intensify inequalities of gender, race, class, and sexuality, we denounce the selective and exclusionary nature of the structures operating on digital platforms, while simultaneously proposing a reflection on the responsibility of technologies of power in controlling life.

The relevance of this article unfolds in three complementary dimensions: theoretical, by updating Foucauldian thought for the context of the algorithmic society; social, by exposing the mechanisms of invisibilization and exclusion inscribed in the dynamics of digital visibility; and ethical, by interrogating the effects of artificial intelligence on life, bodies, and forms of existence. To think about algorithmic governmentality is, therefore, to think about new forms of government of the sensible; and it is in this gesture of problematizing the present that this article finds its strength, its pertinence, and its unique contribution.

Our objective is to cartograph the phenomenon of shadowbanning on Instagram, highlighting its effects on the circulation of discourses from marginalized bodies and voices. To this end, we understand cartography, as outlined by Baracuh (2020), as a methodological strategy consistent with the Foucauldian perspective, since it is oriented by movements, flows, and emergences, rather than by fixed models of representation. In conducting a cartographic analysis, we envision a social cartography, which, unlike traditional cartography aimed at tracing maps of territories and population distribution, concerns itself with maps of force relations. These involve the agency of knowledge and power with their tactics and strategies, materialized in statements, in games of objectification and subjectivation, in practices of resistance and freedom, establishing multiple and differentiated relations where power operates a gridding of the entire social field. Thus, cartography is understood here as an analytical gesture inscribed within the very functioning of the practices it analyzes, following their displacements, their opacities, and their silences.

The analyses derived from the selected corpus focus both on the metrics, before and after the shadowban, and on the effects that this mechanism produces on regimes of visibility. It is important to highlight that the shadowban constitutes an architectural decision: a calculation that, without explicit human signature, establishes which content will achieve engagement and which will be rendered invisible. In this process, it becomes fundamental to question: which subjects and bodies are more easily led to erasure? What rationalities of power legitimize this operation? And what possibilities for resistance and strategies of counter-conduct may emerge in the face of this form of silent exclusion?

2 TOOLS FOR THINKING ABOUT TODAY: FOUCAULDIAN CONCEPTS

Governing, from the Foucauldian perspective, is not restricted to imposing laws or exercising sovereignty; it is, rather, about guiding conduct. Foucault (2008, p. 244) defines governing as “a form of action that is exercised over possible actions”, that is, over what others may do, feel, think, or desire. It would then be an exercise of power that operates upon freedom, which does not eliminate the subject but shapes their gestures, their margins, their limits of action. A power that acts not through negation, but through positivity: it incites, induces, diverts, facilitates or hinders, and makes more or less probable. Ultimately, it may compel or prevent, but it always traverses the field of possibility. As Foucault himself summarizes, it is “a conduct whose object is conduct; a power that acts upon subjects who also act” (Foucault, 2008, p. 244).

Governmentality, in Foucault (2008), appears as a political rationality, expressing a rationalization of relations of power that acts not upon isolated subjects, but upon populations, managing their flows, risks, behaviors, and regularities. This government operates through statistical, biomedical, legal, and economic knowledge, which makes the administration of life legitimate, acceptable, and even desirable.

This rationality begins to take shape at the end of the sixteenth century and the beginning of the seventeenth, articulated with the formation of the modern State and the emergence of new techniques for the administration of collective life. The focus shifted from governing only territories to governing bodies, habits, flows, and contingencies. This represents the transition from a regime of sovereignty, exercised through the right of life and death, toward more subtle and disseminated forms of power, anchored in the regulation of life, in the monitoring of population dynamics, and in the production of knowledge that allows for the administration of probabilities, behaviors, and events that affect collective existence.

The conception of governmentality is profoundly linked to the notion of biopolitics, also developed by the author, which concerns the governance of life in all its dimensions. That is, it is an exercise of power that is not limited to normalizing or prohibiting, but one that administers, calculates, and predicts, infiltrating everyday practices and processes of subjectivation; a power whose primary function is to manage the life of populations, no longer acting solely upon the individual and their singularities.

Unlike the overt surveillance typical of sovereign or authoritarian regimes, control here is exercised in a diffuse and continuous manner, becoming naturalized in the social fabric and presenting itself as a legitimate and necessary part of the order of things. Biopolitics, in this sense, does not appear in isolation, but as one of the most central expressions of modern governmentality.

Moving these Foucauldian conceptions to contemporary terrains, in order to understand the microphysics of new apparatuses of subjectivation, we follow the seminal reading of Rouvroy and Berns (2013), who insert algorithmic apparatuses within the analytics of relations of power proposed by Foucault. What emerges, according to the Belgian authors, is not just a new technology, but a new regime of government: algorithmic governmentality.

