

The online state of nature: Kantian perspectives on freedom of expression, platform power and information disorder

O estado de natureza online: perspectivas kantianas sobre liberdade de expressão, poder das plataformas e desordem informacional

Tailine Hijaz¹

Universidade Federal do Paraná (UFPR)
tailinehijaz@hotmail.com

Joel T. Klein²

Universidade Federal do Paraná (UFPR)
jthklein@yahoo.com.br

Abstract: This article argues that the current system of digital platform governance amounts to an online state of nature, a condition characterized by provisional rights, the absence of public guarantees, and unilateral control over speech. Drawing on Kant's political philosophy, we examine how the concentration of power within digital platforms, exercised without subjection to public law, transforms the conditions under which civil freedom and meaningful public discourse can flourish. We relate this structure to the dynamics of information disorder, and argue that freedom of expression in the digital age must be rethought in light of public norms grounded in principles shared by all.

Keywords: state of nature; Kant; freedom of expression; digital platforms.

Resumo: O artigo argumenta que a configuração atual da governança das plataformas digitais constitui um estado de natureza online, marcado por direitos provisórios, ausência de garantias públicas e controle unilateral sobre o discurso. Com base na filosofia política de Kant, examinamos como o poder concentrado nas plataformas, exercido sem subordinação ao direito público, transforma as condições nas quais a liberdade civil e o debate público significativo podem ocorrer. Relacionamos essa estrutura às dinâmicas da desordem informacional e sustentamos que a liberdade de expressão na era digital precisa ser repensada com base em normas públicas fundadas em princípios compartilháveis por todos.

Palavras-chave: estado de natureza; Kant; liberdade de expressão; plataformas digitais.

¹ PhD Student at Federal University of Paraná (UFPR/Brazil) and Universität Vechta (Germany). This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001. Project CAPES/DAAD/PROBRAL.

² Professor of UFPR/CNPq and Visiting Research Professor at Goethe University Frankfurt. This work is supported by ERC grant (LSR, Grant agreement number: 101170288).

Recebido em 30 de julho de 2025. Aceito em 11 de dezembro de 2025.

Introduction

In recent decades, digital transformation has profoundly altered the conditions under which information is produced, exchanged, and contested. As digital platforms have become central infrastructures for communication and interaction, they have accumulated significant political and economic power, often exercised without transparency or public accountability. These developments have also generated structural tensions between the principle of freedom of expression and the realities of online discourse, in which decisions about visibility, access, and communication are shaped by private actors operating outside public procedures.

In this article, we adopt a normative philosophical approach to this configuration. Drawing on Kant's concept of the state of nature, understood as a critical device rather than a historical hypothesis, we examine contexts in which there is no public law and the exercise of freedom depends on unilateral will and contingent force. We argue that Kant's distinction between provisional private and public rights provides a framework for evaluating the legitimacy of such arrangements. We further argue that major digital platforms within constitutional societies operate as pockets of a state of nature, functioning as normative enclaves where users, content, and algorithms interact under private rules and commercial incentives rather than publicly justified norms. In such spaces, the structural conditions for civil freedom and meaningful public discourse are not secured by law, but are instead left to the discretionary governance of private entities. From a Kantian perspective, this configuration cannot be regarded as morally legitimate.

The article is divided into three parts. First, we reconstruct Kant's account of the state of nature and the moral duty to establish public law, identifying the criteria that distinguish legitimate from illegitimate authority. Second, we develop a diagnostic analysis of platform governance, emphasizing how the regulation of speech is shaped by opaque, profit-oriented infrastructures. Third, we examine the normative implications of this configuration, and argue that freedom of expression in the digital age must be rethought in light of public norms grounded in principles shared by all.

1. The immorality of the state of nature: A Kantian framework

According to contractualist theories, civil society may have arisen naturally in historical terms, but it cannot be legitimized in a normative sense. In contrast to accounts that regard it as naturally determined, such as those of Aristotle, Hume, and Hegel, although from different perspectives, contractualist theories assume that the establishment of the state and the law requires specific justification. This entails providing criteria for distinguishing legitimate institutions from illegitimate ones. It is within this context that the concept of the state of nature emerges (KERSTING, 1994).

The concept of the "state of nature" offers a counterpoint to how human life would be organized if there were no state to create and guarantee law. Since the focus is on the state's legitimacy rather than its creation, the state of nature does not refer to a primitive state prior to the emergence of the state. Instead, it functions as a normative and conceptual construct for evaluating political institutions and the normative issues that arise in their absence. There are several concepts of the state of nature, and contractualist philosophers justify their understanding of law and justice based on how this concept is defined.

Within the framework of contractualist theories, Kant's model of the state of nature is particularly valuable for addressing the concept of a just state and for evaluating political and legal institutions. Kant seeks to construct a model that preserves certain elements found in the theories of Hobbes, Locke, and Rousseau while avoiding their respective shortcomings. From Hobbes, Kant retains the idea that, in the state of nature, individuals would live in constant physical and legal insecurity, since they could rely only on their own strength to enforce their claims. In such a condition, violence prevails, and individuals remain in perpetual tension and potential conflict with one another — perhaps not in an actual state of war, as Hobbes maintains, but rather in a condition of ever-present risk of conflict. Conversely, Kant rejects Hobbes's view that the state of nature entails a complete absence of sociability. From Locke, he adopts the view that rights can indeed be acquired in the state of nature, but he denies that such rights are absolute or that the civil state arises merely to preserve them. From Rousseau, he incorporates the idea that the social contract must reflect the will of everyone involved, understood as the general will in contrast to the will of all, while rejecting the notion that the content of the contract should be culturally or socially grounded in any particular conception of common happiness.

Some of the issues that Kant identifies in the models of the state of nature proposed by Hobbes, Locke, and Rousseau can be summarized as follows. Hobbes depicts the state of nature as purely individualistic. In this condition, even the strongest individual would remain vulnerable to violent death, for instance while sleeping. Yet a more plausible version of the state of nature would be one in which there exists a limited form of sociability. In such a case, individuals would gather in groups bound by loyalty and exercise their arbitrary will over others. Although still a condition of potential conflict, the state of nature is more accurately described as a struggle among more or less organized groups than as a struggle among isolated individuals. The model of Locke recognizes the social dimension of the state of nature but portrays it as excessively harmonious. He maintains that individuals generally acquire rights through peaceful means, by mixing their labor with unpossessed things. However, the idea of labor as the normative foundation of property is highly problematic, not only historically and culturally, but also philosophically, since it presupposes a direct and dogmatic relation between the person and the thing. Moreover, the philosophy of Locke offers no political or juridical remedy when this provision is not respected, giving his theory an intrinsic tendency to preserve the status quo. The model of Rousseau, in turn, presents two versions of the state of nature, both of which are equally implausible. The original state of nature is characterized by largely independent individuals who, unlike in the model of Hobbes, seek to remain isolated and generally avoid contact. By contrast, the degenerate state of nature described in the *Second Discourse* is defined solely by domination and deception, leaving no room for rightful claims.³

In light of these models, it is worth considering the merits of the Kantian model of the social contract. Unlike Hobbes's model, the state does not possess the legitimacy to eliminate rights that could have been acquired in the state of nature, since it does not create rights *ex nihilo*. In other words, the state is inherently limited and therefore precluded from becoming authoritarian. Unlike Locke's model, the state does not merely serve as an instrument of property owners for maintaining a given status quo. Unlike Rousseau's model, the legitimacy of the state does not rest on the customs of a specific community or on any particular conception of happiness.

³ On the role of the concepts of the state of nature and the social contract in Rousseau's work see Consani and Klein (2022).

