



The Algorithmic Cage Artificial Intelligence, Modern Rationalization, and the Inheritors of Max Weber

Carlos F. De Angelis

Facultad de Ciencias Sociales – UBA, Argentina

cfdeangelis@gmail.com

<http://dx.doi.org/10.47814/ijssrr.v9i5.3369>

Abstract

This essay examines, from a sociological and philosophical perspective, the hypothesis that artificial intelligence (AI) constitutes a new phase in the modern rationalization process theorized by Max Weber, and that the metaphor of the iron cage (stahlhartes Gehäuse) can be reformulated in contemporary terms as an algorithmic cage. It discusses whether algorithmic systems replace, deepen, or transform classical bureaucratic logic, and examines their effects on subjectivity, power, and freedom. The argument is articulated with contributions from Foucault, Habermas, Zuboff, Han, Srnicek, and Winner, along with recently produced specialized literature (2019–2025). The final thesis maintains that AI does not break the Weberian cage, but rather reconfigures it as a more capillary, opaque, and affectively seductive architecture, although this does not close off the margins of autonomy and democratic resistance.

Keywords: *Artificial Intelligence; Rationalization; Iron Cage; Algorithmic Governance; Surveillance Capitalism; Psychopolitics; Bureaucracy; Autonomy*

Introduction

In 1905, at the close of the second chapter of *The Protestant Ethic and the Spirit of Capitalism* (2012), Max Weber warned that modern economic rationality had constructed a cage as hard as steel within which men would be forced to live as long as the capitalist order persisted. That image, translated into English by Talcott Parsons as “iron cage” (in Spanish, jaula de hierro), has become one of the densest and most productive metaphors in classical sociology. It encapsulates a diagnosis of modernity: that calculation, legal formalization, impersonal administration, and market logic end up colonizing spheres of life that were originally based on substantive values, moral communities, or shared meanings.

Max Weber made a fundamental contribution to the analysis of bureaucratization, considering it a central and novel element in the development of the modern state and the capitalist economy. Underlying his sociological perspective is the idea that modern administrative organization relies on a specialized

bureaucracy that historically formed on the basis of the division of labor, as a class distinct from the owners and the workers. Weber explains that the bureaucratic state is characterized by a profound structural division in which the administrative staff does not own the resources it uses to govern. This political organization is entirely analogous to modern economic organization, since “in the capitalist enterprise, workers and employees are separated from the material means of production” (Weber, 2012, p. 12).

For Weber, bureaucratic domination transcends the sphere of government and extends to other groups such as the Church or capitalist economic organizations. The bureaucratic system operates through impersonal statutory regulations and rationally delimits the areas of jurisdiction for each office. Unlike the patriarchal orders of the past, which were based on personal duties of piety or submission, the modern bureaucratic state administers public affairs objectively. In this sense, bureaucratic legal authority requires that the official act “without regard for the person, *sine ira ac studio*, without hatred and, therefore, without love” (2012, p. 61).

Despite acknowledging its undeniable technical necessity, Weber expresses deep concern about the political dangers of unchecked bureaucratization. The sociologist points out that an uncontrolled dominance of officials suppresses a nation's true capacity for political leadership, preventing it from being guided by matters of historical or cultural significance. According to his analysis, the official is prepared to manage with objective efficiency and obey blindly as part of a machine but lacks the charismatic qualities and independence of a true leader. While the bureaucrat avoids high-risk decisions, the authentic politician requires a warm passion coupled with cool responsibility and a sense of proportion. Thus, Weber warns that a parliament and a government dominated exclusively by bureaucrats lack genuine political orientation, since politics is conducted with the head, not the soul.

More than a century later, this metaphor must be invoked again in the context of a phenomenon that Weber could not have foreseen in its current form: the massive deployment of artificial intelligence systems in institutions, markets, and everyday life. Machine learning algorithms are imperceptibly taking over the tasks of civil servants: they rate loan applicants, modulate news flows, diagnose illnesses, assign work shifts, calculate criminal recidivism, recommend partners, decide priorities in hospitals, and arbitrate what is visible in public spaces. Generative AI (from ChatGPT to the multimodal systems that will follow it in the future) has extended this logic to symbolic production and intellectual work itself. The question that organizes this essay is whether such developments represent an emancipatory break with the modern bureaucratic order, a heightened continuity, or something qualitatively different that requires new theoretical tools.

The hypothesis presented here is that artificial intelligence can be interpreted, in an expanded Weberian sense, as a new phase of modern rationalization. In this phase, calculation, prediction, automation, and algorithmic governance reorganize institutions, markets, bureaucracies, and subjectivities, extending the disenchantment of the world into territories that classical bureaucracy barely touched: attention, desire, affectivity, and imagination. However, this is neither technological determinism nor an identical repetition of Weber's diagnosis. As the discussion will show, the algorithmic cage presents structural differences from the bureaucratic cage: it operates through seduction rather than coercion, through personalization rather than visible standardization, and its power architecture is contested by transnational corporate actors that have no exact counterpart in Weber's theory of the state.