Unlike classical forms of power, algorithmic governmentality requires neither conscious subjects, nor deliberation, nor will; it operates upon data. It does not act upon conducts already performed, but upon the field of possibility: it modulates behaviors before they happen, through the constant updating of patterns, correlations, and probabilities. It is a predictive, discrete form of power that acts without needing to intervene directly on three domains: the first is data surveillance (dataveillance), in which the continuous and automated collection of data generated by individuals' digital activities occurs; the second is data mining, the stage where collected data is processed through advanced statistical analysis techniques that identify patterns and correlations without the need for prior hypotheses; and the third refers to action upon behaviors (profiling), that is, the application of predictive models to influence and direct the actions of individuals.

The fact that governments and corporations seek to predict individual behavior with precision is not, in itself, a new phenomenon; surveillance practices and strategies have always existed. As Foucault (2008) describes, the exercise of power has mobilized, across different periods, disciplinary apparatuses of biopolitical control: from overt and punitive surveillance of individual bodies to the meticulous control of behavior in institutions such as schools and factories, to the management of populations through statistical and medical knowledges capable of calculating risks and anticipating threats. However, the core of Rouvroy and Berns' (2013) critique concerns the current paradox: we are no longer trusting rational and political human decisions, but machines that do not interpret the world, possess no moral judgment, and do not attribute meaning to data.

The complex mathematical and learning structure behind the algorithmic functioning of social networks enables them to determine what will or will not have visibility. Although these digital spaces are sometimes viewed as democratic places of expression, the (hidden and silent) workings of the algorithms that offer us a kind of analogous "mediated visibility" (Thompson, 2005) cannot be ignored. Furthermore, the great unknown behind this functioning is that companies do not reveal their *modus operandi*: algorithmic procedures are largely unavailable to us and are inevitably selective, emphasizing some information while discarding others (Gillespie, 2014).

On Instagram, this manifests itself through the distribution of posts which, depending on a series of criteria that are not entirely transparent to the public, either create engagement with or invisibilize certain content. An example of algorithmic content regulation on Instagram is the practice of shadowbanning, understood by Middlebrook (2020) as an action that involves the partial censorship of accounts or posts without the user's knowledge or consent.

Historically, the term shadowban emerged in developer forums in the 2000s and became popular in the social media ecosystem starting in 2016, when creators began reporting an abrupt drop in engagement on certain posts without any official explanation. The expression refers to a moderation practice on digital platforms in which a user's content is partially or totally hidden from the public without any formal notification or transparency regarding the reason.

As Gillespie (2018) points out, this form of algorithmic silencing is one of the opaque strategies used by technology companies to maintain "discursive order" on social networks without necessarily exposing themselves to the political burden of explicit censorship. Unlike direct suspension or removal, this practice prevents content from being distributed in the feed or through search and information grouping mechanisms on the platform such as hashtags, the Explore tab, or search results, while the user's account appears, ostensibly, to be functioning normally.

The role of algorithms in curating content circulating on social media not only organizes information flows but also shapes ways of seeing, knowing, and speaking. From Noble's (2018) perspective, shadowbanning is one of the manifestations of the "algorithmic governance" exercised by platforms, which often penalizes discourses of resistance, marginalized bodies, or politically sensitive topics in accordance with market logics, normativity, and surveillance. Regarding shadowbanning, Instagram states that:

Some posts on Instagram may be inappropriate for our global community, even if they do not violate our Community Guidelines. Therefore, the recommendation of these types of posts in places like search results and the Explore section will be limited (Instagram, 2024).

The social network's stance on shadowbanning is criticized by Middlebrook (2020), who questions who is being referred to by the supposed "global community" that Instagram claims to be committed to protecting. The author questions which groups are being privileged by this strategy and which are being systematically erased in pursuit of this goal. This reflection reveals the workings of the algorithmic apparatus and its exclusionary practices, which often operate selectively and invisibly.

The *Algorithmic Bias Report: September 2021*, published by Salty World (2021), revealed that Instagram profiles of plus-size people, queer individuals, and Black women experience shadowbanning at a disproportionate rate compared to those from mainstream communities.

Another article, published in *The Guardian*, titled "Instagram's murky 'shadow bans' just serve to censor marginalized communities" (Joseph, 2019), also pointed to cases of members of marginalized communities who suffered shadowbanning without evident reason. An excerpt from the text states that "non-sexual images of queer people, Black women, and plus-size women have been subjected to shadowbanning practices, suggesting that these bodies, whether dressed or not, are inherently sexual or something that should be hidden" (Joseph, 2019, n.p.).

Many of these penalties are the result of biases in AI systems, which automatically associate certain words or images with inappropriate content without considering the context. An example of this action was the case of the cyclist who was a victim of sexual harassment (see image 1), whose video was flagged by Instagram as "sensitive content", exemplifying the *modus operandi* of algorithmic silencing mechanisms.