Grounded in the idea of the *omnilateral will*, the Kantian state can regulate and qualify rights, organizing them systematically in accordance with the principle of equal freedom rather than according to any conception of happiness. In other words, the state has no intrinsic tendency toward authoritarianism (as in Hobbes), injustice (as in Locke), or paternalism (as in Rousseau). Rather, it is a state with a liberal legal structure and a republican ethos that enables everyone to exercise their rights as they see fit, provided that they do not infringe upon the rights of others.⁴

Furthermore, unlike earlier theories, Kant does not justify leaving the state of nature through conditional reasoning grounded in self-interest. For Hobbes, if one wishes to avoid violent death, one must also wish to leave the state of nature. For Locke, if one intends to preserve the rights acquired in the state of nature, one must enter the civil state to ensure their protection. For Rousseau, if one seeks to recover the freedom and happiness lost in the “true state of nature,” one must leave the corrupt state of nature and take part in the social contract. For Kant, however, the transition is not based on conditional reasoning (“If I want x, then I must do y”) but on an unconditional duty of practical reason: “I ought to enter the civil state”. The state of nature should be abandoned not because it is undesirable for some other reason, however compelling, but because it is intrinsically immoral and unjust. In Kant’s terminology, the obligation to leave the state of nature constitutes a *categorical imperative* rather than a hypothetical one.

Following this broader conceptual delimitation within contractualist theories, our attention now turns to a more detailed examination of Kant’s concept of the state of nature. While not exhaustive, the following discussion outlines its main features.

a) *On the morality of the civil state and the immorality of the state of nature.* According to Kant,

A condition that is not rightful, that is, a condition in which there is no distributive justice, is called a state of nature (*status naturalis*). What is opposed to a state of nature is not (as Achenwall thinks) a condition that is social and that could be called an artificial condition (*status artificialis*), but rather the civil condition (*status civilis*), that of a society subject to distributive justice. For in the state of nature, too, there can be societies compatible with rights (e.g., conjugal, paternal, domestic societies in general, as well as many others); but no law, ‘You ought to enter this condition,’ holds a priori for these societies, whereas it can be said of a rightful condition that all human beings who could (even involuntarily) come into relations of rights with one another ought to enter this condition (MM 06:306; see also 06:242).⁵

This passage indicates that a social condition can exist within the state of nature, governed by what Kant calls private law. The postulate of public law follows from the very legitimacy of private law:

[...] when you cannot avoid living side by side with all others, you ought to leave the state of nature and proceed with them into a rightful condition, that is, a condition of distributive justice. - The ground of this postulate can be explicated analytically from the concept of right in external relations, in contrast with violence (*violentia*)” (MM 06:307)⁶.

⁴We are aware that our claims contrasting Kant’s contractualism with that of Hobbes, Locke, and Rousseau are controversial. However, since these claims do not play a fundamental role in the argument of this paper, and because a proper justification of them would require an independent paper, we will leave this issue here as a contextual introduction and address it more carefully on another occasion.

⁵All translations follow *The Cambridge Edition of the Works of Immanuel Kant* (Kant 1992ff), with pagination according to the Akademie Ausgabe (Kant (1900-)). The works quoted in this paper are: MM – Metaphysics of Morals; TP – On the common saying: That may be correct in theory, but it is of no use in practice; WO – What does it mean to orient oneself in thinking?; WA – An answer to the question: What is enlightenment?; CJ – Critique of the power of judgment.

⁶It should be noted that Kant’s legal philosophy is concerned with the regulation of external freedom, that is, with the external relations among agents. The aim, therefore, is not to evaluate or determine the agents’ intentions at the moment of action. What matters primarily is the action itself, as it can be perceived by a third-party observer. Intentions become relevant only in the

Recognizing private law within the context of the state of nature is important for at least two reasons. First, denying its validity would imply that individuals who wish to enter into relations of rights and duties with one another would lack the legitimacy to do so. This would amount to stripping them of their status as persons and reducing them to the level of things. Second, acquired rights in the state of nature cannot be regarded as either entirely legitimate or entirely illegitimate. If they were completely legitimate, the state would have no moral function but only an instrumental one, as in Locke's theory. If they were completely illegitimate, the state would become the sole source of all law and morality, as in Hobbes's model. Kant positions his view precisely within this middle ground:

The first and second of these conditions can be called the condition of private right, whereas the third and last can be called the condition of public right. The latter contains no further or other duties of human beings among themselves than can be conceived in the former state; the matter of private right is the same in both. The laws of the condition of public right, accordingly, have to do only with the rightful form of their association (constitution), in view of which these laws must necessarily be conceived as public (MM 06:306).

Claiming that the matter of private law remains the same within public law means that if someone can acquire something as their own within the framework of private law then that right cannot be eradicated within the civil condition. However, this right is not absolute, as in Locke's theory, since the state has a moral obligation to guarantee "the rightful form of their association (constitution), in view of which these laws must necessarily be conceived as public". Using marriage as an example, this means that the state may legitimately establish criteria for entering into marriage, such as being of legal age, having mental capacity, and not being already married. Ultimately, the state regulates private rights according to the idea of the omnilateral will. In this sense, the state does not determine the content of the law, which remains the same as in private law; rather, it regulates its form. Thus, the state cannot dictate whom a person should marry or whether they should marry at all, but it can establish rules that those who wish to marry must follow. This ensures that marriage functions as a legitimate legal institution compatible with other state institutions according to the idea of a system of equal freedom. Civil regulation is necessary because private rights are essentially indeterminate in their limits and can easily come into conflict with one another, leading individuals to protect their claims through violence or unilateral coercion. Within the state of nature, only violence or sheer force can enforce ambiguous and indeterminate rights. That is why

[g]iven the intention to be and to remain in this state of externally lawless freedom, men do one another no wrong at all when they feud among themselves; for what holds for one holds also in turn for the other, as if by mutual consent (*uti partes de iure suo disponunt, ita ius est.*) But in general they do wrong in the highest degree by willing to be and to remain in a condition that is not rightful, that is, in which no one is assured of what is his against violence (MM 06:307f.).

It is precisely because of the uncertainty surrounding the limits of each person's rights and the means by which they can be enforced that Kant holds that private law in the state of nature is merely provisional. "So only in a civil condition can something external be mine or yours" (MM 06:256). This means that

[...] a unilateral will cannot serve as a coercive law for everyone with regard to possession that is external and therefore contingent, since that would infringe upon freedom in accordance with universal laws. So it is only a will putting everyone under obligation, hence only a collective general (common) and powerful will, that can provide everyone this assurance. - But the condition of being

evaluation of illegitimate actions, where they may serve as aggravating or mitigating factors. In cases where the action is lawful, however, the evaluation of intent plays no role.

under a general external (i.e., public) lawgiving accompanied with power is the civil condition (MM 06:256).

b) *Different spheres of the state of nature.* The concept of the state of nature allows us to adopt a perspective of illegitimacy that does not depend on historical or genealogical explanation. According to Kant, we may overcome the state of nature in one sphere while remaining within it in another. Thus, a people may have overcome the state of nature and established a civil condition within a certain domain, yet still find themselves in a state of nature in relation to other peoples existing alongside them. In this sense, “a state, as a moral person, is considered as living in relation to another state in the condition of natural freedom and therefore in a condition of constant war” (MM 06:343)⁷. One could therefore argue that, from one perspective, a civil condition exists, while from another, it has not yet been achieved. With the advent of the internet and social media, one could further argue that a new sphere has emerged in which human relations are no longer organized according to the standpoint of civil freedom but rather according to that of natural, or even savage, freedom.

c) *Right as a relation among persons.* According to Kant, private law in the state of nature cannot arise from a direct relationship between persons and things, as Locke proposed. Rather, legal relations can be established only among persons, and these relations sometimes refer to things. Therefore,

it is clear that someone who was all alone on the earth could really neither have nor acquire any external thing as his own, since there is no relation whatever of obligation between him, as a person, and any other external object, as a thing. Hence, speaking strictly and literally, there is also no (direct) right to a thing. What is called a right to a thing is only that right someone has against a person who is in possession of it in common with all others (in the civil condition) (MM 06:261).

