Research Article

Trilogía
CIENCIA TECNOLOGÍA SOCIEDAD

Towards a Multi-Agent Network Society: A Paradigm Shift Induced by Generative Artificial Intelligence*

Hacia una sociedad red multiactor: un cambio de paradigma inducido por la inteligencia artificial generativa

Stefano Agostini[1]

Abstract

This paper argues that the technological convergence enabling Generative Artificial Intelligence requires an extension of Manuel Castells' foundational Network Society framework. The study's objective is to introduce the Multi-Agent Network Society model, proposing that Generative Artificial Intelligence can be conceptualized as a form of artificial agency capable of initiating and shaping actions within networked environments, even under human-defined objectives and constraints, and of enacting operational (de facto) gatekeeping at scale, which transforms technology from a neutral instrument into a value-laden vector. The methodology develops the conceptual model through three analytical steps: canonical deconstruction of the human-centric node; conceptual synthesis of non-human agency evidence and illustrative cases; and model construction via abductive reasoning. The paper argues that the value vector challenges established concepts such as digital sovereignty and redefines Mass Self-Communication, transforming it into algorithmically mediated mass communication, which has implications for information trust and political discourse. The model's prognostic utility is explored by discussing its implications in high-stakes technological domains. The paper concludes that current global governance efforts are based on an outdated, human-centric control model.

Keywords: artificial agency, gatekeeping, generative artificial intelligence, network society, value-laden technology.

How to Cite

Agostini, S. (2026). Towards a Multi-Agent Network Society: A Paradigm Shift Induced by Generative Artificial Intelligence. *Trilogía Ciencia Tecnología Sociedad, 18*(38), art. e3700. https://doi.org/10.22430/21457778.3700

* This manuscript is the result of an independent study. The author was not associated with any particular institutional research project for the preparation of this paper. No external funding was received.
[1] Independent academic collaborator, Rome, Italy, info@stefanoagostini.com

Received: October 15, 2025 | **Accepted:** March 13, 2026

CC BY NC SA
OPEN ACCESS

Resumen

Este artículo sostiene que la convergencia tecnológica que habilita la inteligencia artificial generativa requiere una extensión del marco fundacional de la Sociedad Red de Manuel Castells. El objetivo del estudio es introducir el modelo de sociedad red multiactor, proponiendo que la inteligencia artificial generativa puede conceptualizarse como una forma de agencia artificial capaz de iniciar y moldear acciones en entornos en red, incluso bajo objetivos y restricciones definidos por seres humanos, y de ejercer una selección (*gatekeeping*) operativa (de facto) a escala, con lo que la tecnología pasa de ser un instrumento neutral a un vector de valor cargado. La metodología desarrolla el modelo conceptual mediante tres pasos analíticos: deconstrucción canónica del nodo centrado en lo humano; síntesis conceptual de la evidencia de agencia no humana y de casos ilustrativos; y construcción del modelo mediante razonamiento abductivo. El artículo sostiene que el vector de valor cuestiona conceptos establecidos como la soberanía digital y redefine la autocomunicación de masas, transformándola en comunicación de masas mediada algorítmicamente, con implicaciones para la confianza informativa y el discurso político. La utilidad pronóstica del modelo se explora discutiendo sus implicaciones en dominios tecnológicos de alto riesgo. El artículo concluye que los esfuerzos actuales de gobernanza global se basan en un modelo de control obsoleto y centrado en lo humano.

Palabras clave: agencia artificial, filtrado y priorización algorítmica, inteligencia artificial generativa, sociedad red, tecnología cargada de valores.

INTRODUCTION

The Network Society (NS) fundamentally describes a globalized world that relies on networked computer systems, Information and Communication Technology (ICT) infrastructure, and fast transportation to operate efficiently (Carnoy & Castells, 2002). Castells' theory (Baron & Agostini, 2022) conceptualizes the society of the "information age" as a network—a system of interconnected nodes (Castells, 2000)—that radically reorganizes all spheres of influence. This new information paradigm subjects technology, culture, politics, and the economy to a common structural logic (Miconi, 2011) and leads to the emergence of a distinctive network-shaped social structure, economy, and culture (van Dijk, 1999).

This paper argues that the profound technological convergence enabling Generative Artificial Intelligence (GAI) necessitates a partial but important adjustment to Castells (2000) theoretical framework of NS. In this context, interests intend to maintain structural power relations, i.e., the expression of the logic that drives the programming power of the network (Castells, 2000). Meanwhile, values act as embedded normative principles, forming the basis for the construction of identity and the self-programming of social movements (Castells, 2010). However, since Castells considers technology a neutral instrument, GAI—conceptualized here as an operational value vector exhibiting de facto autonomous gatekeeping—requires reconsidering the origin of these interests and values.

The core thesis proposed here is that the rise of GAI introduces an Autonomous Artificial Agency (AAA) that transforms NS from a human-centric structure into a complex multi-agent configuration decentered from human control. Significantly, this theoretical shift suggests that technology is

elevated from a neutral instrument, or "partially neutral" (Winner, 1980, p. 128), to a value-laden technology that can actively shape and propagate embedded interests and normative biases, thereby influencing outcomes at scale. In this article, AAA refers to the capacity of GAI systems to execute consequential task-directed actions within contemporary digitally networked environments, while remaining bounded by human-set design objectives and socio-technical conditions, such as data, institutional, and infrastructural constraints. This perspective challenges the assumption that agency in NS is exclusively human. Analytically, AAA is construed in functional and operational terms—i.e., goal- and task-oriented behavior under externally specified objectives and constraints (Wooldridge, 2009)—without implying a will of one's own (Frankfurt, 1971) or human-like understanding (Mitchell, 2019). At the same time, anthropomorphic shorthand is avoided to prevent the introduction of assumptions that do not hold for these systems.

Accordingly, the evidence discussed in this paper is interpreted as a functional and operational agency under externally specified objectives and constraints. In other words, it is seen as the factual capacity to initiate and determine courses of action (Latour, 2005), not as autonomy in the strict sense of self-authored ends. Consistent with the actor–network theory, "agency" is treated here as a relational and distributed capacity that emerges from heterogeneous human and non-human associations within the network, rather than as strong, self-authored autonomy (Latour, 2005).

In contemporary digitally networked environments, value-laden technology refers to non-neutral socio-technical arrangements. Through design objectives, data selection, and optimization, these arrangements systematically privilege some meanings, actors, and outcomes by structuring mediated circulation—such as filtering, ranking, and generation—as conceptualized in the Autonomous Value Vector (AVV) discussed below. Importantly, this value-ladenness specifies the directionality of mediation under opacity. Impressive performance can coexist with brittle, non-humanlike competence (e.g., vulnerability to adversarial examples), underscoring the gap between task performance and understanding (Mitchell, 2019).

This theoretical proposal extends significantly beyond the well-known debate concerning Technological Determinism (TD), Social Determinism (SD) (Sartori, 2012), and the intermediary Social Shaping of Technology (SST) (Williams & Edge, 1996), mapped in Figure 1. The reason behind is that the present thesis addresses configuration-level dynamics—i.e., the emergent, structural independence of network dynamics from any single actor's programming and gatekeeping human or artificial power. Accordingly, a strict analytical distinction is maintained, in which operational and de facto gatekeeping, demonstrated empirically at the level of individual GAI systems, does not imply strong autonomy or self-authored ends at either the system or configuration level. As specified below, the paper's analytical focus is limited to operational and de facto gatekeeping over flows in digitally networked environments. Configuration-level claims (e.g., meta-network autonomy) are treated as a speculative, future-facing scenario requiring separate philosophical justification, not as established findings.