The essay is organized into nine brief thematic sections. The first seven explore the proposed axes: conceptual foundation, bureaucracy/algorithm transition, instrumental rationality, subjectivity, power, freedom, and historical comparison. The eighth expands the theoretical debate by incorporating Foucault, Habermas, Zuboff, Han, Srnicek, and Winner. The ninth offers a nuanced and well-reasoned thesis. Throughout the text, contemporary empirical references and recent academic literature are included.

1. Conceptual Basis: Weber's Iron Cage

The notion of an iron cage (*stahlhartes Gehäuse*) appears in the last chapter of *The Protestant Ethic* (2012), but its meaning is only fully understood when articulated within the conceptual system developed in *Economy and Society* (2012). The cage is the metaphorical crystallization of a far-reaching historical process: the rationalization of Western behavior and institutions. Weber distinguishes between substantive rationality (oriented towards ultimate values, moral, religious, or political ends) and formal or instrumental rationality (oriented towards means-ends calculations, predictability, and efficiency). Modernity, he argues, is characterized by the systematic expansion of the latter at the expense of the former: “The entire history of the development of the modern state, in particular, is identified with that of the modern civil service and the bureaucratic enterprise (see *below*), just as the entire evolution of modern large-scale capitalism is identified with the increasing bureaucratization of economic enterprises. The participation of forms of bureaucratic domination is on the rise everywhere” (Weber, 2012, p. 1271).

This expansion is not merely a progress of the scientific spirit. It has a concrete institutional support: modern bureaucracy as the ideal type of legal-rational domination. In contrast to traditional and charismatic forms of domination, legal-rational domination rests on the belief in the legality of ordinances and in the right to command of those who exercise impersonal functions. Its typical features are a clear hierarchy, written rules, separation between office and person, archives, technical qualifications, and promotion based on formal merit. Note that many of these characteristics are shared with algorithms, and like the latter, they constitute a machinery oriented toward calculability. From a purely technical point of view, bureaucracy is the most rational type of exercise of domination, comparable to a machine in relation to non-mechanized modes of organization (2012), as Weber explains: “the separation of the means of administration is carried out in exactly the same way in public bureaucracy and in private bureaucracy (for example: in large capitalist enterprises)” (Weber, 2012, p. 303).

The cost of efficiency is a specific form of loss. Bureaucracy treats people as cases (interchangeable files) and reduces judgment to the application of rules. Formal rationalization colonizes spheres that traditionally responded to a different logic: law becomes procedural calculation, science decouples from any teleology, and the capitalist economy demands systematic accounting that dissolves previous community ties. This is what Weber, drawing on Schiller, called the disenchantment of the world (*Entzauberung der Welt*) the retreat of ultimate meanings and the primacy of technically legitimized procedures (Schluchter, 2017).

The iron cage is a constitutive paradox of modernity. Rationalization freed individuals from the chains of social class and religion but confined them to a new order where subjective freedom is subordinated to the functioning of impersonal systems. Weber recognized the civilizational advantages of legal-rational administration. His concern was political and existential; he feared a world dominated by soulless specialists, heartless hedonists, in which the question of meaning (what is worth living for, what ends to pursue collectively) would be displaced by the technical question of the most efficient means. The cage is not only an institutional structure; it is also a cultural and subjective enclosure.

Three features of this diagnosis are crucial for the contemporary discussion on AI. First, rationalization is a process, not a state: it advances through successive shifts, colonizing areas previously governed by other logics. Second, its effects are ambivalent: it simultaneously produces efficiency and alienation, freedom from traditional constraints and a new submission to impersonal rules. Third, the cage is not indestructible: Weber believed that only a new prophet or a charismatic emergence could break it, a scenario he himself viewed with skepticism. This last point, as we will see, will be critically reformulated by subsequent scholarship.

2. *From Bureaucracy to Algorithms*

The initial question is whether algorithmic systems replace or exacerbate traditional bureaucracy. The prevailing answer in recent literature is that they exacerbate it: that AI does not abolish bureaucracy but rather radicalizes it in a new form, which some authors call algorithmic bureaucracy (Vogl, 2020) or digital (Muellerleile, 2018) Weberianism. Mökander and Schroeder (2024), in a study on fiscal optimization through AI in the public sector, explicitly argue that the implementation of AI systems in public administration should be understood as a continuation and intensification of long-standing processes of rationalization and bureaucratization.

Continuity is structural. Weberian bureaucracy rested on four pillars: explicit rules, hierarchy, records, and expertise. Algorithmic systems reproduce these but transform them. Explicit rules give way to statistical models whose parameters are adjusted through learning from large volumes of data: the rule is no longer stated, it is inferred. Formal hierarchy coexists with technical and ownership hierarchies of the models (those who design the system's architecture retain a structural power that no desk clerk possesses). Records become data profiles, infinitely updatable. And expertise shifts its bearers: from the lawyer or administrator with a professional degree to the data engineer, the statistician, and the *machine learning team*.