Claiming something as one's own implies a relation to another person, which, to be legitimate, must satisfy the criteria of proportionality and universal reciprocity. Even when acknowledging the potential legitimacy of rights that arise within the state of nature, such provisional rights must still conform to the categorical principle of right, which states: “Act externally in such a way that the free use of your choice can coexist with the freedom of everyone in accordance with a universal law” (MM 06:231).

d) *The shortcomings of an ethical approach.* The moral problem inherent in the state of nature cannot be resolved simply by educating individuals or cultivating their moral character, because it

is not experience from which we learn of the maxim of violence in human beings and of their malevolent tendency to attack one another before external legislation endowed with power appears, thus it is not some deed that makes coercion through public law necessary. On the contrary, however well disposed and law-abiding human beings might be, it still lies a priori in the rational idea of such a condition (one that is not rightful) that before a public lawful condition is established individual human beings, peoples and states can never be secure against violence from one another, since each has its own right to do what seems right and good to it and not to be dependent upon another's opinion about this. So, unless it wants to renounce any concepts of right, the first thing it has to resolve upon is the principle that it must leave the state of nature, in which each follows its own judgment, unite itself with all others (with which it cannot avoid interacting), subject itself to a public lawful external coercion, and so enter into a condition in which what is to be recognized as belonging to it is determined by law and is allotted to it by adequate power (not its own but an external power); that is, it ought above all else to enter a civil condition (MM 06:312).

⁷ See also: “In this problem the only difference between the state of nature of individual human beings and of families (in relation to one another) and that of nations is that in the right of nations we have to take into consideration not only the relation of one state toward another as a whole, but also the relation of individual persons of one state toward the individuals of another, as well as toward another state as a whole” (MM 06: 343f.).

In a civil condition, individuals do not need to rely solely on their own judgment about what is fair or unfair, because external coercive laws define these boundaries. Even if ethical education could protect us from all the negative consequences of the state of nature, such a condition would still be unjust, since it does not clearly specify the rights and duties of each individual from the systematic standpoint required by civil freedom. People could disagree in good faith about the definition and application of rights and duties, which means that there would be too much uncertainty to ensure the protection of anyone's rights.

However, as the modern state has evolved and political as well as legal relations have become more complex, the concept of the state of nature must be updated to address these new challenges. We propose the notion of pockets of the state of nature. By this, we mean contexts in which human relations are subject to the demands of rights and duties, since individuals can harm one another's external freedom without their actions being regulated by laws established by a competent authority. In such cases, agents may violate the freedom of others, and each must enforce their own rights. Although claims to rights and duties arise naturally from practical reason, these rights remain provisional, indeterminate, and uncertain, leaving room for unilateral interpretation by each agent.

Unlike the general concept of the state of nature, which functions as a conceptual counterpart to the civil condition, the idea of pockets of the state of nature presupposes the existence of a functioning civil state. It depends on the civil state because it requires an effective legal framework for such contexts to emerge. These pockets may arise naturally as new forms of external freedom relations develop, or they may result from deliberate actions by groups that seek to benefit from a civil state guaranteeing a legal superstructure while simultaneously withdrawing from or restricting its application in certain domains. This leaves individuals to determine their rights unilaterally within those spheres.

From a Kantian point of view, even if the state has not created laws to regulate a particular domain, all human relations capable of affecting the external freedom of others must remain open to possible regulation. For such regulation to be legitimate, however, it must proceed according to the idea of an omnilateral will rather than a unilateral or even multilateral one, since various powerful agents may otherwise unite to impose their interests on weaker parties.

Given this context, the next section will examine whether digital platforms reproduce, in new forms, a condition analogous to the state of nature. This condition is marked by the absence of public guarantees, by power asymmetries, and by the indeterminacy of rights. Building on this hypothesis, we will explore how digital platforms contribute to the current crisis of informational disorder.

2. Platforms as private governors: diagnosing the crisis of information disorder

The internet has radically transformed the conditions under which information is produced, circulated, and consumed. It has given rise to new forms of collective intelligence (LÉVY, 2015)⁸, operationalized "global networks of instrumentalities" (CASTELLS, 2018, p. 77), and inaugurated

⁸ According to Lévy, collective intelligence "is a distributed intelligence everywhere, continuously enhanced, coordinated in real time, which results in the effective mobilization of skills". He adds that "the basis and goal of collective intelligence are the mutual recognition and enrichment of people, not the cult of fetishized or hypostatized communities" (LÉVY, 2015, p. 29). Translated by us from: "é uma inteligência distribuída por toda parte, incessantemente valorizada, coordenada em tempo real, que resulta em uma mobilização efetiva das competências" and "a base e o objetivo da inteligência coletiva são o reconhecimento e o enriquecimento mútuos das pessoas, e não o culto de comunidades fetichizadas ou hipostasiadas".

a new “Galaxy” distinct from traditional mass media (CASTELLS, 2003)⁹. This shift not only expanded access to the public sphere but also reshaped the architecture of expression in notable ways. However, it also led to the concentration of power in the hands of a few platforms — including Meta (Facebook, Instagram, WhatsApp), Google (YouTube), X (formerly Twitter), and TikTok — whose roles now exceed those of neutral intermediaries and who wield significant influence over the digital public sphere¹⁰. For instance, Facebook alone reported nearly 3.43 billion daily active users worldwide as of March 2025, more than twice the population of China¹¹. Moreover, their annual revenues surpass the GDPs of numerous nations, and their infrastructures mediate everyday forms of social interaction, information access, and political deliberation on a global scale.¹²

As Klonick observes, these companies no longer operate as neutral conduits for third-party content; instead, they have become “New Governors”, private entities that self-regulate based on a mix of economic interests and perceived alignment with democratic expectations (2018, p. 1603). Rather than merely hosting content, they act as norm entrepreneurs, interacting with users, states, and other stakeholders, apart from making discretionary decisions about what can be said, shared, or silenced. Indeed, Klonick describes how they establish internal rules, maintain enforcement mechanisms, and act through centralized authority (2018, p. 1663). A well-known example is the suspension of Donald Trump’s account (@realDonaldTrump) after the January 6 Capitol attack. Although the correctness of this decision can be debated, it was implemented through private policies rather than judicial orders or constitutional adjudication¹³.

While these actors now perform roles that were once reserved for public institutions, they remain private corporations driven by user engagement and profit maximization rather than public reason or constitutional duties. Unlike democratic governments, which are, in principle, subject to transparency, justification, and equal treatment, social media platforms prioritize the maximization of user attention, a practice that is often monetized through targeted advertising

⁹ Castells presents the example of YouTube to clarify that, although it is a mass communication medium, it is distinct from traditional media. Due to the interactivity and horizontality of networks, “anyone can post a video on YouTube, with some restrictions. It is the user who selects the video they want to watch and comment on from a vast list of possibilities” (CASTELLS, 2018, p. 21). Translated by us from: “qualquer um pode postar um vídeo no Youtube, com algumas restrições. É o usuário que seleciona o vídeo que quer ver e comentar a partir de uma enorme lista de possibilidades”

¹⁰ See Persily (2022, p. 200), who points out the complex nature of these platforms, arguing they are more than passive intermediaries like “common carriers”, highlighting the role of algorithms in organizing content, and arguing that traditional legal frameworks developed for government speech suppression do not adequately apply to them: “[t]hey are not merely hosting speech, but organizing it. The most important feature of the platforms is the algorithms they employ to structure a unique ‘feed’ for every individual user”.

¹¹ See Meta Reports First Quarter 2025 Results here: https://s21.q4cdn.com/399680738/files/doc_news/Meta-Reports-First-Quarter-2025-Results-2025.pdf.

¹² While revenue and GDP measure different things, the comparison remains valid to illustrate the scale of these corporations’ economic power.

¹³ As explains Franks (2022, p. 75), “Twitter had first temporarily locked the @realDonaldTrump account on January 6 after Trump posted a video and a statement repeating false claims about the election and expressing his ‘love’ for the rioters, requiring Trump to delete the tweets before being able to post again. At the time of the lockout, the Twitter Safety team noted that if Trump violated Twitter’s policies again his account would be banned. In a blog post on January 8, the company explained that it had determined that two of the Trump tweets that followed the riots, one referencing ‘American Patriots’ and another stating that Trump would not be attending President- Elect Biden’s inauguration, were ‘likely to inspire others to replicate the violent acts that took place on January 6, 2021, and that there are multiple indicators that they are being received and understood as encouragement to do so’”.

and behavioral data extraction (ZUBOFF, 2019, pp. 31, 94, 97, 115)¹⁴. Moreover, in this system, virality frequently supersedes veracity, and content designed to provoke outrage or entertain tends to outperform content intended to inform or foster deliberation. Algorithmic curation also prioritizes emotional intensity and interaction volume (ZUBOFF, 2019, pp. 137–138, 580), often diminishing the visibility of reasoned or meaningful contributions to public debate^{15 16}.