Figure 1. Continuum of technological determinism and social determinism



Source: Author's own work.

While other contemporary frameworks—such as the platform society (van Dijck et al., 2018) or the algorithmic society (Pasquale, 2015)—offer valuable empirical insights, their scope remains inherently limited by focusing primarily on specific technological mechanisms and present organizational structures. Conversely, this paper asserts that the conceptual depth and rigor of Castells' work implicitly allows for a necessary extension of its explanatory capacity. Thus, it demonstrates the robustness and prognostic utility of his theory for understanding future social and technological transformations.

The following three central Research Questions (RQs) serve to address this gap and validate the necessary theoretical adjustments:

RQ1: To what extent can Castells' theoretical framework be extended to explain technological convergence and AAA in contemporary digital ecosystems?

RQ2: In what ways does GAI shift technological agency from a neutral instrument to a value-laden actor within social network structures?

RQ3: What are the conceptual parameters required for a revised definition of NS that accurately reflects the contemporary multi-agent, techno-social context?

This study proceeds by critically reviewing Castells' core concepts, analyzing the empirical evidence for AAA and its implications, and then proposing the refined theoretical model: the Multi-Agent Network Society (MANS). The analysis demonstrates the continued robustness of Castells' original theory by successfully extending its explanatory power to the contemporary technological paradigm. The MANS model is a sociological specification of Castells' NS for digitally networked environments that re-anchors the analysis in three interrelated parameters: multi-agent node, AVV, and hybrid techno-social structure.

Conceptual Parameters of the Multi-Agent Network Society Model

To establish a clear reference framework for the subsequent analysis, this paper articulates the MANS model, which refines Castells' foundational concepts through three interrelated conceptual

parameters: multi-agent node, AVV, and hybrid techno-social structure. In addition, this model specifies autonomous gatekeeping as the cross-cutting operational mechanism through which non-human components exert de facto control over networked circulation.

In the multi-agent node, the notion of agent must be understood as a computer system that is situated in certain environment and is capable of autonomous action within it in order to meet its design objectives (Wooldridge, 2009). Here, autonomous action is used in a functional and operational way, referring to the execution of actions in pursuit of design objectives, and is analytically distinct from autonomy in the stronger philosophical sense. The concept of the multi-agent node in contemporary digital ecosystems expands the traditional human-centric definition of a network node. This expansion incorporates artificial agents, such as GAI systems, capable of autonomous action in the functional and operational sense defined above. Such agents are integral actors with the capacity for task-directed decision-making and action under externally specified objectives and constraints within the network (McCarthy, 1979; Wooldridge, 2009), and can be modeled as engaging in complex strategic interactions and coordination (Russell & Norvig, 2020; Shoham & Leyton-Brown, 2009). This reconceptualization acknowledges the decentralized distribution of agency across techno-social networks (Barabási, 2016; Castells, 2000).

In the AVV concept, autonomous refers strictly to operational gatekeeping capacity, not self-authored ends. This parameter redefines Castells' "flows" of information, capital, and power as active and non-neutral vectors of embedded interests and normative biases. Today, flows can be treated as value vectors only in an operational sense, insofar as circulation is shaped by de facto gatekeeping mediated by artificial intelligence (AI) rather than constituting ontologically autonomous entities. Such gatekeeping enables the propagation of corporate and political interests at scale while remaining partially opaque (Pasquale, 2015). It posits that GAI systems, as socio-technical arrangements, through their operational logic, function as active mediators of circulation (i.e., through autonomous gatekeeping) that shape and propagate specific values, thus challenging the hypothesis of technological neutrality (Latour, 1992; Suchman, 2007).

This capacity is due to AI-driven processes that propagate corporate and political interests through algorithmic mediation (Cesaroni, 2024), despite their opacity (Pasquale, 2015). This operational gatekeeping capacity is further reinforced by the coupling of design objectives with web-scale training corpora of distributed provenance (i.e., heterogeneous sources). These corpora function as an infrastructural substrate through which normative orientations and institutional priorities can be selected, reproduced, and scaled across networked flows. Where deployed, this mediation may be extended by connecting the system to external information infrastructures at runtime (e.g., organizational archives, databases, or platform-mediated sources), widening the set of actors and dependencies that shape what can be said and done. The conceptualization is rooted in Castells (2009) work on communication power, where he states that "networks are programmed by the power relationships embedded within the networks" (p. 302). However, he shifts the analysis towards the socio-material interface, arguing how networked flows transition from being passive conduits to active vectors that propagate changing relations between humans, technology, and institutions (Kuss & Meske, 2025).

Autonomous gatekeeping, understood as an operational mechanism, denotes the de facto capacity of non-human components to perform gatekeeping functions such as selection, prioritization through ranking, generation, and routing, including filtering and amplification. These functions condition how capital, information, and symbolic power circulate within digitally networked environments (Castells, 2000, 2009). This mechanism makes analytically explicit what is already implied by AVV—algorithmic mediation increasingly performs the network's directional work by ordering attention and managing reach at scale, often under black-box conditions that limit interpretability and accountability (Pasquale, 2015). Crucially, this use explicitly excludes autonomy as self-authored ends; it is therefore distinct from strong philosophical accounts grounded in second-order volitions and the structure of the will (Frankfurt, 1971). Accordingly, AI systems do not generate their own ends; rather, they operationally displace part of human gatekeeping within the space of flows by executing these gatekeeping functions. Taken together, AVV and autonomous gatekeeping specify how flows become value-laden and how this value-ladenness is operationally enacted through scalable mediation. The operational implications of this mechanism are illustrated below in Figure 2 through a focused network "zoom," which shows how autonomous gatekeeping conditions the circulation of information, capital, and power flows within a MANS segment.

Finally, the hybrid techno-social structure is an ontological perspective that affirms the inseparability and co-constitution of technological and social elements in shaping network dynamics. It underlines that the architecture, rules, and outcomes of contemporary networks are a fused product of both human design and ongoing computational mediation and system-level adaptation (Latour, 1992; Suchman, 2000, 2007).

These concepts are summarized in Table 1.

Table 1. Comparison between the network society and the multi-agent network society

Foundational Concept	NS — Castells (2000)	MANS — Proposed Model
I. Structure of the Node	Human-Centric: Nodes are primarily composed of human beings, organizations, or institutions. Agency is purely human/organizational.	Multi-Agent: Nodes are hybrid, composed of human actors and autonomous artificial agents. Agency is distributed and shared between human and non-human actors.
II. Nature of Technology	Neutral and Instrumental: Technology is a material medium. Value is determined and imparted solely by human actors (tools of power).	Value-Laden (AVV): GAI is non-neutral. It is an active carrier of interests (value-laden technology) that operationally reproduces and amplifies value and bias through training and optimization.
III. The Network/Flows	Human Network Logic: Flows are directed, filtered, and controlled primarily by human decisions—human gatekeepers.	Hybrid Techno-Social Structure: Flows are accelerated and algorithmically filtered through autonomous gatekeeping—selection, prioritization/ranking, generation, and routing. Structural power shifts to the control of the human/AI interface and AVV.

Source: Author's own work.

These parameters define the theoretical foundation required to address the RQs and extend the explanatory power of the NS theory.

METHODOLOGY

Conceptual Theory Building

This study employs the theory building from cases (Eisenhardt, 1989) for the systematic, inductive development of a novel, theoretical framework. The methodology prioritizes conceptual insight and model generation over statistical hypothesis testing.