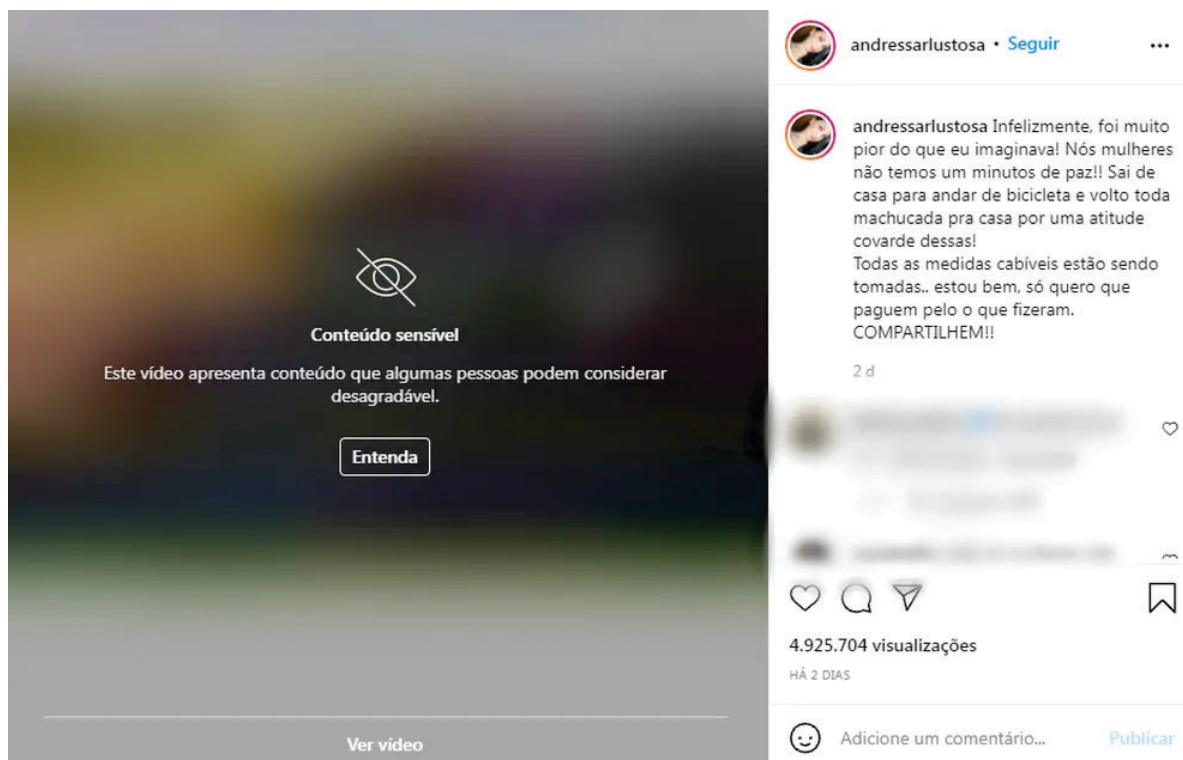


Image 1: Instagram post by Andressa Lustosa²

Source: Lustosa (2021)³

By activating the warning filter that restricts the viewing of content under the argument that it “may be disturbing”, Instagram performs what, from a Foucauldian perspective, can be understood as a discursive requalification: the statement “denunciation of violence” is displaced from its political field and reinscribed in another regime of truth, that of technical moderation. This constitutes a change in the conditions of enunciation, insofar as the event is recomposed under new rules of visibility. What was previously a testimony about the vulnerable body becomes a fact to be softened, so that the user’s gaze on the platform is preserved, while the victim’s body is silenced. The signaling, apparently neutral and protective, operates, however, as yet another of the numerous layers of the apparatus of power that defines what can be seen and under what conditions. This requalification of content highlights how algorithmic devices not only censor, but also produce new forms of enunciation, (re)inscribing discourses in regimes of truth that neutralize their political potential under the pretext of sensitive protection.

Although they can be studied as abstract computational tools, it is essential to recognize algorithms as architectures that are incorporated into the practices of the world, influencing, and being influenced by, the information they process (Couldry, 2012). In this context, their role in social media assumes a crucial part in shaping the contemporary regimes of visibility, since digital platforms are part of the daily lives of billions of people around the world. Considering that algorithms are present in various spheres of everyday life, we draw attention to the structural opacity and lack of transparency in the processes involving them, especially regarding social media.

² Legend: “Unfortunately, it was much worse than I imagined! We women don’t have a moment of peace!! I left home to ride my bike and came back all bruised because of such a cowardly act! All appropriate measures are being taken... I’m okay, I just want them to pay for what they did. Share this!!”

³ The publication was available on the @andressalustosa Instagram profile at the time the corpus was collected in 2021 but was subsequently removed by the author. The image reproduced in this article corresponds to the record made during the data collection period.

