Observatorio (OBS*)

Received: 2026-03-26 | **Accepted:** 2026-07-20 | **Published:** 2026-07-24 | **DOI**: 10.15847/OBS20262872

Communicative Transparency by Design: a conceptual framework for explainable AI as infrastructure for accessible digital communication

Bruno Jose Betti Galasso

Universidade Federal de São Paulo (UNIFESP)

São Paulo, Brazil

Abstract

The opacity of algorithmic systems embedded in digital platforms constitutes a structural barrier to communicative participation and digital inclusion. While Explainable Artificial Intelligence (XAI) has been extensively theorised as a technical and cognitive problem, its communicative implications remain undertheorised within Communication Studies. This article proposes the Communicative Transparency by Design (CTbD) framework, which reconceptualises XAI mechanisms as communicative infrastructure governing how platforms explain their decisions to heterogeneous publics. Synthesising platform studies, human-centred XAI, and algorithmic-literacy and accessibility research through an integrative conceptual analysis, the framework articulates three interdependent dimensions: Inferential Accessibility (the intelligibility of explanations for diverse user profiles), Participatory Agency (users' capacity to contest decisions through communicative acts), and Equitable Intelligibility (the differential distribution of transparency benefits across literacy, disability, and socioeconomic status). The framework's specific novelty is to shift the unit of analysis from the accuracy of an explanation to the communicative relationship it establishes: where prior accounts of algorithmic transparency and accountability ask whether a system discloses its reasoning, CTbD asks to whom that disclosure is addressed, with what affordances for response, and with what equity of effect — recasting transparency as a communicative right rather than a technical or compliance property. Implications for research, platform governance, and inclusive design are discussed.

Keywords: algorithmic transparency, explainable AI, accessible communication, platform studies, communicative infrastructure, digital inclusion

1. Introduction

Every time a user scrolls through a social media feed, applies for a loan through a banking app, or receives a health recommendation from a digital assistant, an algorithm decides what they see, whether they qualify, and what they are told. These decisions are consequential, often irreversible, and — in the vast majority of cases — opaque. The user receives an outcome, not an explanation. They experience an effect, not an account. In the language of communication, they receive a message without the possibility of decoding its logic, contesting its premises, or participating in the meaning-making process that produced it. This is, at its core, a communicative failure — a structural asymmetry between those who encode algorithmic decisions and those who must live with their consequences.

This asymmetry has not gone unnoticed. The last decade has witnessed the consolidation of Explainable Artificial Intelligence (XAI) as a research programme of considerable scope, ranging from mathematical formalisations of model interpretability (Arrieta et al., 2020; Adadi & Berrada, 2018) to human-centred frameworks for designing systems that communicate their reasoning to non-expert users (Shneiderman, 2020; Langer et al., 2021). Regulatory frameworks in the European Union — most notably the AI Act of 2024 and the General Data Protection Regulation — have institutionalised explainability as a requirement for high-risk AI systems, transforming what was once an academic desideratum into a legal obligation. Yet despite this accelerated growth, the dominant framing of XAI remains firmly technical and cognitive: the

problem is how to produce explanations that are mathematically valid and cognitively accessible to the "average user," implicitly understood as a single, undifferentiated epistemic agent.

What this framing misses is the communicative dimension of the problem. Explanation is not merely an information delivery task; it is a communicative act embedded in asymmetric social relations, shaped by platform architectures that distribute legibility unequally, and received by audiences whose capacity to make sense of algorithmic accounts differs systematically across axes of literacy, disability, language, and socioeconomic status. The question is not only whether an explanation is technically accurate — but to whom it is addressed, in what form, with what assumed competencies, and with what affordances for response and contestation.

Communication Studies is exceptionally well-positioned to make this analytical contribution, yet has been surprisingly slow to engage with XAI as an object of inquiry. While platform studies (van Dijck, Poell & de Waal, 2018; Gillespie, 2014; Bucher, 2018) and critical algorithm studies (Pasquale, 2015; Noble, 2018; Diakopoulos, 2019) have produced rich theorisations of algorithmic power and its social consequences, they have been primarily concerned with algorithmic processes as hidden or opaque — with the politics of the black box rather than the politics of its partial opening. The XAI movement, conversely, treats transparency as a value to be maximised within a technical system, without interrogating who benefits from particular forms of transparency, under what social conditions, or through what communicative mechanisms.

Although critical traditions in Communication Studies — from the political economy of communication to platform and critical data studies — have richly theorised the communicative power of platforms, they have done so chiefly by analysing what platforms conceal. What remains absent is a conceptual framework that treats the partial, designed disclosure of algorithmic reasoning — XAI — as communicative infrastructure in its own right: a system of practices, designs, and norms governing what platforms communicate about themselves, to whom, in what modalities, with what affordances for user response, and with what implications for inclusion and exclusion. This article proposes to fill that more precisely specified lacuna. Its contribution is threefold: it (a) reconceptualises XAI as communicative infrastructure rather than a technical property of a model; (b) introduces three evaluative dimensions — Inferential Accessibility, Participatory Agency, and Equitable Intelligibility — through which the communicative quality of any transparency mechanism can be assessed; and (c) reframes transparency as a communicative right whose benefits must be equitably distributed — a question that neither human-centred XAI, platform studies, nor accountability scholarship has yet foregrounded.