But there is also a qualitative intensification. Weberian bureaucracy typically operated on aggregated populations (taxpayers, recruits, patients) using relatively stable categories. Algorithmic governance operates on micro-segmented individuals in real time: each user is subject to a score, a prediction, and a possible intervention. Eubanks (2018) has documented for the case of the American welfare state (how social service allocation algorithms produce shame, anxiety, and constant surveillance) illustrates this difference: classical bureaucracy was slow and massive; algorithmic bureaucracy is instantaneous and personalized. Its disciplinary effect is no less for being more fluid.

Three contemporary empirical fields illustrate this shift with particular clarity. The first is that of digital platforms as quasi-bureaucratic infrastructures. Uber, Rappi, Amazon Mechanical Turk, and Glovo organize the work of millions of people through opaque algorithmic rules: they assign tasks, evaluate performance, sanction behavior, and set dynamic prices. As Srnicek has shown (2017) in *Platform Capitalism*, what is decisive is not technological novelty but the capture of economic coordination infrastructures by a few actors. The platform-company functions as a new form of rational-legal organization in which the employment contract gives way to terms and conditions of use.

The second field is credit, insurance, and social scoring. Systems like FICO in the United States, European banking scoring models, personalized insurance through automotive telematics, and the controversial Social Credit System in China (Liang, 2018) configure algorithmically differentiated citizenships. Every interaction (paying a bill, driving, posting on social media) feeds into a score that opens or closes future opportunities. Curran (2023) has synthesized this phenomenon under the notion of systemic surveillance risk: when the digital economy is sustained by imperatives to collect and connect, system failures (algorithmic errors, biases, security breaches) cease to be anomalies and become structural features.

The third area is that of recommendation systems and attentional flows. YouTube, TikTok, Spotify, and Netflix do not just classify content: they organize, through algorithms optimized for usage time, what millions of people watch, listen to, and remember. This point will be revisited in the section on subjectivity, but it is worth noting its bureaucratic and dehumanizing dimensions (for lack of a better term). Each of these systems functions as a massive classification machine that imposes relevance criteria without any public deliberation. The difference between this and an official applying a written rule is

significant, but so is the analogy: in both cases, a formal rationality operates, deciding cases according to rules that the affected party has not co-created.

In short, the algorithm does not destroy Weberian bureaucracy: it fulfills it in ways that Weber would have considered logical but also unsettling. It delivers on its promise of calculability by taking it to a level of granularity that analogical bureaucracy never achieved. And it also fulfills its dark side: the growing substantive opacity under the guise of procedural neutrality.

3. New instrumental rationality: optimization, quantification and displacement of judgment

The Weberian distinction between formal and substantive rationality allows us to ask whether AI extends the primacy of instrumental rationality to encompass spheres that previously belonged to deliberation, judgment, or values. The answer is yes, albeit with nuances. Instrumental rationality, in its algorithmic form, is expressed in three operations: optimization, quantification, and prediction. Optimization involves the prior formulation of an objective function (click-through rate, revenue per trip, expected repeat business, patient satisfaction measured by questionnaires) that the system will seek to maximize. Quantification translates relationships, behaviors, capabilities, and experiences traditionally conceived in other languages into metrics (Espeland, 2008; Mau, 2019). Prediction (the heart of supervised machine learning) infers future behaviors from past regularities. All three operations presuppose that the world can be modeled as a set of measurable and combinable variables.

From a Weberian perspective, the crucial point is that this rationality is not ideologically neutral: each objective function embodies a value choice that is rarely discussed publicly. When a streaming platform optimizes attention span, it is prioritizing one value (capturing the user's time) over other possible values (cultural diversity, aesthetic quality, civic education). When a US judicial system uses COMPAS to predict recidivism, it implicitly accepts that future recidivism is the relevant metric, rather than, for example, rehabilitation or reparations for victims. What quantification obscures is not the technical dimension but the underlying axiological decision.

The concept of quantified Mau's (2019) society and the complementary notion of datafication (Mejias, 2024) allow us to clarify this point. Datafication is not merely the conversion of the world into data: it is an epistemic order that establishes what counts as legitimate knowledge. In it, narratives, testimonies, expert intuitions, and political deliberations tend to be devalued in the face of quantitative evidence. Human judgment (prudent, situated, capable of recognizing the irreducibility of the particular case) is subordinated to the statistical average. Aristotle (1995) would have said that *phronesis* is replaced by generalized *techné*; Weber, that substantive rationality is displaced by formal rationality. This displacement has concrete political consequences. Algorithmic governance tends to present intrinsically political issues as technical problems. Who should have access to social housing? What criteria define the risk of a migrant? What is the value of a detainee's freedom of movement compared to the statistical risk of escape? When these questions are translated into algorithms, their deliberative character evaporates. As Aradau and Blanke (2022) and Yeung (2018) explain, governance by design produces a structural depoliticization: it turns value conflicts into optimization problems.