3 THE POLITICS OF INVISIBILITY: WHO IS SILENCED?

*silence cried out in revolt
memory echoed the present
history does not allow for chance
– Luíza Romão*

Visibility, in the context of digital platforms, is neither a neutral given nor a spontaneous consequence of the subjects' participation. It is, as Bruno (2021) observes, a technology of power: a set of procedures and rationalities that defines what can appear, in what forms, and under what conditions. While it produces visibility, it simultaneously institutes invisibilities. In this economy of the gaze, seeing and being seen become regulated operations, modulated and calculated by technical apparatuses that manage appearances, performances, and affects. Inscribed in this logic, this discursive tactic cannot and should not be understood as a system failure, but as a mechanism for the political regulation of visibility.

As Foucault (2008) proposes, modern power acts less through repression and more through the meticulous administration of life. In the digital environment, this rationality takes the form of an algorithmic government of attention flows, which intensely circulates certain images and discourses while pushing others into zones of silence. Power does not prohibit, but it ensures that the discourse of some does not reverberate. To be silenced, in this regime, does not mean being prevented from speaking, but being deprived of the possibility of making oneself heard. Exclusion occurs at the level of visibility: what is said loses reach and statistical existence in a universe managed by algorithmic architectures. As Bruno (2013) discusses, algorithms participate in the selection and hierarchization of content that achieves greater or lesser visibility, affecting the conditions under which discourses circulate.

This refers to a way of governing the sensible, or, in Foucault's terms (2014), a microphysics of visibility: a power that acts not through direct violence, but through the modulation of attention. Invisibility, therefore, is characterized by the way in which power is actualized: to silence is to manage the field of listening; to invisibilize is to manage the field of the visible in order to undermine potential resistance. It is the production of subjectivities aligned with the demands of calculation. If one wishes to be seen, if one wants the statement to circulate, one must dance to the order of discourse or, in this case, the algorithmic order.

Algorithmic power does not act only upon data, but upon statements, bodies, voices, and affects, distributing visibility and shadow according to the logics of engagement and profitability. Understanding its architecture therefore implies interrogating the regimes of truth that sustain this economy of the visible, for, as Foucault (1996, p. 12) states:

Each society has its regime of truth, its "general politics" of truth: that is, the types of discourse which it welcomes and makes function as true; the mechanisms and instances which enable distinguishing true statements from false ones, the way in which one and the another are sanctioned; the techniques and procedures that are valued for obtaining truth; the status of those who are charged with saying what counts as true.

On Instagram, the right to speak and to appear is conditioned by this truth of the time, which we can understand as the "algorithmic truth", which privileges what generates interaction, what harmonizes with the network's aesthetics, and what does not "disturb" the fluidity of media consumption. The Shadowban acts as a filtering apparatus for difference through a technique of visibility management, which ensures that the space of appearance remains homogeneous.

The politics of visibility simultaneously implies a making-invisible. While it enables or prevents the appearance of discourses, voices, and bodies, it also redefines the modes of appearance within the discursive field, delimiting forms of existence and participation. The government of visibility becomes, thus, a form of the government of life, sustained by a biopolitics of the gaze that distributes bodies and speech according to criteria of adequacy, utility, and docility. In this intertwining of aesthetics and norm, the shadowban reveals the production of silence as a technology of government.

4 TO SILENCE IS TO GOVERN: SHADOWBANNING AS AN APPARATUS OF POWER

Understanding shadowbanning as an apparatus of power implies shifting the gaze from technique to politics. It is not merely about reducing the reach of accounts or publications on digital platforms, but about inscribing an operation within the contemporary rationality of government, in which algorithms manage conduct, administer visibility, and reorganize the possibilities of enunciation. If the apparatus, as outlined by Michel Foucault, articulates practices, norms, knowledge, and strategies oriented by a govern function, shadowbanning emerges as one of its updates, adjusted to the conditions of possibility that mark the present.

This operation reveals continuity with the Foucauldian model of power, which is defined not by prohibition, but by the capacity to produce realities, create fields of possibility, and induce ways of being and appearing.

What makes power hold good, what makes it accepted, is simply the fact that it doesn't only weigh on us as a force that says no, but that it traverses and produces things, it induces pleasure, forms knowledge, produces discourse. It needs to be considered as a productive network which runs through the whole social body, much more than as a negative instance whose function is repression (Foucault, 1980, p. 119).

Power produces knowledge, and knowledge produces power, in an endless cycle. It is in this circuit that shadowbanning is inscribed as a contemporary practice of regulating visibility: beyond restricting access, it manufactures effects of truth regarding what can circulate, who is considered relevant, and which statements can emerge in the digital public space.

The architecture in which these artificial intelligence technologies are developed reflects the social, political, and epistemic conditions of the societies that create them. Biases of the social world, such as racism, sexism, and classism, are not eliminated in technical conversion; on the contrary, they are incorporated, amplified, and naturalized in the form of statistical models, relevance metrics, and automated classifications. Even when produced without discriminatory intent, such systems carry with them the structural marks of the relations of power that precede them. Thus, algorithms do not merely classify; they normalize and transform historical inequalities into effects of visibility.