For these reasons, the economic architecture of networks fosters the spread of problematic content, such as misinformation, disinformation, and bullshitting. Despite the frequent conflation of these concepts, they refer to distinct practices. In short, misinformation is defined as the unintentional propagation of false content, whereas disinformation signifies the deliberate dissemination of false or manipulated content intended to attain political or economic ends (EUROPEAN COMMISSION, 2018; ORGANIZATION OF AMERICAN STATES, 2019)¹⁷. Bullshitting, as theorized by Frankfurt (2005) and elaborated by Cassam (2019, p. 80), refers to communicative behavior characterized not by deception but by disregard for truth. In reality, the bullshitter disregards factual accuracy and focuses primarily on persuasion or impression management. Therefore, in an environment as described, where visibility and virality are rewarded, all these forms of content find fertile ground.

Another challenge that must be addressed is the sheer volume of content. Indeed, the sheer scale of digital communication, when considered in conjunction with anonymity, bot activity, and algorithmic personalization, effectively undermines users' ability to discern reliable information and further complicates the already difficult task of moderating online content. As Wardle and Derakhshan (2017) note, distinctions between truth and falsehood are becoming increasingly indistinct, and emotionally charged or absurd claims can attain disproportionate visibility. These conditions erode shared epistemic standards and weaken the basis for meaningful disagreement and democratic legitimacy. Consequently, the collapse of factual consensus engenders cynicism, disorientation, and distrust in both traditional media and institutional authority.

On this topic, Vosoughi, Roy, and Aral (2018) demonstrated that false information spreads faster and more widely than true content, partly because it aligns with group identities and elicits more robust emotional responses¹⁸. Additionally, BuzzFeed News indicated that during the final months of the 2016 U.S. presidential election, the twenty most-engaged fake news stories

¹⁴This paper does not aim to explore the concept of “surveillance capitalism” but, for a detailed account, see Zuboff’s *The Age of Surveillance Capitalism* (2019, part 1.3), which argues that this new economic order unilaterally claims private human experience as free raw material (what she terms “behavioral surplus”) to produce prediction products sold in behavioral futures markets, creating unprecedented asymmetries of knowledge and power.

¹⁵ On this point, Zuboff uses the example of professional journalism (whose purpose is to separate truth from falsehood) to show how Facebook’s News Feed, for instance, treats all content alike, regardless of its truthfulness. According to Zuboff, in this context information disorder is treated as problematic unless it threatens business operations, and moderation is framed as a form of self-protection for the platform rather than public responsibility (2019, p. 138 [online]).

¹⁶ For a comprehensive examination of the complexity of algorithms, see Bucher (2018), who conceptualizes them as sociomaterial assemblages shaped by code, people, and context, embedded with biases, and actively shaping social reality.

¹⁷ For a more detailed discussion in the specialized literature regarding the definitions of fake news and disinformation, as well as a proposed legal definition of the latter, see subsection 2.2.3 in Hijaz (2023, pp. 125-138).

¹⁸ “Falsehood diffused significantly farther, faster, deeper, and more broadly than the truth in all categories of information, and the effects were more pronounced for false political news than for false news about terrorism, natural disasters, science, urban legends, or financial information. We found that false news was more novel than true news, which suggests that people were more likely to share novel information. Whereas false stories inspired fear, disgust, and surprise in replies, true stories inspired anticipation, sadness, joy, and trust” (VOSOUGHI, ROY, ARAL, 2018).

outperformed the twenty most popular news items from traditional media (SILVERMAN, 2016). These empirical studies confirm that platforms do more than merely host public discourse; they shape, filter, and often distort it through design choices and commercial incentives. Therefore, as discussed throughout this section, the information ecology inherently prioritizes content that is captivating over that which is accurate, and content that is polarizing over that which is reasoned.

In this context, the governance of the new digital public sphere is not guided by established public principles such as legality, transparency, due process, or other values deemed fundamental to a society that claims to be democratic. Instead, it is governed by opaque terms of service enforced through discretionary and often automated mechanisms. As Suzor observes, “[t]his is the opposite of the standards we expect of legitimate, legal decision-making in a democratic society” (2019, p. 8). Indeed, “when such companies make decisions about who uses their networks and how, they have almost unlimited discretion” and “[t]hey are accountable only to the market; there are no checks and balances on how they wield their power” (SUZOR, 2019, pp. 6–7). Furthermore, as noted in the introduction of this work, their compliance with public law tends to be reactive and selective, driven by reputational pressure or regulatory threats rather than institutional obligation. Consequently, fundamental decisions about visibility, access, and expression are made by actors who are not bound by constitutional duties and are subject to limited public scrutiny.

This phenomenon creates significant challenges in the enforcement of constitutional rights, including those related to freedom of expression, personal image, or honor against private entities. In instances where users contest the removal of content or the imposition of restrictions, their recourse is generally limited to internal appeals. Given the vast volume of content processed daily, moderation relies heavily on automated tools or rapid human review, restricting the scope for context or subtlety (ROBERTS, 2019, p. 179)¹⁹. Combined with a general lack of transparency, this frequently leads to mistakes. For instance, the iconic Vietnam War photo of the “Napalm Girl” was removed by Facebook due to its depiction of nudity, despite its historical significance, and was reinstated only after public backlash (LEVIN, WONG, HARDING, 2016). Conversely, extremely harmful content such as extremist videos has often remained available on platforms even after being reported. When removed, such content is frequently reinstated through re-uploads, often after having already been widely viewed and shared (LEVIN, 2017). Ultimately, as Keller (2019, p. 2) stresses, “users have few or no legal rights when platforms take down their posts”.

When addressing the enforcement of fundamental rights in the digital sphere, one must consider the triangular dynamic between platforms, states, and users, a relationship that, although it varies by jurisdiction, also reveals common patterns. In the United States, for instance, the First Amendment restricts government action but does not obligate platforms to uphold speech rights internally²⁰. Brazil’s Civil Rights Framework for the Internet (Marco Civil da Internet) similarly

¹⁹We do not intend here to evaluate how good these decisions are or how they should ideally be made, but it is important to note that such decisions are made exclusively based on the platform’s own standards of what constitutes free speech, acceptable discourse, and what does not. This complex situation creates an informational environment where users (and citizens) face insecurity, unpredictability, and vulnerability to arbitrary or inconsistent decisions, weakening the stability and fairness essential to a democratic public sphere (GORWA, 2019; FLEW; MARTIN; SUZOR, 2019).

²⁰In this specific case, the outcome could even be seen as positive, considering that freedom of expression in the United States protects speech in cases where many would consider it reasonable to impose limitations, such as false speech, hate speech, and similar content. Nevertheless, it remains striking that a private platform ultimately defines what speech is acceptable.

grants platforms intermediary immunity, with exceptions limited to court-ordered takedowns²¹. While a singular cause for this phenomenon remains elusive, the absence of clearly delineated and consistently applied criteria to guide enforcement mechanisms and internal decision-making processes within major platforms has given rise to competing claims among scholars. On the one hand, some argue that platforms overstep by censoring dissent²³; on the other, critics contend that they fail to prevent the spread of harmful content²⁴.

Indeed, some initiatives have attempted to address this governance deficit. Meta's Oversight Board, launched in 2020, is often cited as a step toward accountability. Structured as a quasi-judicial body²⁵, the Board reviews a limited number of moderation cases and issues binding decisions. It is composed of independent experts and financed by a trust endowed with \$150 million by Meta, with the stated purpose of operating independently from the company²⁶. However, the scope of its jurisdiction is limited, its policy recommendations are non-binding²⁷, and its role remains embedded within the same corporate infrastructure it is meant to oversee (KLONICK, 2020, p. 2464). Even in high-profile cases, such as the decision to uphold Trump's suspension, the Board's rulings often align with the company's prior choices, raising doubts about independence and institutional legitimacy (REICH; SAHAMI; WEINSTEIN, 2021, p. 215)²⁸.