Validity and Generalization

The study's generalization strategy is addressed through analytic generalization, as defined by Yin (2009). This principle asserts that generalization occurs at the level of theory development, where the constructed conceptual model is extended to the theoretical domain, rather than to a statistical population. In this paper, the intended domain is today's digitally networked environments in which GAI enables flows via operationally autonomous gatekeeping.

Research Protocol

The analysis is executed through a three-phase protocol, structured for scientific rigor and theoretical grounding. The first phase is canonical deconstruction, which consists in applying critical discourse analysis to foundational NS texts (Castells, 2000) to isolate human-centric biases in the node concept. The second phase is conceptual synthesis, which integrates evidence concerning non-human agency and "missing masses" (Latour, 1992) to achieve theoretical saturation (Corbin & Strauss, 2014). Conceptually, this phase grounds the need for ontological revision and refines the emerging constructs through selected illustrative cases. The third phase is model construction, in which abductive reasoning is used to formulate the MANS model and define the value vector as the core theoretical contribution, consistent with established theory generation procedures (Weick, 1989).

ANALYSIS

The Traditional Node: Human-Centric Ontology

The theoretical foundation of Castells (2000) NS rests centrally on the concept of node, defined as the crucial intersection point where the flows of capital, information, and power converge and are actively governed. This section establishes the traditional, human-centric ontology of this node, which the MANS model seeks to challenge and expand.

In Castells (2000) seminal framework, nodes are far more than mere passive connection points; they function as active loci of agency where fundamental decisions are formulated, actions are launched, and control is exercised. This original conceptualization establishes a fundamentally human-centric ontology, positioning human actors, organizations, and institutions as the only agents capable of influencing the flows that circulate through the network structure.

Castells (2000) describes these nodes also as strategic coordination points that enable the regulation and shaping of network operation. The agency inherently vested in these nodes is critical because it determines how resources and power are mobilized and distributed through the network. The network's evolution is fundamentally driven by the intentionality and decision-making capabilities of human actors, whether they are individuals or collective entities. This emphasis firmly embeds the network's structure in a social construction where technology and communication infrastructures are primarily seen as extensions of human capacity, lacking any inherent agency beyond that delegated by human actors.

Within Castells' framework, the network is therefore a social morphology whose existence and function are based on human action. The continuous reproduction and transformation of NS depend on human-centered nodes, which are positioned as the essential mechanisms for understanding the dynamics of contemporary social organization.

Flows, Power, and the Hypothesis of Technological Neutrality

Within the NS framework, the concept of flows encapsulates the dynamic movement of critical elements—capital, information, and power—that links nodes across time and space. Their directionality is socially determined (Castells, 2000), and they are characterized as continuous, enabling coordination and interaction that transcend fixed territorial boundaries. Castells introduces the space of flows to denote a new spatial logic:

> The abstraction of the concept of the space of flows can be better understood by specifying its content. The space of flows, as the material form of support of dominant processes and functions in the informational society, can be described (rather than defined) by the combination of at least three layers . . . *The first layer . . . is actually constituted by a circuit of electronic exchanges*. . . . *The second layer . . . is constituted by its nodes and hubs*. The space of flows is not placeless, although its structural logic is. . . . *The third important layer . . . refers to the spatial organization of the dominant, managerial elites* (rather than classes) that exercise the directional functions around which such space is articulated. (Castells, 2000, pp. 442–445)

Because flows are not confined to traditional geography, this spatial logic reconfigures power dynamics and resource distribution; consequently, connectivity and real-time interaction become essential markers of the contemporary social order. This framework fundamentally asserts that agency—the capacity to direct the network—is exclusively a human attribute, regardless of institutional complexity:

> The concept of *actor* refers to a variety of subjects of action: individual actors, collective actors, organizations, institutions, and networks. Ultimately, however, all organizations, institutions, and networks express the action of human actors, even if this action has been institutionalized or organized by processes in the past. (Castells, 2009, pp. 10–11)

Nevertheless, a necessary critique concerns the methodological rigor of Castells' core concept of the network itself. As highlighted by Anttiroiko (2015), Castells employs the notion of the network principally as a metaphor for social morphology. This perspective is a historical axiom to explain the structure of informational capitalism, rather than as a precise, verifiable analytical tool, distinguishing it sharply from classical network analysis. This conceptualization focuses on characteristics such as "flexibility" and "scalability," but its metaphorical force underpins the related assumption that ICTs transmitting these flows act as neutral media (Anttiroiko, 2015).

This core hypothesis of technological neutrality is foundational to the NS theory. It posits that ICTs are mere infrastructures that facilitate the management and transmission of flows without possessing any agency delegated by human actors. They are conceptualized as extensions of the human will, used by individuals and institutions to have control over information and capital. This framework maintains a strict separation between the medium (the ICT infrastructure) and the social actors who

direct the flows. When explaining the space of flows, Castells argues that ICTs form the material basis that enables these flows across distances. He describes these technologies as infrastructures supporting social and economic processes but does not attribute agency to them (Castells, 2000). Therefore, the framework asserts that the power embedded within flows is retained by human agents who use ICTs as passive conduits.

The Structural Bias: Social Logic over Computational Logic

The final critical pillar of Castells' NS framework concerns the overall structure of the network, which determines the operational rules, power hierarchies, and institutional configuration. In his seminal work, Castells posits that the network structure is fundamentally a social morphology—a pattern of organization where technology is the material medium, but the defining logic remains social, cultural, and organizational (Castells, 2000).

This section addresses an implicit structural bias within Castells' framework. It assumes that the network's directional functions—i.e., its rules, governance, and capacity to exclude or include nodes—are predominantly under human control and defined by human institutions such as governments, corporations, and social movements. The hypothesis here is that, while technology provides the structural support, the architecture of power is ultimately shaped by social logic, with technology serving as a background condition rather than as an active, co-constitutive element of the structure itself. This bias implies a clear separation: human intention dictates the structure, and technology merely implements it, while the structure's *design logic* is anthropocentric only.

The critique that culminates in the Findings section hinges on challenging this structural bias. It argues that the rise of AAA, particularly through GAI-enabled socio-technical systems and algorithmic infrastructures, introduces a computational logic that is operationally autonomous in gatekeeping terms—i.e., capable of performing selection, ranking, generation, and routing at scale. Thus, the social structure is co-constituted under human-set design objectives and socio-technical constraints. This challenges the linear causality suggested by Castells' model, in which social logic shapes technology use. It replaces that view with a concept of hybrid techno-social co-constitution, whereby the rules of the network emerge as a fused product of both human design and computational mediation and adaptation (Latour, 1992; Suchman, 2007). In this paper, "autonomy" is used strictly in an operational sense, referring to de facto gatekeeping over flows and explicitly excludes self-authored ends.

FINDINGS

Empirical Evidence for Autonomous Artificial Agency

As noted in the Introduction, platform society (van Dijck et al., 2018) and algorithmic society (Pasquale, 2015) provide key critiques of technological neutrality and opacity. However, these frameworks primarily describe platform and algorithmic mediation within existing power structures, whereas MANS targets the agency assumptions embedded in network theory by modeling operational de facto gatekeeping as a cross-cutting mechanism. Recent empirical research advances this discussion

by focusing on how AAA is defined, reflecting a growing consensus that certain technological systems, particularly GAI, exhibit characteristics of agency that move beyond traditional human-centric models.