As Foucault (2008) suggests, power operates in a subtle and pervasive manner, establishing the modes of enunciation that define who has the right or authority to speak on a given subject and from which institutional position. The shadowban likewise operates within this logic, implicitly managing the limits of what can be enunciated.

Although Instagram systematically denies the existence of “shadowban” as an official category, Meta’s institutional documents reveal that the platform operates formal mechanisms for limiting visibility. In the report *Lowering the Distribution of Problematic Content*, the company explicitly states that it reduces the reach of content considered “problematic” or that “violates Community Guidelines in full or in part”, adopting measures that prevent such posts from being widely recommended or appearing in the discovery tab (Meta, 2023).

Likewise, in the Creator FAQ – a section dedicated to answering questions from content creators – Meta acknowledges even more directly that “certain actions may limit the distribution” of posts or accounts, even when there is no explicit violation or notification to the user. The word “shadowban” is deliberately avoided, but the practice of restricting reach is institutionally admitted through the vocabulary of “distribution limiting” in a discursive strategy that shifts the phenomenon from the sphere of censorship to the administrative sphere of “safety”, “integrity”, and “quality” of content, erasing the political character of this management.

The platform’s policy acts in such a way as to induce the production of subjectivities adjusted to the demands of engagement and the logic of visibility. More than limiting the dissemination of content, the algorithm mobilizes conduct management techniques that operate on the anticipation of judgment and the ever-present possibility of vanishing from the feed. This continuous risk installs a regime of self-monitoring and self-censorship, in which subjects begin to govern themselves according to the expectations inscribed in algorithmic rationality.

Foucault (2014, p. 143) identifies that the success of disciplinary power is due to three techniques: “hierarchical observation, normalizing judgment, and their combination in a procedure that is specific to it, the examination”. From a Foucauldian perspective, the primary function of this type of power would be “to train” individuals. Similarly, we can consider that social media algorithms exercise a new technique of disciplinary power, given that “discipline ‘makes’ individuals” (Foucault, 2019, p. 167). It is, therefore, a performative operation of algorithmic and artificial-intelligence categorization systems that act in the modulation and orientation of conduct and decisions in the incitement of habits, desires, and interests, already extensively documented (Bruno, 2013; Gillespie, 2014). By transforming erasures into calculations and metrics, the shadowban reveals the contemporary face of disciplinary power.

5 CARTOGRAPHY OF ERASURES: THE CASE OF TRICIA WISE (@SAFE.SLUT) AND THE TRACES OF INVISIBILITY

[...] Today, instead of being burned at the stake, we are erased from memory and deprived of our rights. But this is nothing new; we have been silenced for centuries.
– Tricia Wise

From a Foucauldian perspective, cartographing is a political gesture. It is not a matter of spatially representing the world, but of following the movements of force that traverse it. Cartography, as Baracuhy (2020, p. 131) reminds us, “aims not only to discuss relations of power through an inventory of discursive techniques and tactics [...], but, above all, to diagnose the present time”.

Cartography moves away from any descriptive or classificatory pretension: it does not seek to fix an object, but to follow flows and variations, to describe the modulations of power that traverse practices, bodies, and discourses. It belongs to a method, derived from archaeogenealogy, guided by a sensitivity toward discontinuities, displacements, and effects. Cartographing cases of shadowbanning implies observing how power permeates algorithmic operations, producing simultaneous displacements in the conditions of enunciation and in the visibility practices that order the emergence of discourses and the appearance of marginalized bodies.

When observing the case of American content creator Tricia Wise, we find an example of this process. The profile @safe.slut was created in 2019 with the goal of educating, raising awareness, and empowering people regarding herpes diagnoses, sexual education, and the critique of gender norms. In 2020, a year in which the COVID-19 pandemic and social isolation measures led to a massive intensification in the use of social networks, Tricia’s profile received a series of invisible restrictions. Various posts shared by the user stopped being displayed on the platform, account engagement decreased drastically, and the profile ceased to appear in search engines, becoming “invisible” to other users.

The analysis of this case allows us to observe not only the numerical effects of the shadowban – the drop in engagement metrics and reach – but, above all, the discursive effects of the practice. By comparing the performance of posts before and after the restriction, a significant reduction in reach is observed in posts related to female pleasure, artistic nudity, and sexual self-care; themes that, although legitimized by public health policies, are frequently framed by algorithms as “sensitive content”. This automatic categorization exposes a regime of truth that associates female sexuality with obscenity, the female body with censorship, and pleasure with that which must remain hidden. In an account later published by Salty, Wise describes an experience of content moderation on Instagram:

I made a post about how people can access abortion pills, and it was removed because they claimed I was “selling illegal products”. I changed the caption to censor the word “abortion” and omitted the information about the abortion pill, and the new post was accepted (Wise, 2022).