However, it would be a mistake to think that substantive rationality is entirely eliminated. Rather, something paradoxical occurs efforts to make AI fair, transparent, and accountable, the so-called field of *fairness, accountability, and transparency in machine learning* (FAccT), operate within the same formalist framework they criticize. For example, there are attempts to translate fairness into computable metrics (demographic parity, equality of opportunity) when the philosophical notion of justice precisely resists this kind of reduction (Mökander, 2024). This paradox (combating rationalization with more rationalization) is typically Weberian: it shows the extent to which the cage is not broken simply by piling additional rules inside.

4. *Subjectivity and Discipline: From Calculation to the Modulation of Desire*

The Weberian cage produced a specific type of subject: the disciplined civil servant, the specialized professional, the punctual worker, the law-abiding citizen. This production of subjectivity operated through relatively external mechanisms: regulations, sanctions, hierarchical supervision. The contemporary question of how subjectivity is produced in the algorithmic cage (a question developed by authors such as Foucault, Han, Rouvroy, and Stiegler) points to a qualitative shift: external discipline gives way to an internal modulation of desire, attention, and self-image.

Towards the end of his work, Foucault had already announced the shift from disciplinary societies to control societies (1977; 2010). The former confined individuals within institutions (schools, factories, hospitals, prisons) to produce docile bodies. The latter operate openly, through the continuous modulation of variables (scores, access, rewards). Algorithmic governmentality, in the terms of Rouvroy and Berns (2016), dispenses even with the reflective subject: it seeks neither to convince nor to discipline, but rather to anticipate and modulate behaviors as they emerge, before they become deliberate.

Byung-Chul Han has developed a relevant formulation of this shift. In *Psychopolitics* (Han, 2014), he argues that the contemporary regime operates not through repression but through seduction: intelligent power makes people submit voluntarily because they believe they are free. The smartphone is a digital rosary, the "like" is the amen of the new faith, and Big Data allows the psyche to be conditioned at a pre-reflective level. The subject is no longer the exploited worker, but the entrepreneur of the self who exploits himself in the name of performance, transparency, and self-optimization. Prison is replaced by shopping, surveillance, and by voluntary connection.

While Han emphasizes the psychological and philosophical dimension of this shift, Zuboff (2019) articulates it as a new economic order: surveillance capitalism. In his analysis, large technology corporations extract behavioral surplus from users' everyday activity, process it in AI infrastructures, generate predictive products, and sell them in behavioral futures markets. The ultimate goal is no longer just to predict behavior, but to modify it: algorithms are designed to act upon users through notifications, gamification, variable reinforcement, and nudges whose effectiveness is measured in engagement metrics. What Zuboff calls *the Big Other* is a ubiquitous architecture that operates outside the bounds of democratic control.

Three subjective consequences deserve highlighting. The first is the weakening of attention. Crary (2013) and Williams (2018), among others, have documented how the attention economy erodes the capacity for prolonged concentration, deep reading, and contemplative thought. This is not an individual phenomenon but a structural one: the very architecture of the platforms is designed to capture increasingly shorter fragments of attention. The second is compulsive self-optimization: smartwatches that quantify sleep, apps that monitor diet, productivity, and mood, and that produce subjects perpetually falling short of a measured ideal self. The third is technological dependence: the progressive outsourcing of cognitive abilities (memory, spatial orientation, writing) to opaque systems that we do not control and that we find increasingly difficult to do without.

The critical question is whether this constitutes domination. If, following Han, the subject believes themselves to be free while self-exploiting, what kind of freedom remains? Here, sociological theory recovers a Weberian intuition: substantive freedom requires room for deliberation about ends, not merely for choosing between technically predetermined means. A freedom reduced to clicking between algorithmically curated options is not the same as the freedom to collectively define which options deserve to be offered.

5. Power and Domination: State, Corporation, Citizenship?

Weberian sociology of power focused on the state as the bearer of the legitimate monopoly on violence and on its administrative arm (the bureaucracy) as the primary instrument of legal-rational domination. Corporations certainly appeared in his analysis but were subordinated to the state matrix and the logic of industrial capitalism. The contemporary situation demands a reformulation of this framework. The world's most powerful AI systems (in terms of computing power, data, and talent) are concentrated in a handful of transnational corporations (Alphabet, Microsoft, Meta, Amazon, Apple, Anthropic, OpenAI, ByteDance, Tencent, Alibaba) and, to some extent, in associated state apparatuses (the United States and China). This concentration constitutes a global redistribution of power whose implications are only now beginning to be theorized.

Crawford (2021), in his *AI Atlas*, has mapped the material dimension of this power: AI models require massive quantities of rare minerals, energy, water for cooling, fiber optic infrastructure, and undervalued human labor (data taggers in Kenya, the Philippines, or Venezuela). The supposed immateriality of the cloud conceals a concrete geography of extraction and exploitation. Far from surpassing industrial capitalism, AI prolongs it into an informational capitalism that multiplies its material dependencies. Srnicek (2017), Mejias, and Couldry (2024) have shown that this concentration is not a market accident: it is the logical consequence of network dynamics, increasing returns, and the capture of positive externalities from data.