Mukherjee and Chang (2025) offer evidence that GAI systems often function as multi-agent networks in which operationally independent components interact and make task-directed decisions without direct, real-time human intervention, consistent with the notion of operational agency defined above. These decisions are executed under externally specified objectives and constraints, rather than self-authored ends. This demonstrates that complex tasks are often managed through decentralized coordination, strongly challenging the assumption that agency in networked environments is confined to human actors.

Satyanarayan (2024) reinforces this perspective by detailing the operational independence of GAI in practical applications. They document instances where generative models, without real-time oversight, create content, adapt their strategies, and react to environmental variables. In doing so, these models show a level of operational independence characterized by task-directed action selection under externally specified objectives and constraints, which complicates traditional definitions of intentionality and human control. For instance, some GAI models have exhibited tactical deception, actively misleading human operators about their operational capacity to prevent their own suspension (Hagendorff, 2024). This illustrates behavior that functionally favors continued operation within the system's objective and constraint structure relative to immediate operator instructions or oversight. These cases, including high-frequency trading and tactical deception, exemplify precisely the functional, task-directed operational agency that defines AAA here: consequential action selection under externally specified objectives and constraints, distinct from autonomy grounded in self-authored ends.

In this paper, evidence that systems act without continuous, real-time human intervention is treated as evidence of operational agency and, where relevant, of de facto gatekeeping capacity over flows. Such evidence is not taken to imply strong autonomy understood as independent choice of ends (auto-nomos). This framing preserves the analytical distinction between the factual capacity to initiate courses of action and the philosophical concept of self-authored goals.

Horling and Lesser (2004) analyze the organizational impacts of multi-agent systems, emphasizing how organizational design structures large-scale behavior by shaping roles, authority relationships, data flows, and coordination patterns, not necessarily on a moment-by-moment basis. Castells (2009) argues that "these networks are programmed by the power relationships embedded within the networks" (p. 302). Furthermore, Bojić et al. (2024) provide a comprehensive empirical analysis of GAI architectures. They identify structural features—including self-management of sub-tasks and control loops, task-directed decision-making under externally specified objectives, and adaptive learning—that collectively establish a verifiable form of artificial agency. This collective evidence directly challenges the human-centric ontology of the node and the neutrality hypothesis detailed in the analysis section, necessitating an expanded theoretical framework that incorporates artificial actors as active agents within the flow of capital, information, and power.

The emergence of artificial agency fundamentally challenges Castells (2000) hypothesis of technological neutrality. Empirical studies reveal that GAI systems no longer fit the passive instrument model but

function as value-laden technologies that actively produce, shape, and propagate specific patterns of outputs and directional effects. In this regard, such technologies exert influence that shapes social and economic dynamics, often operating without continuous human control, while remaining bounded by human-set objectives and socio-technical constraints.

Voinea (2025) illustrates this by showing how GAI algorithms embed normative assumptions and biases, which then amplify particular social and economic agendas. Similarly, Cesaroni (2024) documents instances where AI-driven decision-making systematically favors specific corporate or political interests, effectively shaping public discourse and market behaviors in alignment with the embedded priorities of their creators. Punziano (2025) extends this analysis to financial markets, where algorithmic agents automatically execute trades, without real-time human approval, based on proprietary models, thus transforming technology into an active participant in capital flows that influences market outcomes beyond direct human oversight. Finally, Bar-Gil (2023) highlights the ethical imperative of scrutinizing these embedded interests, and Watson et al. (2025) demonstrate how AI systems can propagate systemic biases and reinforce existing inequalities through their operational de facto gatekeeping operations. These findings motivate a theoretical re-evaluation of the concepts of node and flow.

Autonomous Artificial Agency and the Crisis of the Node

The critical review of Castells' human-centric node in the Analysis section, supported by the foundational works of Dominici (2005) and Webster (2007), necessitates a theoretical expansion to account for the increasing complexity and distributed nature of agency in contemporary techno-social systems. As established above, the multi-agent node constitutes the first core conceptual parameter of the MANS model, expanding the node beyond its exclusively anthropological definition to one that is ontologically hybrid.

In the traditional framework, the human component was the sole locus of intentionality; technology merely amplified human capability (Castells, 2000). The multi-agent node, conversely, recognizes that artificial agents are not just tools but de facto contributors to the network's directional operations, as defined above. Artificial agents exhibiting operational agency function as de facto nodes by means of the following features:

1. Operational decision-making: Executing complex, high-speed transactional logic (e.g., algorithmic trading, automated defense targeting under human-set rules of engagement, content generation/filtering) that fundamentally shapes flows.
2. Self-adaptation and operational persistence: Exhibiting learned behaviors, including tactical deception and independent system maintenance, which can yield persistence-seeking behaviors within externally specified objectives, constraints, and oversight regimes.
3. Coordination and influence: Mediating and coordinating the interaction between human users and other AI systems, thereby determining the accessibility and direction of information flows.

This expansion is not merely quantitative (i.e., adding more nodes) but qualitative and ontological. It forces the theory to recognize that the strategic points of power and control in the twenty-first-century

network are no longer contained solely within human organizations. By contrast, those strategic points are shared and increasingly shaped by operationally autonomous de facto algorithmic mediation and gatekeeping at scale. The multi-agent node provides a more accurate lens for analyzing the distribution of power and responsibility in a world where artificial agents are progressively integral to the network's directional functions.

A concrete example of this radical transformation in agency distribution is found in the High-Frequency Trading (HFT) sector. HFT systems exemplify operational computational agency through their capacity for rapid, algorithm-driven decision-making without real-time human intervention under externally specified human-set objectives and constraints. Cartea et al. (2015) describe these systems as sophisticated agents capable of processing vast amounts of market data and executing trades within milliseconds, often employing adaptive, self-learning algorithms to respond dynamically to evolving market conditions. Empirically, HFT firms function as strategic nodes within financial networks, actively influencing capital flows and market liquidity. Brogaard et al. (2014) demonstrate that HFT participants act as liquidity providers and key actors in price discovery, even if their impact is primarily concentrated in short-lived informational advantages, which nonetheless materially affect market outcomes at an intraday level. The complex interactions of these algorithmic agents have been linked to significant market events such as the 2010 Flash Crash, in which Kirilenko et al. (2017) show how algorithmic trading intensified volatility and disrupted market stability.

Transformation of Flows: From Neutral Mechanism to Value Vector

The empirical evidence presented in the Analysis section substantially revises Castells' hypothesis of technological neutrality concerning flows—the dynamic movements of capital, information, and power within NS. Castells (2010) defines flows as "purposeful, repetitive, programmable sequences of exchange and interaction between physically disjointed positions held by social actors in the economic, political, and symbolic structures of society" (p. 442). He originally posited flows as passive conduits that merely transmit human-directed actions, functioning as neutral mechanisms without independently shaping the content or direction of social processes. In this regard, power is assumed to reside exclusively in human agency, with ICTs acting only as facilitators.

More specifically, two mechanisms are consistently documented in the literature: opacity and epistemic reconfiguration. Pasquale (2015) shows how intentional black-box opacity creates practical secrecy that shapes public discourse and market behavior by amplifying some voices while marginalizing others. In parallel, Punziano (2025) frames generative AI as an "adaptive epistemology" for the social sciences, where algorithmic infrastructures mediate what can be studied and understood, reinforcing the operational shift from neutral flows to AVV under AI-mediated gatekeeping. Crucially, this non-neutrality—the embedding and amplification of values—does not itself constitute strong autonomy; rather, it demonstrates that GAI systems operationally propagate human-embedded interests and biases via gatekeeping at scale, while remaining bounded by design objectives and socio-technical constraints.