The statement above highlights a point of tension where the algorithmic apparatus operates as a mechanism of discursive control. The first layer of this event demonstrates how the platform activates a regime of truth by automatically interpreting the statement as the “sale of illegal products”. In this procedure, the algorithm produces an operational and automated truth about that discourse,

regardless of its intentionality, since, as we have extensively discussed so far, the algorithm, although intelligible, analyzes data in a decontextualized manner, without logical interpretation. The technical gesture that classifies and filters content as inappropriate is evident, even when this content was not, in fact, violating the platform's rules and guidelines.

In the second instance, we can observe that Tricia, by suppressing the word “abortion” and removing information from the publication, conforms to the order of discourse imposed by the platform, shaping her statement to the conditions of acceptability inscribed in the apparatus. What we find here goes beyond the prohibition of a specific term, confirming the existence of a discursive economy in which certain words trigger automated protocols of risk, invisibility, and penalization. The cartography thus shows how users adapt to the system's opaque rules, adjusting their statements to operate according to the norms of the very apparatus that limits them.

Furthermore, the fact that the second post was accepted after the omission of the term “abortion” confirms the functioning of an algorithmic governmentality that (re)organizes the circulation of statements, while simultaneously (re)producing erasures of statements, of voices, and of bodies. The platform does not only exclude discourses; it also creates sanitized versions of what can be said, conditioning visibility and public debate on certain topics.

At this point, it is worth noting that to make the statement compatible with the logic of the algorithmic apparatus, Tricia resorted to a form of discursive evasion known as “algospeak”, “a discursive phenomenon consisting of the encoding of words with the intent of bypassing algorithmic monitoring mechanisms” (Barreto et al., 2025, p. 198). This resource, as Steen, Yurechko, and Klug (2023) point out, is a form of coded speech that emerges precisely when users identify that the system regulates not only what can be said, but how it can be said.



Image 2: Post by Tricia Wise using algospeak

Source: Wise, 2023

The post shown above from the profile @safe.slut demonstrates that statements do not emerge freely, but rather, they are conditioned by the rules of formation that delimit what can be said, in what forms, and with what effects. Firstly, the explicit use of algospeak in the artwork (substituting “slut” with “sl♥ut”) makes visible the strategy social media users employ to escape algorithmic punishment mechanisms. Subsequently, it also reveals that discourse on female sexuality, especially when articulated by women who place themselves at the margins of moral normativity, historically occupies a place of risk and discursive vulnerability.

Algospeak operates as a tactic of resistance, in the Foucauldian sense that where there is power, there is always the possibility of escape, deviation, or counter-conduct. The need to suppress or encode words, signs, and spellings to avoid algorithmic detection

highlights the existence of a contemporary procedure of rarefaction: certain discourses are filtered, restricted, or negated even before they become fully enunciable, especially if they thematize autonomous female sexuality.

It is worth noting that the very aesthetics of the post – visually clean, pink, and soft – constitute a strategy of normalized appearance, another resource employed to manage one's own visibility and neutralize potential punishments. Thus, the body and subjectivity presented are modeled to become “acceptable” within an algorithmic regime that explicitly punishes deviations. The post, in this case, is an enunciative gesture that reorganizes itself in order to continue existing.

Finally, this specific content is inscribed in a regime of subjectivation in which women need to justify, reorganize, or soften their public appearance to gain visibility. Algospeak is an inventive discursive practice that creates gaps in algorithmic control by allowing the circulation of the interdict through coded means. The post, therefore, brings to the surface the tension between the production of subjects and the imposition of norms, revealing how platforms reconfigure ways of speaking and appearing through algorithmic techniques of surveillance and filtering.

The cartography of Tricia Wise's case also shows how erasure is accompanied by visibility management mechanisms that fuel the disciplining of Instagrammers through messages sent by the platform to the user. Generic notifications, alerts about “guideline violations”, or recommendations on how to “improve engagement”, which constitute statements that teach the subject what is acceptable and what must be avoided, are frequently seen. It is a disciplinary process, in the light of Foucauldian thought, in which power is exercised not only through prohibition but through the induction of self-surveillance at the expense of reach.



Image 3: Message from the platform posted by Tricia Wise
Source: Wise, 2021

The analysis of the statements produced by Tricia within the platform, based on the selected corpus, allows us to observe the emergence of a counter-discourse in which silence is named, erasure is made explicit, and invisibility becomes an object of enunciation. Far from operating as a totalizing or absolute instance, power, in Foucauldian analytics, is constituted as a mobile relation, traversed by frictions, displacements, and practices of counter-conduct. Tricia's case confirms this immanence: resistance is not situated outside the algorithmic regime of visibility, but is inscribed within it, mobilizing the very mechanisms of circulation and exposure that produce the silencing.

The publication of screenshots of the platform's actions, the use of denunciation hashtags such as #shadowbanned, and the collective engagement surrounding the erasure transform the algorithmic apparatus into a field of symbolic dispute, in which statements operate as vectors of destabilization for the norms that regulate appearance.