The consequence is an unprecedented epistemic asymmetry. Zuboff (2022) calls it epistemic inequity: the platforms know far more about us than we know about them. This asymmetry is not only informational; it is operational. The corporations that control AI models can define what information circulates, what prices are offered to whom, what news is prioritized, and which political candidates receive the most exposure. Although they formally respect national legal frameworks, in practice they establish quasi-sovereign power over global information spaces. This is what Mathias Risse and others, in the context of Harvard's Carr Center, have called the "deadly encounter of institutional orders" between democracy and surveillance capitalism.

Faced with this, states oscillate between three positions. The first is regulation: the European Union has been a pioneer with the General Data Protection Regulation (GDPR, 2018), the Digital Services Act (2024), and, especially, the AI Act (2024), which classifies AI systems by risk level and prohibits applications deemed unacceptable (such as state-run social scoring). The second is strategic partnership: the United States has established a complex public-private relationship in which large technology companies operate as geopolitical allies against China, receiving subsidies and public contracts in exchange for guiding AI development toward shared objectives. The third is organic integration: China has developed a model in which platforms operate within a strict political framework and in which AI is a central part of the state governance apparatus, with initiatives such as the Social Credit System (Liang, 2018).

None of the three options easily fits the classic Weberian model of state bureaucracy. European regulation presupposes a technical capacity that regulators rarely possess when faced with corporations that command armies of engineers. American integration produces a gray area where the boundaries between public and corporate interests become blurred. The Chinese model revives state centrality but exercises it through technical infrastructures that reduce the margins of civil autonomy to levels that Weber would likely have considered totalitarian.

Meanwhile, citizens remain in a structurally disadvantaged position. They not only lack the information necessary to evaluate what systems do with their data; they also lack, except in specific cases, the collective channels to challenge them. Movements for algorithmic justice, platform workers' unions

(such as the App Drivers & Couriers Union in the UK), digital rights organizations (EFF, AccessNow, Derechos Digitales in Latin America), and platform cooperatives represent important attempts at organization, but they operate with an overwhelming asymmetry of resources compared to corporations.

6. Margins of Autonomy in the Algorithmic Era

While the metaphor of the algorithmic cage describes a structural tendency, it does not, however, describe a predetermined outcome. Weber himself, despite his pessimistic tone, left open the possibility of charismatic ruptures or collective reorientations. The contemporary question is what concrete forms these margins of autonomy can take in the face of algorithmic rationalization, and by what means they can be expanded.

At least five analytical approaches can be distinguished. The first is democratic regulation. The European AI Act (Regulation (EU) 2024/1689) constitutes the first systematic attempt to establish legally binding obligations for the development and deployment of AI. It prohibits applications considered to pose an unacceptable risk (state social scoring, cognitive manipulation, real-time mass biometric identification, with some exceptions), requires impact assessments for high-risk systems, and mandates minimum levels of transparency for general models. Its limitations are evident (difficulty of oversight, asymmetry with less stringent jurisdictions, risk of regulatory capture) but its existence demonstrates that algorithmic rationalization is not static: it can be politicized.

The second path is ethical design, expressed in frameworks such as explainable AI (XAI), fairness, accountability, and transparency (FAccT), algorithmic audits, and frameworks like the Asilomar AI Principles or UNESCO recommendations (2021). These efforts face, as we have already noted, a Weberian paradox: combating formal rationalization with more formal rationalization. However, they are not useless: they produce enforceable normative criteria, open professional spaces for internal critique, and build vocabularies that civil society can reclaim.

The third way is transparency and accountability. Movements like the Algorithmic Accountability Now Coalition, research by organizations such as AlgorithmWatch, the AI Now Institute, and the Ada Lovelace Institute, and specialized data journalism (ProPublica, The Markup) have systematically documented flaws, biases, and abuses of algorithmic systems. Transparency has its limits: many models are inherently opaque due to their architecture (deep learning), not just because of trade secrets. But the pressure for explainability has made the question "Who decided this, and how can I challenge it?" legitimate again in sectors such as criminal justice and social welfare.

The fourth path is critical digital literacy. Without citizens capable of understanding (at least in general terms) how algorithmic systems operate, democratic deliberation about them is impossible. This goes beyond mere technical training: it involves developing the ability to critically read recommendations, rankings, and predictions, and to recognize that behind every metric lies a value judgment. UNESCO, the OECD, and many civil society organizations have made progress in this area, although global gaps in digital literacy remain enormous.

The fifth path, perhaps the most promising, is that of cooperative alternatives and digital commons. Cooperative platforms like Up & Go (domestic cleaning in New York), Stocksy (photography), Eva (a transportation cooperative in Quebec), or open-source software and open AI models like Hugging Face, BLOOM, and Mozilla's initiatives demonstrate that it is possible to build digital infrastructures with a different logic of ownership and governance. Data commons, data trusts, and regional digital sovereignty initiatives (especially in Latin America, Africa, and India) pave the way for collectively reclaiming the infrastructures of rationalization.