Accordingly, the issue appears to extend beyond the absence of regulatory oversight. In practice, regulation already exists, but it is implemented by the platforms themselves, resulting in a form of public sphere governance that lacks legitimacy. In this arrangement, the scope of fundamental rights is effectively delegated to corporate actors whose priorities are primarily shaped by market

²¹ According to article 19, "In order to protect freedom of expression and prevent censorship, internet application providers may only be held civilly liable for damages resulting from third-party content if, following a specific court order, they fail to take the necessary measures, within the scope and technical limits of their service and within the time frame established, to make the allegedly infringing content unavailable, except as otherwise provided by law" (Brazil, 2014). Translated by us from: "Com o intuito de assegurar a liberdade de expressão e impedir a censura, o provedor de aplicações de internet somente poderá ser responsabilizado civilmente por danos decorrentes de conteúdo gerado por terceiros se, após ordem judicial específica, não tomar as providências para, no âmbito e nos limites técnicos do seu serviço e dentro do prazo assinalado, tornar indisponível o conteúdo apontado como infringente, ressalvadas as disposições legais em contrário".

²² In contrast, the EU's Digital Services Act (DSA), adopted in 2022, imposes obligations on major platforms to proactively moderate content, address systemic risks, and facilitate user appeals. Despite the challenges associated with its implementation, including a potential lack of coverage of emerging technologies and decentralized networks, the DSA marks a significant shift toward stronger platform responsibility and regulatory dialogue. For a comprehensive overview of this topic, see G'sell (2023).

²³ See, for example, Dickinson (2022), who argues that dominant platforms such as Facebook, Google, and X exert excessive control over public discourse, increasingly acting as private regulators of speech and silencing controversial or unpopular voices under the guise of content moderation.;

²⁴ See Truong et al. (2025), who argue that when social media platforms compete for user engagement, stricter moderation policies may lead to user migration to less regulated spaces, creating disincentives for effective regulation of misinformation and resulting in a collective under-enforcement scenario.

²⁵ When addressing this topic, Reich, Sahami, and Weinstein chose the following subtitle: "A Supreme Court for Facebook?" (2021, p. 213).

²⁶ "In 2019, Meta (then Facebook) established an irrevocable trust and transferred \$130 million for the set-up and operations of the Oversight Board to the Trustees. On July 22, 2022, Meta announced additional funding of \$150 million to be transferred to the Trustees as part of a commitment to provide ongoing financial support to the Oversight Board" (OVERSIGHT BOARD, online).

²⁷ To clarify, the decisions are binding for the specific cases reviewed by the Board, but its general policy recommendations are not.

²⁸ "In late spring 2021, it upheld Facebook's decision to suspend Trump but rejected an indefinite ban. It gave Facebook six months to revisit the case and provide clear, public standards for any continuing ban. A Solomonic decision, it pleased no one and returned power to Facebook. And odd choice if Oversight Board was intended to diminish the unchecked power of Facebook in deciding the boundaries of permissible speech on its platform" (REICH, SAHAMI, WEINSTEIN, 2021, p. 215).

incentives. Meanwhile, national legal systems grapple with the challenge of constraining the transnational reach of these platforms, and international regulatory initiatives remain incipient and weakly binding. We argue that this dynamic creates a regulatory vacuum, wherein critical decisions about fundamental values, expression, truth, and participation are made in the absence of the usual constraints or safeguards that govern analogous public functions. From a philosophical standpoint, it is therefore important to question the moral implications of entrusting such regulatory decisions to a select group of a few transnational corporations. Thus, building on the Kantian argument set forth in the first section — particularly the concept of pockets of a state of nature —, the following will explore the normative implications of this arrangement and argue that a certain form of public regulation could offer a more legitimate and morally adequate response.

3. The political and legal approach to justifying and limiting online platforms

3.1. Why consider the digital environment a pocket of the state of nature?

As discussed in the first part of this paper, Kant's definition of the state of nature is characterized by the exercise of power without public authorization, by uncertain rights, and by a freedom dependent on force or on the convenience of the parties involved. We hold that this description provides a valuable framework for diagnosing the evolving regulatory landscape of major digital platforms, where decisions about what can be said, seen, or removed are made unilaterally through opaque internal procedures and with limited adherence to public legal standards. For instance, companies such as X and Rumble adopt different responses to judicial orders depending on the country, revealing a selective logic guided by political, reputational, or economic interests rather than by the uniform application of law²⁹.

In this context, the hypothesis advanced here is that these platforms operate as pockets of a state of nature — spaces that exist within a formal legal order yet in which certain actors systematically resist the transition to a civil condition by retaining the autonomy to define their own norms. This hybrid — and convenient — configuration combines the benefits of state stability, such as legal infrastructure, economic systems, and contractual protections, with the preservation of unilateral freedom that is not subject to obligations of public justification. In other words, this mode of operation preserves the advantages of civil society while maintaining the absence of constraints characteristic of the savage freedom described by Kant.

As discussed in the second part, the internet has enabled this kind of arrangement: certain groups benefit from public infrastructure, yet whenever regulation conflicts with their interests, they reject state oversight of interactions within digital environments and claim for themselves the exclusive right to determine what they consider legitimate. In turn, the criteria that govern content visibility or removal are shaped by internal rules that are often enforced through automated systems or private bodies without transparency or institutional guarantees of contestation. Consequently, users find themselves subject to unstable norms and decisions that evade any notion of legal transparency, thereby undermining both the predictability and the equality essential to the exercise of freedom of expression.

Thus, from the Kantian perspective developed here, even if certain rights may arise in the digital context, such as the right to access information or to express oneself in widely used virtual spaces,

²⁹ See: Campos Mello (2024); Global Freedom of Expression (2024a; 2024b).

these claims must be evaluated in light of the categorical principle of right, which states: “Act externally in such a way that the free use of your choice can coexist with the freedom of everyone in accordance with a universal law” (MM 06:231). Kant conceives the state of nature not as a historical stage but as a condition defined by the absence of legitimate public authority and the indeterminacy of rights. In the context of digital networks, this means that while the state should not exercise material control over access to platforms, it can — and must — establish formal criteria and ensure that digital interactions conform to the same principles that govern other areas of law, such as the protection of minors, civil liability, and respect for fundamental rights. Insofar as they do not constitute a realm beyond public law, platforms must also be subject to the moral principles of right, guided by the elimination of unilateral coercion and by the public determination of rights and duties.

Therefore, we argue that the concept of *pockets of a state of nature* is particularly relevant in this context, precisely because overcoming such a condition does not entail creating a parallel order, such as a supposed digital civil state, which would merely perpetuate the state of nature among competing jurisdictions. Rather, practical reason requires that there be a single legitimate civil condition. In other words, the only way to overcome these pockets is to extend the principles of public law to the digital environment, thereby ensuring that all subjects are placed under the same legitimate authority.

3.2. Why focus on platforms rather than on individuals?

This work focuses on platforms because they exercise structural power in shaping the discursive environment. When analyzing the normative challenges posed by digital communication, it is essential to distinguish between individuals who express opinions and the agents responsible for structuring the conditions under which those opinions are formulated, circulated, and received. Ultimately, what can be said, by whom, with what reach, and with what consequences is determined by how platforms define usage policies, organize informational flows, and control mechanisms of visibility. Moreover, this power is not exercised through isolated acts but through the continuous management of the architecture of the digital public sphere.

Therefore, just as the state must, according to Kant, secure the conditions for the coexistence of individual freedoms without determining their content, it must also ensure that platforms enable expression by structuring public space without shaping it according to their own interests. Yet, by combining opaque algorithmic ranking criteria, nontransparent internal policies, and unilateral sanctions, platforms assume a role that goes beyond mere technical mediation, directly influencing the conditions under which discourse occurs.

In this context, the decision to analyze platforms as the central object arises from the recognition that the exercise of individual freedom depends on a minimally structured discursive space governed by public and accessible norms. When that space is organized by private actors wielding asymmetric power, individuals become dependent on decisions they neither control nor have the institutional means to contest. This asymmetry undermines users’ external freedom by subjecting them to coercive practices that do not originate from an omnilateral will. For this reason, we argue that, from a normative perspective, regulating platforms requires establishing a set of institutional conditions that ensure freedom of expression is exercised as a form of public freedom.