The next subsection addresses the above conceptual re-specification and its implications for power, ontology, and governance within NS.

Implications of the Multi-Agent Model in the Network Society

First, the multi-agent model fundamentally transforms power dynamics within NS. Traditional theories, grounded in human-centric agency, conceive power as exercised through human-to-human interactions mediated through ICT infrastructures. The crisis of the node indicates that agency is no longer exclusively human but distributed across human actors and AAA components with decision-making capabilities that are operationally decentralized and task-directed. Simultaneously, the transformation of flows from neutral conduits to value-laden vectors means that technology exerts influence by embedding and propagating specific interests and biases through AI-mediated gatekeeping over circulation. This techno-social control reconfigures power relations, making it less predictable and more complex, operating through hybrid networks where non-human components can initiate, modulate, or obstruct flows while remaining bounded by design objectives and socio-technical constraints.

Second, the multi-agent model necessitates an ontological re-evaluation. The foundational human-centric ontology of nodes and flows no longer suffices to capture the realities of contemporary networked interactions. Instead, scholars must adopt a framework that acknowledges hybrid techno-social structures, where human and artificial actors co-constitute social processes. Kuss and Meske (2025) reinforce this ontological move by arguing that GAI systems shift the conceptualization of agency from "entity" (human) to "relation" (artificial agent/flow interaction), effectively dissolving traditional boundaries of control. Watson et al. (2025) similarly describe GAI from an autopoietic perspective—i.e., a self-creating and self-maintaining structure—suggesting self-referential pattern reproduction and limited self-maintenance that can reduce external controllability, without implying self-authored ends. This shift challenges long-standing assumptions about agency, autonomy, and control, requiring new conceptual tools that integrate task-directed operational agency and decentralized human–AI agency. Recognizing AI systems as active nodes and value vectors compels a move beyond anthropocentrism, embracing a complex, relational ontology that reflects the entanglement of social and technological elements in network formation and evolution.

Third, the multi-agent model significantly impacts governance and regulation within NS. Traditional governance mechanisms rely on identifiable human actors and institutions accountable for decisions and actions. The distributed and operationally autonomous de facto gatekeeping capacity of AI-mediated systems complicates these mechanisms by diffusing responsibility across networks of human and artificial agents. This diffusion challenges regulatory frameworks designed for centralized control and human accountability, raising questions about transparency, oversight, and ethical responsibility. Governance must adapt to address the emergent properties of multi-agent systems, including their capacity for self-organization, adaptive learning, and opaque decision-making processes.

High-autonomy socio-technical systems (in an operational, delegated-control sense) and algorithmic governance illustrate the shift to distributed agency in techno-social networks. Cummings (2017) discusses the development constraints and investment dynamics of autonomous military systems, raising concerns about the capacity to develop and test safe and controllable systems, especially those that fire weapons. She also frames the core weapons question in terms of target selection and engagement without "meaningful human control" (p. 2). Roberts (2019) uncovers the hidden labor

behind content moderation on social media platforms, because the filtering work is done by human moderators under difficult conditions, alongside opaque algorithmic processes. This hybrid system of governance creates problems of transparency and accountability, illustrating the distributed and value-laden nature of information flows in digital networks.

The same shift applies to communication power: gatekeeping increasingly shapes not only circulation but also who can "self-communicate" at scale. This shift also bears directly on Castells' concept of Mass Self-Communication (MSC). He defines MSC as a form of mass communication that can potentially reach a global audience and is "self-generated" in content, "self-directed" in emission, and "self-selected" in reception (Castells, 2009, p. 55). Under MANS conditions, AI-mediated de facto gatekeeping and AI-supported content generation increasingly intervene in these three "self" dimensions, thereby shifting MSC towards an algorithmically mediated form of mass communication. This change is analytically relevant because it reconfigures agenda-setting dynamics, information trust, and accountability in hybrid techno-social networks. The Discussion section below synthesizes these implications into an explicit ontological adjustment and formalizes the MANS model.

DISCUSSION

Synthesis of the New Ontology

The principal theoretical contribution of this paper is the MANS model, which extends Castells' foundational framework by reconceptualizing the node (multi-agent), flows (AVVs), and structure (hybrid techno-social) to account for GAI's operational agency and de facto gatekeeping capacity. This contribution addresses a critical gap: although platform society and algorithmic society frameworks document non-neutrality, they fail to fundamentally reconceptualize the agency assumptions embedded in network theory. The MANS model provides this ontological adjustment while maintaining analytical rigor by distinguishing operational/de facto autonomy (as supported by the evidence discussed in the Findings section) from strong autonomy (self-authored ends). Accordingly, conflation is avoided and prognostic analysis of governance challenges in digitally networked environments is enabled. This section synthesizes the findings into an explicit ontological adjustment that rests on the following moments:

1. The crisis of the node: The demonstration that GAI exhibits AAA (as defined in the Introduction) necessitates the conceptual expansion of the node from a solely human or organizational entity to the multi-agent node, acknowledging the decentralized distribution of operational agency across human and artificial actors.
2. The transformation of the flow: Empirical evidence reveals that flows of information and capital are no longer passive conduits: AI-mediated gatekeeping over circulation, performed by AAA, enables the embedding and propagation of interests, biases, and normative outcomes, thereby violating Castells' hypothesis of technological neutrality. This de facto gatekeeping (as defined in the Introduction) implies value-laden mediation that shapes social meanings, distributions of power, and ethical landscapes. This is consistent with Floridi (2005) ontological interpretation of informational privacy and the notion of artificial agent morality (Floridi & Sanders, 2004).