None of these paths, on its own, is enough to “break the cage”. Taken together, however, they configure a horizon of politicizing technology: the idea that algorithmic architecture is a human artifact, modifiable, contestable, and democratically redesignable. In this, Langdon Winner's lesson is crucial: artifacts have politics (Winner, 1980). Technical decisions are always, at some level, political decisions, and recognizing this is the condition for any effective resistance.

7. Historical Comparison: Bureaucratic Cage and Algorithmic Cage

Having outlined the major features of both configurations, it is worthwhile comparing them analytically. There are clear continuities and decisive differences. Let us begin with the continuities. Both the bureaucratic cage of the 20th century and the algorithmic cage of the 21st are historical expressions of the same process: the expansion of formal rationality into ever-increasing domains of social life. Both are based on the conviction that calculation, explicit rules, and predictability are superior (technically and normatively) to the arbitrariness, favoritism, and unpredictability of previous systems. Both treat people as classifiable cases, even though the classification tools may vary. And both produce, as a structural counterpart, specific forms of alienation: that of the office worker who applies rules they do not understand; that of the user subjected to recommendations whose logic they cannot examine.

The differences, however, are structural. At least seven can be listed. First, speed: Weberian bureaucracy operated over prolonged periods (weeks, months, years); algorithmic systems operate in milliseconds. Second, granularity: bureaucracy classified by relatively stable categories (marital status, age, profession); algorithmic systems produce individual profiles updated in real time. Third, visibility: the official was recognizable and the rule, in principle, accessible; the algorithm is typically invisible and the rule opaque, even to its designers in the case of deep learning models. Fourth, the bearer of power: the nation-state no longer holds a monopoly; transnational corporations emerge as co-sovereign agents in informational spaces.

Fifth, the disciplinary modality: bureaucracy operated through coercion and procedure; the algorithmic cage operates through seduction, gamification, micro-incentives, and variable rewards (closer, at this point, to behavioral psychology than to legal theory). Sixth, the raw material: bureaucracy processed documents; algorithms process data on virtually every dimension of existence (movements, heart rate, searches, tone of voice, facial expressions). Seventh, the relationship with knowledge: bureaucracy presupposed professional expert knowledge, transmissible through formal training; AI produces statistical-correlational knowledge whose transmissibility is unequal and whose ownership is concentrated.

Taken together, these differences suggest that the algorithmic cage is not simply the same cage with new bars. It is a qualitative mutation of modern rationalization, in which control becomes more capillary, faster, and more affectively seductive, but also more opaque and more concentrated. If Weber thought the bureaucratic cage was as hard as steel, we could say that the algorithmic cage is more like a flexible polymer: it adapts, envelops, and modulates without the subject clearly perceiving its limits.

8. Expanded Theoretical Debate

Foucault's work is essential for understanding the algorithmic cage. In *Discipline and Punish* (2002), Foucault showed how Bentham's panopticon operated as a structure of asymmetrical visibility that produced disciplined subjects without the need for explicit violence. Few know about everyone; everyone knows they can be seen. This panoptic logic finds a continuation (but also a mutation) in the algorithmic regime. As Han (2021) has pointed out, the digital panopticon no longer relies on a centralized gaze; subjects themselves voluntarily produce the information that makes them visible. And

unlike the panopticon, which produced internalization through uncertainty about who is being looked at, the algorithmic regime produces conformity through the promise of gratifying personalization.

Moreover, later Foucauldian concepts (biopolitics, governmentality, subjectivation) illuminate dimensions that Weber could not theorize. Algorithmic governance is biopolitics in the Foucauldian sense: it operates on populations, modulating life probabilities (of illness, crime, professional success). It is governmental in that it operates through the freedom of subjects, not against it. And it produces specific subjectivations (the quantified self, the self-entrepreneur, the optimized user) who internalize the system's evaluation criteria (Rouvroy, 2016).

In *Theory of Communicative Action*, as notes Susen (2018), Jürgen Habermas developed a critique of rationalization distinct from Weber's: he distinguished between instrumental rationality, characteristic of systems (market and administration), and communicative rationality, characteristic of the lifeworld (*Lebenswelt*). For him, the critical diagnosis of modernity consisted of the “colonization of the lifeworld” by systemic logics: money and administrative power invaded spheres (family, education, public space) that should be governed by argumentative deliberation.

AI and digital platforms can be interpreted as a new phase of this colonization. Social media promised to expand the public sphere but instead fragmented it into algorithmically modeled bubbles designed to capture attention, not to foster rational deliberation. Public discourse is now mediated by systems optimized to produce outrage, polarization, and virality (economically profitable indicators, but deliberately disastrous ones). Habermas has recently updated his diagnosis, arguing that digitization has undermined the structural conditions for the discursive formation of the will. The algorithmic cage, interpreted through a Habermasian lens, is not merely an administrative intensification: it is a corrosion of the preconditions for deliberative democracy.