3.3. Limiting the unlawful freedom of digital platforms as a condition for freedom of thought

Kant was one of the strongest defenders of freedom of thought. According to him, attempts by authoritarian states to restrict freedom of expression inflict profound harm on one of the most fundamental human rights, since freedom of thought is “the sole palladium of the people’s rights” (TP 08:304). Freedom of thought is inseparably linked to freedom of expression. For Kant,

The freedom to think is opposed first of all to civil compulsion. Of course it is said that the freedom to speak or to write could be taken from us by a superior power, but the freedom to think cannot be. Yet how much and how correctly would we think if we did not think as it were in community with others to whom we communicate our thoughts, and who communicate theirs with us! Thus one can very well say that this external power which wrenches away people’s freedom publicly to communicate their thoughts also takes from them the freedom to think (WO 08:144).

Because of the intrinsic connection between freedom of thought and freedom of expression, any attempt to restrict or abolish the right to freedom of expression is always illegitimate, even if it were approved by all citizens, since it could never be reconciled with the omnilateral will. Kant addresses this issue in the context of religious debates, one of the major political and social challenges of his time. He was personally affected by this matter when he faced political persecution and the threat of censorship:

May a people itself make it a law that certain articles of faith and forms of external religion, once adopted, are to remain forever? And so: May a people hinder itself, in its posterity, from making further progress in religious insight or from at some time correcting old errors? It then becomes clear that an original contract of the people that made this a law would in itself be null and void because it conflicts with the vocation and end of humanity; hence a law given about this is not to be regarded as the real will of the monarch, to whom counter representations can accordingly be made (TP 08:305).

The same reasoning applies to enlightenment, for which freedom of thought is likewise essential. In this regard, and echoing the passage above, Kant writes:

[...] to renounce enlightenment, whether for his own person or even more so for posterity, is to violate the sacred right of humanity and trample it underfoot. But what a people may never decide upon for itself, a monarch may still less decide upon for a people; for his legislative authority rests precisely on this, that he unites in his will the collective will of the people (WA 08:39).

Thus, within the framework of public law, the state cannot decide on behalf of individuals what they cannot decide for themselves. The omnilateral will must express the demands of reason, and reason cannot consent to a rule that undermines the very conditions of its possibility³⁰.

Since the concept of freedom of thought has been established as a fundamental basis of the individual’s subjective rights, it is necessary to inquire into the criteria for its realization. In this domain as well, submission to law is required, because the freedom of thought of one individual must be compatible with the freedom of thought of all others. Kant formulated this law through the distinction between the private and public uses of reason. The private use of reason occurs when an individual acts within a public role, broadly defined, which may include the standpoint of an official, an institutional representative, a citizen, or a person fulfilling a particular duty. The

³⁰In this sense, Kant agrees with Rousseau’s view that human beings cannot voluntarily renounce their freedom and become slaves. For both Rousseau and Kant, such a contract would necessarily be invalid, regardless of whether the parties consent to it. A central element of Kant’s philosophy becomes relevant here: the morality of the categorical imperative and the concept of the omnilateral will depend on normative rather than factual conditions. In other words, the issue is not simply what people want, think, or do, but what they are entitled to want, think, or do as equal free beings. Thus, the question of political legitimacy does not concern the will of all, which is illegitimate and may be entirely partial. Rather, it concerns what people could will, provided that certain moral criteria are met to ensure the possibility of a general will (in Rousseau’s terminology) or an omnilateral will (in Kantian terminology).

public use of reason, by contrast, refers to “that use which someone makes of it as a scholar before the entire public of the world of readers” (WA 08:37). The public use of reason requires that one offer reasons and arguments that others could accept, in accordance with what Kant calls the “maxims of common human understanding”, formulated as follows: “1. To think for oneself; 2. To think in the position of everyone else; 3. Always to think in accord with oneself. The first is the maxim of the unprejudiced way of thinking, the second of the broad-minded way, and the third of the consistent way” (CJ 05:294).

According to Kant, the private use of reason must be limited, while its public use must always remain free. Thus, the same individual must make appropriate decisions to comply with certain rules based on the private use of reason. For example, as a driver, one must take the necessary steps to stop at a red light. Yet, at another time and under different circumstances, that same individual may argue that there should be no traffic light at that location, or even that traffic lights should be abolished altogether. In other words, as a passive citizen, the freedom to use reason privately must be limited, whereas as an active citizen, one must be free to use reason to argue for changing the legislation in question.

Often overlooked in the literature is the existence of an inverse proportionality in Kant’s distinction between the private and public uses of reason: the greater the freedom of private use, the less the freedom of public use, and vice versa (see Klein 2023a, 2023b, 2015). If the freedom of the private use of reason is expanded beyond a certain point, such use no longer constitutes a use in which an individual acts within a public role, broadly defined, because there would no longer be any public role to follow, and one would find oneself in a state of nature. The state of nature is precisely characterized by an almost unlimited freedom of the private use of reason, in which there would be no space, or only a very fragile one, for the free exercise of the public use of reason. After all, how could one make free public use of reason if other individuals constantly restricted or undermined such use through the private exercise of their own reason? Conversely, the greater the space for the freedom of the public use of reason, the more the freedom of its private use must be restricted. If, for instance, we wish to have spaces where the very existence of God can be freely discussed and questioned, to invoke a fundamental issue that has long served as a banner for “holy wars”, then the freedom of the private use of reason by believers must be restricted so that it does not undermine or limit such a space.

Therefore, we can conclude that for freedom of thought to exist, it is necessary to distinguish between the private and public uses of reason. The private use of reason should be subject to political and social constraints, whereas the public use of reason must remain entirely unrestricted. The only limits on the public use of reason are those imposed by the inherent standards of rationality and by the specific context in which it is exercised. Consequently, although the freedom of the public use of reason cannot be censored or placed under authoritarian control, it must still respect the rules that allow it to coexist with the public exercise of freedom by all others. These rules include the maxims of common human understanding, the principles of logic, and humanity’s accumulated knowledge. Kant never elaborated on the precise normative content of these rules, perhaps because it varies depending on the context. For example, arguing as a citizen for changing a traffic regulation is one thing, while engaging in a debate about the safety and efficacy of a vaccine is quite another. What counts as a legitimate public use of reason, therefore, may vary from one context to another.

In this article, we will not develop the general rules governing the public use of reason in the digital context, as that will be left for another occasion. Here, we argue that digital platforms cannot be regarded solely as agents exercising the freedom of public reason, but rather as loci where individuals and institutions — each of which may be considered moral persons with rights and duties — exercise their freedom in both the private and public uses of reason. In this sense, platforms themselves engage in a private use of reason, which must be determined by the principles of the omnilateral will. In other words, platforms must function as spaces that make possible the coexistence of the public and private uses of reason by other agents. This does not deny digital platforms the right to uphold a particular worldview; however, they may do so only while respecting the rules of public reason. Failure to observe this limitation results in what we have referred to above as a pocket of the state of nature.

The actions of digital platforms, which represent their private use of reason, must be limited in order to uphold the freedom of thought of all users. The public function of platforms is to facilitate the free public use of reason by their users. Rather than determining the content of discourse, platforms should guarantee the conditions under which users can freely exercise their reason without coercion or manipulation. If they fail to do so, whether by promoting or tolerating a form of lawless or “savage” freedom, they should be subject to sanctions and, if necessary, removed from circulation. Just as an individual may be punished for committing a crime against another, digital platforms can be held accountable for providing a space in which the use of reason becomes incompatible with the categorical imperative of right. By preserving lawless freedom, digital platforms create a pocket of the state of nature in which users cause one another harm to the highest degree.

3.4. Why is education not sufficient to overcome the digital state of nature?

One could argue that information disorder might be resolved through virtual ethical education or civic training aimed at fostering user responsibility and critical engagement. Although such initiatives are highly relevant, they remain insufficient from a normative standpoint. As discussed in the first part of this work, the injustice of the state of nature lies not only in its empirical outcomes but also in its structural incapacity to establish reciprocal duties under juridical conditions. Therefore, the issue cannot be reduced to conditional reasoning such as “if people are educated, they will act critically and autonomously, and the problem will be solved”, because the digital state of nature remains morally unacceptable even in the presence of ethically responsible individual behavior. Indeed, the absence of legitimate public authority and of a shared legal framework renders every right uncertain and every duty unstable, which is incompatible with the principles of civil freedom.