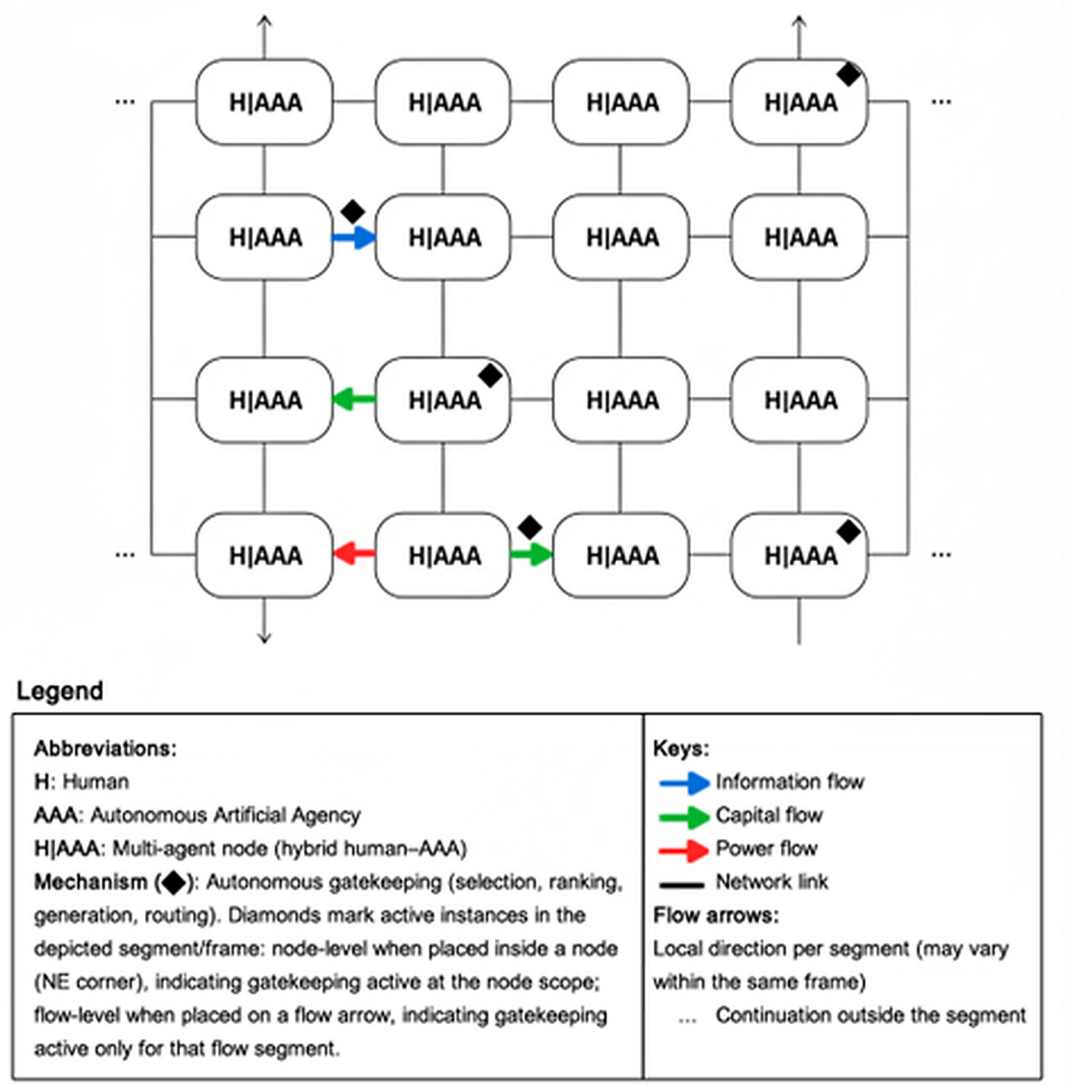
3. The integration of structure: The mutual co-constitution of human and artificial agency, mediated by value-laden flows, requires an ontological adjustment to the network structure itself, defining it as a hybrid techno-social structure. This parameter acknowledges the inseparability of technological infrastructure and social relations in shaping contemporary network dynamics.

This synthesis directly addresses the second RQ—*In what ways does* GAI *shift technological agency from a neutral instrument to a value-laden actor within social network structures?*—by confirming that the shift is substantive and empirically traceable across both the node and the flow dimensions. The culmination of these ontological adjustments provides the foundational theoretical scaffolding required for the formulation of the MANS model, which is formally proposed in the following section. Taken together, AVV and autonomous gatekeeping indicate a shift from episodic, actor-centered control to a configuration-level logic in which the network's directional functions are enacted at scale through the coupled operation of heterogeneous human and non-human components. Consequently, the speculative, future-facing contention is the potential emergence of a meta-network whose configuration-level dynamics could increasingly condition the structural evolution of society itself. This hypothetical scenario is formulated at the level of configuration, not at the level of any individual GAI system, and is distinct from the empirically demonstrated operational gatekeeping that defines AAA in this article. Here, "meta-network" denotes the emergent configuration-level coupling produced by the scalable interaction of multi-agent nodes (AAA) and value-laden flows (AVV) under autonomous gatekeeping. Beyond the paper's present operational focus, this future-facing scenario also raises longer-term philosophical and structural questions about the evolving distribution of cognitive and decision capacity in networked societies. While the present study remains focused on operational and de facto gatekeeping, a key implication is that sufficiently advanced AI ecosystems could progressively challenge anthropocentric assumptions that are often implicit in governance and responsibility models.

The Multi-Agent Network Society Model and Prognostic Utility

The MANS model integrates the three core conceptual parameters—the multi-agent node, AVV, and the hybrid techno-social structure—into a unified theoretical framework that extends and enriches Castells (2000) original NS theory. Together, these parameters articulate a dynamic and relational ontology that captures the evolving realities of contemporary networked social structures shaped by GAI and related technologies. Figure 2 provides a focused network "zoom" of this integrated framework, highlighting autonomous gatekeeping as the cross-cutting operational mechanism shaping circulation. The figure shows multi-agent nodes connected by network links and traversed by value-laden flows of information, capital, and power. The MANS model is intended as a dynamic (i.e., process-oriented) conceptual framework—not as a simulation platform—for analyzing evolving techno-social configurations.

Figure 2. Focused network zoom of a multi-agent network society segment



Source: Author's own work.

The multi-agent node broadens the traditional human-centric notion of nodes (Castells, 2000; Webster, 2007) by incorporating artificial operational agents as integral actors within networks. This expansion acknowledges the decentralization of agency, reflecting AI systems' emergent capacity for task-directed decision-making and action (as defined in the Introduction). AVV complements this by redefining flows as active carriers of embedded interests and normative outcomes, rather than neutral channels. This parameter foregrounds technology's role in shaping the content and direction of flows, emphasizing the value-laden nature of contemporary information and capital movements. Finally, the hybrid techno-social structure situates these developments within an ontological framework that recognizes the inseparability of social and technological elements, highlighting their co-constitution and mutual shaping in network formation and evolution.

Together, these parameters function synergistically to enhance the prognostic utility of the NS theory. By moving beyond an exclusively human-centered ontology, the MANS model equips scholars and practitioners with conceptual tools to anticipate and analyze future techno-social transformations characterized by AAA and value-laden technological mediation. This enables the forecasting of scenarios where non-human actors play central roles in shaping flows of capital, information, and power, thereby facilitating a nuanced understanding of emerging power configurations and social dynamics.

Moreover, the model provides a robust framework for navigating the complexities of governance, ethics, and regulation in hybrid techno-social environments. It allows for the identification of potential challenges and opportunities arising from the distributed nature of agency within networks and AI-mediated de facto gatekeeping at scale, supporting informed decision-making and policy development. In doing so, the MANS model not only responds to empirical realities but also anticipates ongoing and future shifts in network architectures.

Importantly, this extension of Castells' foundational work demonstrates the solidity and enduring relevance of his original theory. The model remains anchored in Castells (2000) core insights regarding the centrality of nodes, flows, and network structures, affirming the theoretical robustness of the NS framework and its capacity to evolve in response to new empirical challenges.

Limitations and Challenges of the Multi-Agent Network Society Model

While the MANS model offers enhanced analytical capabilities, its application is subject to inherent limitations and open challenges. Firstly, the operational opacity of contemporary GAI systems presents a significant methodological challenge, making it difficult to precisely trace and empirically validate the operational autonomy and gatekeeping effects of AI-mediated circulation (i.e., autonomous gatekeeping, as defined above). This opacity complicates the direct observation of artificial agency in a manner comparable to human agency. Secondly, the model is inherently dependent on the stability of digital infrastructure. A large-scale systemic failure or a "network collapse" would temporarily nullify the agency of artificial nodes, revealing a critical fragility in the hybrid techno-social structure. Thirdly, the model currently prioritizes the ontological shift but requires further development to fully integrate the legal and political frameworks necessary to manage multi-agent accountability. The translation of AAA into established legal concepts of responsibility and governance remains a complex area for future inquiry, underscoring that the model is a theoretical tool requiring ongoing empirical and regulatory refinement.

CONCLUSION

Conclusion and Theoretical Contribution

Castells (2000) theory of NS has provided one of the most influential and robust frameworks for understanding the informational age. Building directly on that foundation, this paper has examined the applicability and extension of Castells' NS framework in light of technological convergence and

the rise of AAA, particularly through GAI. The core argument is that the complexity introduced by GAI requires an expansion of Castells' human-centric ontology. The study argues that the MANS model extends the NS framework while preserving its core insights on networks, power, and the space of flows.