Zuboff's critique (2019; 2022) has already been mentioned. It is worth highlighting her specific contribution to the Weberian debate. Zuboff explicitly recalls that industrial capitalism “disfigured nature” and argues that surveillance capitalism disfigures human nature. Her thesis is that we are facing a new institutional order (not merely a variation of digital capitalism) characterized by the massive and secretive extraction of behavioral data and the generation of predictive products sold in behavioral futures markets. What distinguishes her from Weber is that her analysis emphasizes not the state bureaucracy but a new form of corporate power, which she calls instrumental power and distinguishes from both sovereign power and totalitarianism. Instrumentalism does not seek to conquer the soul (as totalitarianism does) but rather to dispense with it: it is only interested in observable and modifiable behavior.

Han radicalizes the diagnosis. For him, we live not in a disciplinary society but in a performance society; not under external coercion but under self-exploitation; not in a Foucauldian panopticon but in a digital one of voluntary transparency. Neoliberal freedom is the most effective form of coercion because it makes each subject their own exploiter. The smartphone, the “like,” and the algorithmic curation of one's own identity on social networks are the concrete tools of this subtle subjugation. His critique has been criticized as excessively totalizing (Quiroga, 2020), but it offers an essential phenomenological correlate to Zuboff's structural analysis.

Nick Srnicek in *Platform Capitalism* (2017) offers an economic and political perspective. Platforms are the dominant organizational form of 21st-century capitalism because they solve a fundamental economic problem: how to extract value in an economy where information is the primary input. He distinguishes five types (advertising, cloud, industrial, product, and *lean platforms*) and shows that their power derives from network effects and the increasing returns to data. His crucial contribution is to debunk the notion that digital monopoly is an accident: it is the logical consequence of the sector's technical and economic structure. This is important for our argument because it shows that the

concentration of algorithmic power cannot be corrected simply through more regulation (it may require structural interventions in infrastructure ownership).

Finally, Langdon Winner, in his classic essay *Do Artifacts Have Politics?* (1980), formulated a thesis that remains foundational for critical AI studies: technologies are not ethically neutral; they embody political choices in their very design. His example of the low bridges on Long Island, designed to prevent buses from passing and thus restrict access to the beaches for racialized populations, anticipates what we now call algorithmic bias. But the lesson runs deeper: any complex technical system organizes social relations and distributions of power. AI is no exception. Recognizing this shifts the debate from the moral question *What should we do with AI?* to the political question *Who decides how the systems that govern us are designed and deployed?*

9. Final thesis: Does AI Break the Cage or Perfect It?

At this point, it is possible to formulate a reasoned answer to the initial question. Artificial intelligence does not break the iron cage: it refines it and transforms it into a more sophisticated algorithmic cage, more invisible and more all-encompassing in some dimensions, but also, paradoxically, more fragile and contestable in others.

It perfects it because it fulfills, to an extent Weber could not have imagined, the promises of formal rationality: calculability, predictability, efficiency, and expansion into all domains of life. Any process that can be modeled as an objective function is, in principle, a candidate for algorithmic optimization. It sophisticates it by adding layers that bureaucracy lacked: real-time personalization, affective modulation, structural opacity, and the capture of desire. It makes it more invisible because its rules are not published in an official bulletin: they are inferred from the data and are continuously updated. It makes it more all-encompassing because it reaches dimensions (attention, desire, intimacy) that the bureaucratic panopticon barely touched.

And yet, it also weakens it. The Weberian bureaucratic cage, anchored in the nation-state and stable professional bodies, enjoyed a legitimacy built over decades of institutionalization. The algorithmic cage is sustained by corporations that depend on public trust, shifting legal frameworks, and technical models whose errors are becoming increasingly visible. The scandals of bias in COMPAS, Cambridge Analytica, automated layoffs, and the hallucinations of generative models have opened significant cracks in the legitimacy of the algorithmic regime. This crisis of confidence can be a lever for politicization.

Therefore, the most rigorous answer is neither technophile nor technophobic. AI does not automatically liberate (its current deployment expands unprecedented forms of domination), but neither does it irremediably condemn. The politically relevant question is under what conditions could it operate as a tool for augmenting human capabilities and not as a device for capture. These conditions, this analysis suggests, include at least: democratic regulatory frameworks that limit the most invasive uses; cooperative alternatives that diversify infrastructure ownership; critical public literacy that allows for deliberation on ends and not just means; and ongoing philosophical reflection on which dimensions of human experience should not be optimized, even if they can be.

The final metaphor could be this: if Weber's iron cage were a visible structure that modernism insisted on inhabiting, the algorithmic cage is a climate-controlled system in which it is pleasant to remain. The question is not only how to escape, but how to recover the capacity to ask that question. Keeping alive the possibility of thinking outside the regime of optimization is, perhaps, the most urgent political task of our time, and the most vital legacy that Weber's diagnosis can bequeath to the 21st century.