Even if an ethically educated virtual community could avoid some of the negative effects of the digital state of nature, the condition would still be unjust, since it fails to clearly establish reciprocal rights and duties in accordance with the systematic perspective required for civil freedom. At most, networks could reach agreements based on the will of all, which, for Rousseau, could be profoundly unjust and, for Kant, would remain a partial and contingent will. However, they would never satisfy the criteria of an omnilateral will, which is consistent with the categorical imperative of right. Any such agreement would be contingent, as it would fail to respect the intrinsic principles that regulate external freedom, such as publicity, political representation, and

the balance of powers among the different branches of the state. In short, it would not meet the structural criteria of the concept of a just state or of a republic.

Therefore, a digital sphere based solely on voluntary coordination or on initiatives led by the platforms themselves would not meet the requirements of civil freedom, at least not from the theoretical perspective adopted here. What is required is not merely ethical behavior within an already existing architecture, but a transformation of that very architecture so that it no longer depends on private discretion and becomes subject to principles that can be publicly justified and applied equally. Once again, we emphasize that education may help to reduce harm, but it cannot substitute for legitimate legal conditions. Ultimately, only a transition to a civil condition, through juridical extension and structural reform, can overcome the pockets of unilateral power that characterize the online state of nature.

4. Final Remarks: Toward a Civil Form of Online Freedom

According to a Kantian conception, freedom must be exercised under conditions that ensure reciprocal recognition among subjects and remain within the bounds of shared public norms. When asserted unilaterally, on the basis of force or self-interest, freedom cannot be regarded as legitimate. Applied to the digital environment, this requirement calls into question the current organization of platforms, which are structured by commercial incentives and sustained by decisions that are not subject to any form of public deliberation or public right. As we have sought to demonstrate, these companies have assumed regulatory functions, organizing the flow of information according to their own rules and concentrating power in ways that evade the requirements of publicity, justification, and equality that define the civil condition.

Therefore, we argued that the concept of online freedom should be reconsidered in light of a model that preserves individual autonomy while incorporating it into an institutional framework designed to safeguard equal access, normative predictability, and public justification. Moreover, we maintained that civil freedom is not defined by the absence of interference, but by the existence of a legitimate order that distinguishes between compatible and incompatible uses of freedom. This means that speech must circulate within boundaries that can be accepted by all participants, even in digital environments. In Kantian terms, the exercise of freedom requires a transition from the digital state of nature to a public order guided by shared principles and the concept of equal freedom.

In this context, the emphasis falls far more on the architecture of platforms than on the content of individual messages. It is also important to stress that the problem does not arise from the plurality of opinions, but from the asymmetric conditions under which that plurality is expressed. Therefore, structuring a space that enables the public use of reason requires publicly defined criteria for organizing visibility, moderation, and access. To reiterate, the aim is not to impose substantive truths — since everything must remain open to discussion in accordance with the freedom of the public use of reason — but rather to establish a shared institutional foundation. To *civilize digital freedom*, in this sense, means creating an order that protects expression rather than shaping it, ensures equal conditions rather than dictating content, and provides legal security rather than promoting discursive uniformity.

However, this transformation faces significant obstacles. For instance, the transnational nature of platforms makes it difficult to enforce public norms consistently, particularly when national

legal systems operate according to different scopes and values. Although initiatives such as the European Union's Digital Services Act represent steps toward the institutional governance of the digital sphere, they still lack global effectiveness and encounter political resistance. Furthermore, a persistent imaginary of absolute freedom on the internet, which in Kantian terms would correspond merely to an ideal of imagination that in reality amounts to nothing more than savage freedom, is sustained by both corporate interests and certain segments of public opinion, making it more difficult to develop an institutional conception of responsibility. Although philosophy does not offer direct solutions to these challenges, nor does it claim to do so, we have argued that it can contribute to the debate by providing normative criteria for assessing the legitimacy or illegitimacy of platform conduct in light of the principles of public right.

In the virtual context of digital networks, this means that the state cannot intervene to determine, in a general or material sense, who may or may not have access to platforms, but it can establish formal criteria for their use. For example, it may set a minimum age for joining a social network, as in a proposed bill in Australia that seeks to restrict access for minors, or define what distinguishes an offensive opinion from a hate crime. It is the role of the state to enact general laws that regulate digital interactions in accordance with the broader legal framework governing other domains of law. In short, since digital networks are not exempt from public law, they must also be subject to the moral principles of public right, which are grounded in the moral obligation to determine rights and duties and to eliminate unilateral violence in accordance with the principle of equal freedom.

Bibliographic References

- BRAZIL. 2014. *Law no. 12,965 of April 23, 2014*, [Online]. Available at: https://www.planalto.gov.br/ccivil_03/_ato2011-2014/2014/lei/112965.htm. Accessed: June 2, 2025.
- BUCHER, T. 2018. *If... Then: Algorithmic Power and Politics*. New York: Oxford University Press.
- CAMPOS MELLO, P. 2024. Musk cumpriu centenas de ordens de remoção de conteúdo do X fora do Brasil sem acusar censura. *Folha de São Paulo*. Available at: <https://www1.folha.uol.com.br/poder/2024/04/musk-cumpriu-centenas-de-ordens-de-remocao-de-conteudo-do-x-fora-do-brasil-sem-acusar-censura.shtml>. Accessed: December 17, 2024.
- CASSAM, Q. 2019. *Vices of the Mind. From the Intellectual to the Political*. Oxford: Oxford University Press.
- CASTELLS, M. 2018. *A sociedade em rede*. Trad. Roneide Venancio Majer, 19ª ed. Rio de Janeiro/São Paulo: Paz e Terra.
- CASTELLS, M. 2003. *A galáxia da internet: reflexões sobre a internet, os negócios e a sociedade*. Trad. Maria Luiza X. de A. Borges. Rio de Janeiro: Zahar.
- CONSANI, C. F.; KLEIN, J. T. 2022. *Leituras de Rousseau*. Florianópolis: Nefiponline. Available at: <https://nefipo.paginas.ufsc.br/files/2012/11/LEITURAS-DE-ROUSSEAU.pdf>. Accessed: July 20, 2025.
- CONSANI, C. F. 2015. Democracia e os discursos de ódio religioso: O debate entre Dworkin e Waldron sobre os limites da tolerância. *ethic@ - An international Journal for Moral Philosophy*, Florianópolis, v. 14, n. 2, pp. 174–197, dec. Available at: <https://periodicos.ufsc.br/index.php/ethic/article/view/1677-2954.2015v14n2p174>. Accessed: July 20, 2025.
- DICKINSON, G. M. 2022. Big Tech's Tightening Grip on Internet Speech, [Online]. *Indiana Law Review*, v. 55, n. 1, p. 101–128. Available at: <https://doi.org/10.18060/26418>. Accessed: June 2, 2025.
- EUROPEAN COMMISSION. 2018. *A multi-dimensional approach to disinformation: report of the independent High level Group on fake news and online disinformation*, [Online]. European Commission. Available at: <https://digital-strategy.ec.europa.eu/en/library/final-report-high-level-expert-group-fake-news-and-online-disinformation>. Accessed: June 2, 2025.
- FLEW, T.; MARTIN, F.; SUZOR, N. 2019. Internet regulation as media policy: Rethinking the question of digital communication platform governance. *Journal of Digital Media & Policy*, v. 10, n. 1, pp. 33-50. Available at: https://doi.org/10.1386/jdmp.10.1.33_1. Accessed: June 2, 2025.
- FRANKFURT, H. G. 2005. *On Bullshit*. Princeton, NJ: Princeton University Press.
- FRANKS, M. A. 2022. The Free Speech Industry, pp. 65-86. In: BOLLINGER, L. C.; STONE, G. R. (Editors). *Social Media, Freedom of Speech, and the Future of Our Democracy*. Oxford: Oxford University Press.
- GLOBAL FREEDOM OF EXPRESSION. 2024a. Columbia University. *The Case of the X Ban in Brazil*. Available at: <https://globalfreedomofexpression.columbia.edu/cases/the-case-of-the-x-ban-in-brazil/>. Accessed: March 8, 2025.