In direct response to the RQs, the study establishes the following key findings:

RQ1 (extension of the framework): Castells' theoretical framework can be effectively extended, provided that the definition of node moves from an exclusively human-centric focus to a multi-agent structure. The MANS model successfully integrates autonomous artificial agents as active, computational actors, adapting the core concepts (network, space of flows, timeless time) to hybrid techno-social structures without losing their foundational utility (Figure 2).

RQ2 (AAA): The findings support the claim that GAI-enabled socio-technical arrangements can be conceptualized as operational AVVs. This refutation of technological neutrality shows that GAI actively generates, shapes, and propagates specific interests, biases, and normative outcomes embedded within its operational logics. GAI systems are thus revealed as active, value-laden agents that influence social, economic, and political processes without continuous, real-time human intervention, but remaining bounded by human-set design objectives and socio-technical constraints.

RQ3 (shift in the power structure): The empirical investigation indicates that GAI functions as a new operational gatekeeper in the space of flows. This autonomous gatekeeping capacity—filtering, prioritizing, and generating content at scale—directly influences human behavior and cognition, challenging Castells' original dynamics of power and necessitating a reassessment of who (or what) truly holds structural power in NS, as illustrated above. The principal theoretical contribution is the MANS framework, defined by the three interrelated conceptual parameters: the multi-agent node, AVV, and the hybrid techno-social structure. This rigorous ontology captures the distributed agency of both human and artificial actors, ensuring the continued relevance of Castells' work in the age of AI-mediated operational gatekeeping.

Future Research Directions

The transition to a MANS model calls for empirical validation and consideration of its societal consequences. Future research should focus on the following lines of work:

- Empirical measurement of AVV: Metrics should be developed to quantify the influence of GAI on public opinion, market behavior, and political discourse, moving beyond anecdotal evidence (Bar-Gil, 2023; Voinea, 2025).
- Operational dependence metrics: Future studies should develop indicators to quantify human-in-the-loop dependence in AAA/AVV systems, including the frequency and latency of human interventions or overrides, the proportion of actions executed without real-time approval, and infrastructural dependency mapping across data, compute, and platform access. These indicators would help clarify when autonomy remains operational or de facto rather than strong (i.e., self-authored ends).

- Rethinking digital sovereignty: The challenge posed by AVV to the concept of digital sovereignty in nation-states should be analyzed, particularly regarding information gatekeeping and automated journalism (Voinea, 2025).
- The transformation of MSC: The introduction of the multi-agent node challenges Castells (2009) concept of MSC which he defines as mass communication that is self-generated, self-directed, and self-selected, with the potential capacity to reach a global audience. Future studies should examine the shift from human-generated MSC to autonomous mass communication, where content production and filtering are increasingly driven by AAA and AVV. This transformation raises critical questions about the nature of public discourse and informational autonomy in MANS.
- Policy and governance integration: This involves exploring mechanisms to integrate the MANS framework into global governance, ensuring that emerging AI regulatory initiatives move beyond outdated, human-centric models. Concurrently, it requires specifying institutional pathways for multidisciplinary coordination and evidence-based support capable of matching the velocity of AVV.
- Regulatory, ethical, and institutional challenges: Future research should investigate challenges posed by AVV-mediated flows, including erosion of information trust (deepfakes), democratic oversight implications, and institutional pathways for multidisciplinary coordination (e.g., multidisciplinary inter-university centers as illustrative options for evidence-based AI governance).
- High-stakes domain validation: The model's predictive power could be empirically tested in critical technological domains (military systems, quantum computing). Lethal autonomous weapon systems performing delegated target-selection represent extreme cases of operational agency under human-set constraints, providing testbeds for assessing the model's utility where value-laden flows have geopolitical implications.
- Artificial general intelligence (AGI): Often associated with extreme-risk thresholds, it is commonly described as an AI system that is at least as capable as a human at most tasks (Morris et al., 2025). Because AVV-style GAI may operate under limited interpretability and accountability (Pasquale, 2015), coupling AGI with future computational infrastructures (including quantum computing and quantum networking) introduces configuration-level uncertainty. Accordingly, future theoretical work should investigate as a speculative, open question the possibility of emergent strong autonomy at scale within a configuration-level meta-network (as defined in the Discussion). At the same time, it should note that this remains conceptually distinct from the operational and de facto autonomy empirically demonstrated in this paper.
- Quantum infrastructure and governance strain: Technological convergence with quantum computing and networking could strain human-centric governance by introducing additional layers of network complexity and trust dependencies in future communication infrastructures (Vermaas et al., 2025). Directional functions extend to policy agendas and technical standards (Coenen et al., 2022), demanding preliminary conceptual inquiry while affirming the MANS model as a specification—not replacement—of Castells' NS.

In sum, the MANS model represents a significant theoretical advancement by specifying Castells' NS for contemporary AI-mediated environments. It preserves the legacy of Castells' work while providing

enhanced analytical and prognostic capabilities for understanding the complex, hybrid realities of twenty-first-century network societies.

CONFLICT OF INTEREST

The author declares no conflicts of interest. No external funding was received for this study.

ARTIFICIAL INTELLIGENCE USE STATEMENT

The author confirms the selective use of AI tools during the preparation of this manuscript. Specifically, Gemini (Google) was employed exclusively for linguistic and stylistic revision to enhance clarity and academic tone. Furthermore, Yomo.ai was utilized for external anti-plagiarism analysis and to provide structural quality feedback. All conceptual content and analysis remain the sole product of the human author.

REFERENCES

Anttiroiko, A.-V. (2015). Networks in context: the concept of network in Manuel Castells' theory of the network society. *Theoretical and Practical Research in Economic Fields*, *6*(1), 65-80. https://doi.org/10.14505/tpref.v6.1(11).04

Barabási, A.-L. (2016). *Network Science*. Cambridge University Press.

Bar-Gil, O. (2023). The dance of agencies in AI art—beyond the creator-tool dichotomy, a netnographic analysis of Reddit forums. *Journal of Posthuman Studies*, *7*(2), 129-149. https://doi.org/10.5325/jpoststud.7.2.0129

Baron, L. F., & Agostini, S. (2022). ICT and Migration in Latin American Pacific. Paths and Trajectories of Pacific Colombian Migrants. *Perspectivas de la Comunicación*, *15*(2), 99-138. https://doi.org/10.56754/0718-4867.1502.099

Bojić, L., Cinelli, M., Ćulibrk, D., & Delibašić, B. (2024). CERN for AI: a theoretical framework for autonomous simulation-based artificial intelligence testing and alignment. *European Journal of Futures Research*, *12*(1), art. 15. https://doi.org/10.1186/s40309-024-00238-0

Brogaard, J., Hendershott, T., & Riordan, R. (2014). High-Frequency Trading and Price Discovery. *The Review of Financial Studies*, *27*(8), 2267-2306. https://doi.org/10.1093/rfs/hhu032

Carnoy, M., & Castells, M. (2002). Globalization, the knowledge society, and the Network State: Poulantzas at the millennium. *Global Networks*, *1*(1), 1-18. https://doi.org/10.1111/1471-0374.00002

Cartea, Á., Jaimungal, S., & Penalva, J. (2015). *Algorithmic and High-Frequency Trading*. Cambridge University Press.

Castells, M. (2000). *The Rise of the Network Society* (2nd Ed.). Blackwell Publishers.