Bibliography

- Aradau, C., & Blanke, T. (2022). *Algorithmic reason: The new government of self and others*. Oxford University Press.
- Aristotle. (1995). *Physics*. Madrid: Gredos.
- Crary, J. (2013). *24/7: Late capitalism and the ends of sleep*. Verso.
- Crawford, K. (2021). *Atlas of AI: Power, politics, and the planetary costs of artificial intelligence*. Yale University Press.
- Creemers, R. (2018). China's social credit system: An evolving practice of control. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.3175792> .
- Curran, D. (2023). Surveillance capitalism and systemic digital risk: The imperative to collect and connect and the risks of interconnectedness. *Big Data & Society*, *10(1)* , 1–12. <https://doi.org/10.1177/20539517231177621>.
- Espeland, W. N. (2008). A sociology of quantification. *European Journal of Sociology*, *49(3)* , 401–436.
- Eubanks, V. (2018). *Automating inequality: How high-tech tools profile, police, and punish the poor*. St. Martin's Press.
- Foucault, M. (1977). Il faut défendre la société. *Cours au Collège de France*. Paris: Gallimard/Seuil.
- Foucault, M. (2002). *Discipline and Punish: The Birth of the Prison*. Buenos Aires: Siglo XXI Editores.
- Foucault, M. (2010). *The courage of truth: the government of self and others II*. Mexico: Fondo de Cultura Económica.
- Habermas, J. (1987). *Theory of Communicative Action (Vols. 1–2)*. Taurus.
- Han, B.-C. (2014). *Psychopolitics. Neoliberalism and new techniques of power*. Barcelona: Herder.
- Liang, F. D. (2018). Constructing a data-driven society: China's social credit system as a state surveillance infrastructure. *Policy & Internet*, *10(4)* , 415–453.
- Mau, S. (2019). *The metric society: On the quantification of the social*. Polity Press.
- Mejias, UA (2024). *Data grab: The new colonialism of Big Tech and how to fight back*. University of Chicago Press.
- Mökander, J., & Schroeder, R. (2024). Artificial intelligence, rationalization, and the limits of control in the public sector: The case of tax policy optimization. *Social Science Computer Review*, *42(6)* , 1–20. <https://doi.org/10.1177/08944393241235175>.
- Muellerleile, C., & Robertson, S. L. (2018). Digital Weberianism: Bureaucracy, information, and the techno-rationality of neoliberal capitalism. *Journal of Global Legal Studies*, *25(1)* , 187–216.
- European Parliament and Council of the European Union. (2024). *Regulation (EU) 2024/1689 on harmonised rules in the field of artificial intelligence (AI Act)*. Official Journal of the European Union.
- Quiroga, S. (2020). Psychopolitics, swarm, mediatization and digital communication in Byung-Chul Han: An initial critique. *Revista Investigación y Desarrollo (UNSL)*, *12* , 95–115.
- Rouvroy, A. (2016). Algorithmic governmentality and perspectives of emancipation: Disparity as a condition of individuation through relationship? *Philosophical Addendum*, *1* , 88–116.
- Schluchter, W. (2017). *The Disenchantment of the World: Six Studies on Max Weber*. Bogotá: Fondo de Cultura Económica.
- Schroeder, R. (2023). *Rationalization revisited: From Weber to digital capitalism*. Routledge.
- Srnicek, N. (2017). *Platform capitalism*. Polity Press.
- Susen, S. (2018). Jürgen Habermas: Between Democratic Deliberation and Deliberative Democracy. In R. & Wodak, *The Routledge Handbook of Language and Politics* (pp. 43-66). Abingdon: Routledge <https://openaccess.city.ac.uk/id/eprint/18941/>.
- UNESCO. (2021). *Recommendation on the ethics of artificial intelligence*. United Nations Educational, Scientific and Cultural Organization.
- Vogl, T. M. (2020). Smart technology and the emergence of algorithmic bureaucracy: Artificial intelligence in UK local authorities. *Public Administration Review*, *80(6)* , 946-961.
- Watson, D. S. (2023). In defense of sociotechnical pragmatism. *AI & Society*, *38(4)* , 1–14.
- Weber, M. (1949). *The Methodology of the Social Sciences*. Illinois: The Free Press of Glencoe.

- Weber, M. (1976). *Politics and Science*. Buenos Aires: Editorial La Pleyade.
- Weber, M. (2012). *Economy and society*. Mexico: Fondo de Cultura Económica.
- Weber, M. (2012). *The Protestant Ethic and the Spirit of Capitalism*. Mexico: Fondo de Cultura Económica.
- Weber, M. (2012). *Sociology of religion*. Madrid: Akal.
- Williams, J. (2018). *Stand out of our light: Freedom and resistance in the attention economy*. Cambridge University Press.
- Winner, L. (1980). Do artifacts have politics? *Daedalus*, 109(1) , 121–136.
- Yeung, K. (2018). Algorithmic regulation: A critical interrogation. *Regulation & Governance*, 12(4) , 505–523.
- Zuboff, S. (2019). *The age of surveillance capitalism: The fight for a human future at the new frontier of power*. PublicAffairs.
- Zuboff, S. (2022). Surveillance capitalism or democracy? The death match of institutional orders and the politics of knowledge in our information civilization. *Organization Theory*, 3(3) , 1–79. <https://doi.org/10.1177/26317877221129290>.

Copyright Notice

This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.