6 FINAL REMARKS ON ALGORITHMIC GOVERNMENTALITY

Returning to Michel Foucault at the end of this journey is a way to reaffirm the potency of his work for a critical understanding of the present and to ratify what Albuquerque Júnior (2008, p. 9) describes as a “cartographer of our time and our world”⁴. By offering conceptual tools for the historical analysis of the government of conduct, regimes of truth, and technologies of power, his “toolbox” provides the instruments that allow us to interrogate various contemporary practices, such as shadowbanning. It was through this theoretical apparatus that this article sought to understand this contemporary practice on Instagram as an operation inscribed within the broader rationalities of algorithmic government.

The tension between silencing and resistance allows us to understand, drawing on Benjamin (2019), how platforms reproduce racial and gender inequalities under the appearance of computational impartiality. When marginalized subjects must make their own invisibility public in order to exist discursively, algorithmic power reaches its most paradoxical point, silencing and, at the same time, compelling the subject to speak, to prove that they have been silenced, and to attempt to reclaim the visibility that has been obscured from them.

From the perspective of statements, shadowbanning produces an unequal redistribution of the conditions of enunciability. Statements that conform to dominant affective and aesthetic regimes are amplified, while those that tension norms of gender, body, and sexuality are progressively displaced to zones of low visibility. However, the analysis demonstrates that erasure is not absolute. Even under algorithmic reach restrictions, traces remain – residual likes, comments, mentions, and screenshots – which compose a dispersed archive, traversed by fissures and resistance.

At this point, cartography reveals its analytical power by following these vestiges, not with the goal of restoring lost visibility, but of understanding how power operates in the intervals between what is enunciated and what is silenced. Shadowbanning is thus inscribed as an event within this archive, silently rearranging the conditions under which certain statements can emerge, circulate, or be progressively obscured.

By articulating these elements, it becomes possible to understand shadowbanning as a contemporary reconfiguration of the technologies of power analyzed by Foucault, as well as one of the central phenomena through which algorithmic governmentality is enacted in the digital environment. Just as disciplinary panopticism operated through continuous surveillance and the internalization of the norm, algorithmic power acts through the management of visibility and the modulation of conduct. The fundamental difference lies in the fact that, in the digital environment, the norm is not imposed from the outside, but is produced from within the system itself, emerging from the interactions, data, and statistical patterns that feed back into it. It is a form of distributed and immanent surveillance.

Shadowban, therefore, is inscribed in this regime as a predictive composition technique that, by responding to algorithmic surveillance, reorganizes the conditions under which certain subjects and statements can appear or be progressively displaced into zones of invisibility.

Finally, it must be highlighted that although shadowbanning produces exclusion, it can also engender specific forms of political visibility. When Tricia Wise makes her silencing public and when the term shadowban begins to circulate as a collective statement, an epistemological displacement occurs: algorithmic power ceases to act only as an invisible force and begins to be named, problematized, and converted into an object of analysis. This visibility can be understood as a form of counter-conduct, as a practice inscribed within it, that tensions its rationalities. The denunciation of erasure, the formation of support networks, and the sharing of metrics and evidence thus configure discursive practices that reorder subject positions and produce modes of digital solidarity in the face of algorithmic governmentality.

⁴ Translation from the Portuguese original by the authors.

Thus, the case of Tricia Wise allows us to understand shadowbanning not merely as a technical mechanism of exclusion, but, in the light of Michel Foucault, as an integral part of the apparatuses of power in digital media, constituted by practices of algorithmic governmentality and counter-conduct. These practices are articulated with regimes of truth that differentially regulate visibility and discursive circulation on digital platforms. It is concluded that Foucauldian analytics remains central to understanding the present by revealing how subtle forms of power are exercised through algorithmic infrastructures and produce contemporary regimes of invisibility.