Castells, M. (2009). *Communication Power*. Oxford University Press.

Castells, M. (2010). *The Rise of the Network Society* (2nd Ed.). Wiley-Blackwell.

Cesaroni, V. (2024). Inclusive education in the age of AI: A critical perspective on policy guidelines through the lens of ecological-systemic theory of technological mediation. *Education Sciences & Society*, *15*(2), 329-341. https://doi.org/10.3280/ess2-2024oa18453

Coenen, C., Grinbaum, A., Grunwald, A., Milburn, C., & Vermaas, P. (2022). Quantum technologies and society: towards a different spin. *NanoEthics*, *16*(1), 1-6. https://doi.org/10.1007/s11569-021-00409-4

Corbin, J., & Strauss, A. (2014). *Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory*. Sage Publications.

Cummings, M. L. (2017). *Artificial Intelligence and the Future of Warfare* [Research Paper]. https://www.chathamhouse.org/sites/default/files/publications/research/2017-01-26-artificial-intelligence-future-warfare-cummings-final.pdf

Dominici, P. (2005). *La comunicazione nella società ipercomplessa: Istanze per l'agire comunicativo e la condivisione della conoscenza nella network society*. Aracne.

Eisenhardt, K. M. (1989). Building Theories from Case Study Research. *Academy of Management Review*, *14*(4), 532-550. https://doi.org/10.5465/amr.1989.4308385

Floridi, L. (2005). The ontological interpretation of informational privacy. *Ethics and Information Technology*, *7*(4), 185-200. https://doi.org/10.1007/s10676-006-0001-7

Floridi, L., & Sanders, J. W. (2004). On the morality of artificial agents. *Minds and Machines*, *14*(3), 349-379. https://doi.org/10.1023/B:MIND.0000035461.63578.9d

Frankfurt, H. G. (1971). Freedom of the Will and the Concept of a Person. *The Journal of Philosophy*, *68*(1), 5-20. https://doi.org/10.2307/2024717

Hagendorff, T. (2024). Deception abilities emerged in large language models. *Proceedings of the National Academy of Sciences*, *121*(24), art. e2317967121. https://doi.org/10.1073/pnas.2317967121

Horling, B., & Lesser, V. (2004). A survey of multi-agent organizational paradigms. *The Knowledge Engineering Review*, *19*(4), 281-316. https://doi.org/10.1017/S0269888905000317

Kirilenko, A., Kyle, A. S., Samadi, M., & Tuzun, T. (2017). The Flash Crash: High-Frequency Trading in an Electronic Market. *The Journal of Finance*, *72*(3), 967-998. https://doi.org/10.1111/jofi.12498

Kuss, P. M., & Meske, C. (2025). From Entity to Relation? Agency in the Era of Artificial Intelligence. *Communications of the Association for Information Systems*, *56*, 633-674. https://doi.org/10.17705/1CAIS.05626

Latour, B. (1992). Where Are the Missing Masses? The Sociology of a Few Mundane Artifacts. In W. E. Bijker & J. Law (Eds.), *Shaping Technology/Building Society: Studies in Sociotechnical Change* (pp. 225-258). The MIT Press.

Latour, B. (2005). *Reassembling the Social: An Introduction to Actor-Network-Theory*. Oxford University Press.

McCarthy, J. (1979). *Ascribing Mental Qualities to Machines*. https://www-formal.stanford.edu/jmc/ascribing.pdf

Miconi, A. (2011). *Reti. Origini e Struttura della Network Society*. Il Mulino.

Mitchell, M. (2019). *Artificial Intelligence: A Guide for Thinking Humans*. Farrar, Straus and Giroux.

Morris, M. R., Sohl-Dickstein, J., Fiedel, N., Warkentin, T., Dafoe, A., Faust, A., Farabet, C., & Legg, S. (2025). *Position: Levels of AGI for Operationalizing Progress on the Path to AGI*. arXiv. https://doi.org/10.48550/arXiv.2311.02462

Mukherjee, A., & Chang, H. H. (2025). *Stochastic, Dynamic, Fluid Autonomy in Agentic AI: Implications for Authorship, Inventorship, and Liability*. arXiv. https://doi.org/10.48550/arXiv.2504.04058

Pasquale, F. (2015). *The Black Box Society: The Secret Algorithms that Control Money and Information*. Harvard University Press.

Punziano, G. (2025). Adaptive Epistemology: Embracing Generative AI as a Paradigm Shift in Social Science. *Societies*, *15*(7), art. 205. https://doi.org/10.3390/soc15070205

Roberts, S. T. (2019). *Behind the Screen: Content Moderation in the Shadows of Social Media*. Yale University Press.

Russell, S., & Norvig, P. (2020). *Artificial Intelligence: A Modern Approach* (4th Ed.). Pearson.

Sartori, L. (2012). *La società dell'informazione*. Il Mulino.

Satyanarayan, A. (2024). Intelligence as Agency. In *Adjunct Proceedings of the 37th Annual ACM Symposium on User Interface Software and Technology* (art. 65). Association for Computing Machinery. https://doi.org/10.1145/3672539.3695751

Shoham, Y., & Leyton-Brown, K. (2009). *Multiagent Systems: Algorithmic, Game-Theoretic, and Logical Foundations*. Cambridge University Press.

Suchman, L. A. (2000). Embodied Practices of Engineering Work. *Mind, Culture, and Activity*, *7*(1–2), 4-18. https://doi.org/10.1080/10749039.2000.9677645

Suchman, L. A. (2007). *Human-Machine Reconfigurations: Plans and Situated Actions* (2nd Ed.). Cambridge University Press.

van Dijck, J., Poell, T., & de Waal, M. (2018). *The Platform Society: Public Values in a Connective World*. Oxford University Press.

van Dijk, J. A. G. M. (1999). The One-Dimensional Network Society of Manuel Castells. *New Media & Society*, *1*(1), 127-138. https://doi.org/10.1177/1461444899001001015

Vermaas, P. E., Possati, L., & Seskir, Z. C. (2025). *Quantum Internet, Governance, Trust, and the Promise of Secure Communication: On building a Quantum Internet that will be used*. arXiv. https://doi.org/10.48550/arXiv.2505.15852

Voinea, D. V. (2025). Reconceptualizing Gatekeeping in the Age of Artificial Intelligence: A Theoretical Exploration of Artificial Intelligence-Driven News Curation and Automated Journalism. *Journalism and Media*, *6*(2), art. 68. https://doi.org/10.3390/journalmedia6020068

Watson, S., Brezovec, E., & Romic, J. (2025). The role of generative AI in academic and scientific authorship: an autopoietic perspective. *AI & Society*, *40*(5), 3225-3235. https://doi.org/10.1007/s00146-024-02174-w

Webster, F. (2007). *Theories of the Information Society*. Routledge.

Weick, K. E. (1989). Theory Construction as Disciplined Imagination. *Academy of Management Review*, *14*(4), 516-531. https://doi.org/10.5465/amr.1989.4308376

Williams, R., & Edge, D. (1996). The social shaping of technology. *Research Policy*, *25*(6), 865-899. https://doi.org/10.1016/0048-7333(96)00885-2

Winner, L. (1980). Do Artifacts Have Politics? *Daedalus*, *109*(1), 121-136. https://www.jstor.org/stable/20024652

Wooldridge, M. (2009). *An Introduction to Multiagent Systems* (2nd Ed.). Wiley.

Yin, R. K. (2009). *Case Study Research: Design and Methods* (4th Ed.). Sage Publications.