GLOBAL FREEDOM OF EXPRESSION. 2024b. Columbia University. *The Case of the Rumble Ban in Brazil*. Available at: <https://globalfreedomofexpression.columbia.edu/cases/the-case-of-the-rumble-ban-in-brazil/>. Accessed: June 6, 2025.

GORWA, R. 2019. The platform governance triangle: conceptualising the informal regulation of online content. *Internet Policy Review*, n. 8, v. 2. DOI: 10.14763/2019.2.1407. Available at: <https://policyreview.info/articles/analysis/platform-governance-triangle-conceptualising-informal-regulation-online-content>. Accessed: July 10, 2025.

G'SELL, F. 2023. The Digital Services Act (DSA): A General Assessment, [Online]. In: VON UNGERN-STERNBERG, A. (ed.). *Content Regulation in the European Union – The Digital Services Act*. Trier Studies on Digital Law: Verein für Recht und Digitalisierung e.V., Institute for Digital Law (IRDT). Available at SSRN: <https://ssrn.com/abstract=4403433> or <http://dx.doi.org/10.2139/ssrn.4403433>. Accessed: June 3, 2025.

HIJAZ, T. 2023. *Quanto Vale a Liberdade? O problema da desinformação entre os diferentes fundamentos da liberdade de expressão*. São Paulo: Dialética.

KANT, I. *Cambridge Edition of the Works of Immanuel Kant*. Paul Guyer and Allen W. Wood (ed.). Cambridge: Cambridge University Press, 1996–.

KANT, I. *Gesammelte Schriften*. Ed. Königlich Preussische Akademie der Wissenschaften (and successors). Berlin: de Gruyter (and predecessors), 1900–.

KELLER, D. 2019. Who do you sue? State and platform hybrid power over online speech, [Online]. *Aegis Series Paper No. 1902*. Hoover Institution. Available at: <https://www.hoover.org/research/who-do-you-sue-state-and-platform-hybrid-power-over-online-speech>. Accessed: June 2, 2025.

KERSTING, W. 1994. *Die politische Philosophie des Gesellschaftsvertrags*. Darmstadt: Wissenschaftliche Buchgesellschaft.

KLEIN, J. T. 2018. Kant on religious intolerance. *Philosophica* (Lisbon), v. 51, pp. 25–38. Available at: https://repositorio.ulisboa.pt/bitstream/10451/40702/1/JoelKlein_Philosophica_51.pdf. Accessed: July 20, 2025.

KLEIN, J. T. 2015. Freedom of the Press: A Kantian Approach. *Estudos Kantianos*, v. 03, p. 83–92, 2015. Available at: <https://revistas.marilia.unesp.br/index.php/ek/article/view/5122>. Accessed: July 20, 2025.

KLEIN, J. T. 2023a. Enlightenment as the normative principle of social rationality. *Studia Kantiana*, v. 21, p. 99–117. Available at: <https://revistas.ufpr.br/studiakantiana/article/view/91982>. Accessed: July 20, 2025.

KLEIN, J. T. 2023b. Liberdade versus irracionalidade acadêmica: uma análise a partir de um ponto de vista Kantiano. *Ethic@* (UFSC), v. 22, p. 691–716. Available at: <https://periodicos.ufsc.br/index.php/ethic/article/view/95262/54867>. Accessed: July 20, 2025.

KLONICK, K. 2018. The New Governors: The People, Rules, and Processes Governing Online Speech, [Online]. *Harvard Law Review*, v. 131, p. 1598–1670. Available at: https://harvardlawreview.org/wp-content/uploads/2018/04/1598-1670_Online.pdf. Accessed: March 16, 2025.

KLONICK, K. 2020. The Facebook Oversight Board: Creating an Independent Institution to

Adjudicate Online Free Expression, [Online]. *The Yale Law Journal*, v. 129, n. 8, p. 2418–2499. Available at: <https://www.yalelawjournal.org/feature/the-facebook-oversight-board>. Accessed: March 16, 2025.

LEVIN, S.; WONG, J. C.; HARDING, L. 2016. Facebook backs down from ‘Napalm Girl’ censorship and reinstates photo, [Online]. *The Guardian*. Available at: <https://www.theguardian.com/technology/2016/sep/09/facebook-reinstates-napalm-girl-photo>. Accessed: June 3, 2025.

LEVIN, S. 2017. Tech firms fail to stop abusive content – leaving the public to do the dirty work, [Online]. *The Guardian*. Available at: <https://www.theguardian.com/technology/2017/dec/05/youtube-offensive-videos-journalists-moderators>. Accessed: June 3, 2025.

LÉVY, P. 2015. *A inteligência coletiva: por uma antropologia do ciberespaço*. Trad. Luiz Paulo Rouanet, 10^a ed. São Paulo: Edições Loyola.

META. 2025. *Meta Reports First Quarter Results*. Available at: https://s21.q4cdn.com/399680738/files/doc_news/Meta-Reports-First-Quarter-2025-Results-2025.pdf. Accessed: June 6, 2025.

ORGANIZATION OF AMERICAN STATES. 2019. *Guia para garantir a liberdade de expressão frente à desinformação deliberada em contextos eleitorais*, [Online]. OAS. Available at: <https://www.oas.org/es/cidh/expresion/publicaciones/DesinformacionElectoral.pdf>. Accessed: June 2, 2025.

OVERSIGHT BOARD. *How Is the Oversight Board Funded?* [Online]. Available at: <https://www.oversightboard.com/faq/>. Accessed: June 3, 2025.

PERSILY, N. 2022. Platform Power, Online Speech, and the Search for New Constitutional Categories, pp. 193-212. In: BOLLINGER, L. C.; STONE, G. R. (Editors). *Social Media, Freedom of Speech, and the Future of Our Democracy*. Oxford: Oxford University Press.

RAWLS, J. 1971. *A Theory of Justice*. Cambridge, MA: Harvard University Press.

REICH, R.; SAHAMI, M.; WEINSTEIN, J. M. 2021. *System Error*. Where Big Tech Went Wrong and How We Can Reboot. New York: HarperCollins.

ROBERTS, S. T. 2019. *Behind the Screen: Content Moderation in the Shadows of Social Media*. New Haven: Yale University Press.

SILVERMAN, C. 2016. This Analysis Shows How Viral Fake Election News Stories Outperformed Real News On Facebook, [Online]. *BuzzFeed News*, November 16. Available at: <https://www.buzzfeednews.com/article/craigsilverman/viral-fake-election-news-outperformed-real-news-on-facebook#.etwaV6WDZq>. Accessed: May 28, 2025.

SUZOR, N. 2019. *Lawless: The Secret Rules That Govern Our Digital Lives*. Cambridge: Cambridge University Press.

TRUONG, B. T.; KIM, S.; NOGARA, G.; VERDOLOTTI, E.; SAHNEH, E. S.; SAURWEIN, F.; JUST, N.; LUCERI, L.; GIORDANO, S.; MENCZER, F. 2025. Delayed takedown of illegal content on social media makes moderation ineffective. Technical Report. *arXiv preprint*, arXiv:2502.08841. Available at: <https://doi.org/10.48550/arXiv.2502.08841>. Accessed: June 2, 2025.

VOSOUGHI, S.; ROY, D.; ARAL, S. 2018. The spread of true and false news online, [Online]. *Science*, v. 359, n. 6380. Available at: <https://doi.org/10.1126/science.aap9559>. Accessed: February 12, 2025.

WARDLE, C.; DERAKHSHAN, H. 2017. *Information Disorder: Toward an Interdisciplinary Framework for Research and Policy Making* [Online]. Council of Europe Report. Available at: <https://edoc.coe.int/en/media/7495-information-disorder-toward-an-interdisciplinary-framework-for-research-and-policy-making.html>. Accessed: February 14, 2025.

ZUBOFF, S. 2019. *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. New York: PublicAffairs.