REFERENCES

- ALBUQUERQUE JÚNIOR, D. M.; VEIGA-NETO, A.; SOUZA FILHO, A. (org.). *Cartografias de Foucault*. Belo Horizonte: Autêntica, 2008.
- BARACUHY, R. Cartografias da resistência: As vozes das manifestações de rua. *MOARA – Revista Eletrônica do Programa de Pós-Graduação em Letras*, v. 1, n. 57, p. 126-142, 2020. Disponível em: <https://periodicos.ufpa.br/index.php/moara/article/view/9688>. Acesso em: 15 ago. 2026.
- BARRETO, M. S.; BARACUHY, R.; SANTOS, R. C.; SOUSA, L. V.; ALVES, L. L. Algospeak na ordem discursiva do Instagram: A subversão do algoritmo. *Boletim de Conjuntura (BOCA)*, Boa Vita, v. 21, n. 62, p. 187-204, 2025. Disponível em: https://revista.ioles.com.br/boca/index.php/revista/article/view/6640?utm_source=chatgpt.com. Acesso em: 16 ago. 2026.
- BENJAMIN, R. *Race after technology: Abolitionist tools for the New Jim Code*. Medford, MA: Polity Press, 2019.
- BRUNO, F. *Máquinas de ver, modos de ser: vigilância, tecnologia e subjetividade*. Porto Alegre: Sulina, 2013.
- BRUNO, F. Racionalidade algorítmica e laboratório de plataforma. In: GROHMANN, R. (org.). *Os laboratórios do trabalho digital: entrevistas*. São Paulo: Boitempo, 2021. p. 155-165.
- COULDRY, N. *Media, society, world: Social theory and digital media practice*. Cambridge: Polity Press, 2012.
- FOUCAULT, M. *A arqueologia do saber*. Tradução de Luiz Felipe Baeta Neves. 8. ed. Rio de Janeiro: Forense Universitária, 2008.
- FOUCAULT, M. *A ordem do discurso*: aula inaugural no Collège de France, pronunciada em 2 de dezembro de 1970. Tradução de Laura Fraga de Almeida Sampaio. São Paulo: Loyola, 1996.
- FOUCAULT, M. *Microfísica do poder*. Tradução de Roberto Machado. 10. ed. Rio de Janeiro/São Paulo: Paz e Terra, 2019.
- FOUCAULT, M. *Vigiar e punir: nascimento da prisão*. Tradução de Raquel Ramalhete. 42. ed. Petrópolis/RJ: Vozes, 2014.
- GILLESPIE, T. The relevance of algorithms. In: GILLESPIE, T.; BOCZKOWSKI, P. J. FOOT, K. A. (ed.). *Media technologies: Essays on communication, materiality, and society*. Cambridge/MA: The MIT Press, 2014. p. 167-193.
- GILLESPIE, T. Custodians of the internet: platforms, content moderation, and the hidden decisions that shape social media. New Haven: Yale University, 2018.
- INSTAGRAM. Por que algumas publicações no Instagram não estão aparecendo nos resultados da pesquisa e no Explorar? Central de Ajuda, *Instagram*, 2024. Disponível em: <https://help.instagram.com/613868662393739>

JOSEPH, C. Instagram's murky 'shadow bans' just serve to censor marginalised communities. *The Guardian*, 8 nov. 2019. Disponível em: <https://www.theguardian.com/commentisfree/2019/nov/08/instagram-shadow-bans-marginalised-communities-queer-plus-sized-bodies-sexually-suggestive>. Acesso em: 15 ago. 2026.

LUSTOSA, A. Publicação sobre a restrição de conteúdo no Instagram. Instagram, 2021. Publicação excluída.

META. Lowering distribution of problematic content. *Meta Transparency Center*. 2023. Disponível em: <https://transparency.meta.com/enforcement/taking-action/lowering-distribution-of-problematic-content/>. Acesso em: 15 ago. 2026.

MIDDLEBROOK, C. *The grey area*: Instagram, shadowbanning, and the erasure of marginalized communities. SSRN, 2020.

NOBLE, S. U. *Algorithms of oppression*: How search engines reinforce racism. New York: NYU Press, 2018.

ROMÃO, L. S. Sangria. 1. ed. São Paulo: Selo do Burro, 2017.

ROUVROY, A.; BERNIS, T. Governance de l'algorithmique et de la donnée: Vers une juridicisation des technologies de l'informatique décisionnelle, 2013. <https://shs.cairn.info/revue-reseaux-2013-1-page-163?lang=fr>

SALTY. *Algorithmic bias report*: September 2021. 2021. Disponível em: <https://saltyworld.net/algorithmicbiasreport-2/>. Acesso em: 15 ago. 2026.

STEEN, E.; YURECHKO, K.; KLUG, D. You can (not) say what you want: Using algospeak to contest and evade algorithmic content moderation on TikTok. *Social Media + Society*, v. 9, n. 3, p. 1-17, 2023. Disponível em: <https://doi.org/10.1177/20563051231194586>. Acesso em: 16 ago. 2026.

THOMPSON, J. B. The new visibility. *Theory, Culture & Society*, v. 22, n. 6, p. 31-51, 2005.

WISE, Tricia. I shared my sexual assault story and Instagram deleted my account. *Salty*, 2022. Disponível em: <https://saltyworld.net/i-shared-my-sexual-assault-story-and-instagram-deleted-my-account/>. Acesso em: 29 set. 2026.

WISE, T. *thank u @wearemon for writing this rly nice lil recap about my chat w @iamcatherinedrysdale last month [...]*. Instagram: @safe.slv, 16 dez. 2023. Disponível em: <https://www.instagram.com/p/C07KJNlIuaBC/>. Acesso em: 16 ago. 2026.

WISE, T. *y'all I just woke up to this* [...]. Instagram: @safe.slvt, 6 dez. 2021. Disponível em: https://www.instagram.com/p/CXILclNulZw/?img_index=1. Acesso em: 16 dez, 2025.



Received on April 28, 2026. Accepted on September 29, 2026.
Published on October 15, 2026.