The Communicative Transparency by Design (CTbD) framework develops from the premise that explanatory mechanisms in AI-powered platforms are not supplementary features added to an otherwise communicationally neutral system; they are constitutive of how those platforms participate in public communication. When a recommendation engine displays "Why am I seeing this?" in a collapsible menu that requires two steps to access, written in English, using technical vocabulary, with no option for feedback — it is making communicative choices that structure who can meaningfully engage with the account and who cannot. When an adaptive learning system informs a student with a cognitive disability that "your learning path has been adjusted based on your performance profile," without specifying what was adjusted, why, or how to contest it, the explanation is communicatively present but substantively exclusionary.

As an ensemble, the three dimensions serve as both an analytical instrument — a lens through which communication researchers, platform designers, and policy analysts can evaluate, critique, and improve the communicative quality of XAI mechanisms — and a normative agenda, elaborated in full in Section 4.

The argument proceeds as follows. Section 2 develops the theoretical foundations of the framework through an integrative engagement with platform studies, critical algorithm studies, human-centred XAI, and accessible communication research. Section 3 details the methodological approach — an integrative conceptual analysis. Section 4 presents the CTbD framework in full, operationalising each of its three dimensions with concrete analytical criteria and candidate empirical indicators. Section 5 applies the framework through an integrated worked example, demonstrating how the three dimensions operate together and drawing out its role as both an analytical instrument and a normative agenda. Section 6 sets out the contributions and limitations of the framework, proposes a research agenda, and articulates the normative implications for platform governance and inclusive design.

2. Theoretical Foundations: Algorithmics, Communication, and the Politics of Explanation

The CTbD framework draws from three bodies of literature that have developed in relative parallel and whose convergence remains underexplored: platform studies and critical algorithm studies, human-centred XAI research, and accessible communication and disability studies. This section develops the theoretical engagement with each, identifying the specific contributions and tensions that motivate the framework.

2.1. Platform Studies: Algorithms as Communicative Architectures

The "platform turn" in Communication Studies — crystallised in the work of José van Dijck, Thomas Poell, and Martijn de Waal (2018) — proposed a shift in analytical focus from media content to the infrastructural logics that govern the production, circulation, and monetisation of content at scale. For van Dijck and colleagues, platforms are not neutral conduits for communication; they are active organising structures that encode commercial values, datafication imperatives, and selection mechanisms into the architectures through which users interact. The platform, in this account, is a communicative actor — not in the intentional, agentive sense, but in the structural sense: it makes choices that shape what is sayable, to whom, and under what conditions of visibility and accountability.

This infrastructural reading has a political-economic counterpart. Srnicek's (2016) analysis of platform capitalism locates the power of platforms in their position as data-extracting intermediaries whose business model depends on enclosing the very interactions they mediate — a structural condition that gives platforms a material interest in selective, rather than full, disclosure. Read through this lens, algorithmic governance is not a neutral administrative function but an exercise of infrastructural power: platforms govern communication by controlling the conditions under which their own operations become legible at all. Scholarship on algorithmic governance and accountability (Diakopoulos, 2019) has begun to specify the institutional mechanisms — auditing, disclosure mandates, contestation procedures — through which this power might be held to account, yet it stops short of theorising the communicative conditions under which disclosure becomes meaningful to differently situated publics — precisely the question this article takes up.

Tarleton Gillespie's (2014) foundational essay on the "relevance of algorithms" sharpened this insight by directing analytical attention to the role of algorithms in defining what counts as relevant, credible, and visible in public communication. Gillespie argued that algorithmic selection is not a neutral process of matching content to user preferences; it is a political act of categorisation that embeds editorial judgements within technical procedures, making them appear objective and thus harder to contest. The algorithm's

power, he noted, lies precisely in its invisibility: it produces effects without legibility, decisions without accountability.

Taina Bucher's (2018) account of "algorithmic power and politics" extended this analysis by examining the experiential dimension of algorithmic governance. Bucher proposed the concept of the "algorithmic imaginary" — the culturally distributed beliefs, expectations, and fears that users develop about how algorithms operate — to argue that algorithmic power is not only exercised through opaque technical decisions but also through the anticipatory behaviours that users develop in response to their (often inaccurate) models of how the system works. The user, in this account, is not simply subjected to algorithmic power; they participate in their own governance through the adaptations they make to maximise algorithmic favour.

What emerges from these accounts is a conception of platforms as communicative architectures that systematically disadvantage users who lack the resources, literacies, or informational access necessary to navigate algorithmic governance. Frank Pasquale's (2015) diagnosis of the "black box society" — in which consequential decisions about credit, employment, insurance, and access to services are made by algorithms whose logic is proprietary and opaque — extended this analysis to its most critical register, arguing that algorithmic opacity constitutes a form of informational domination that operates through secrecy rather than force. Safiya Umoja Noble (2018) demonstrated, through the analysis of search engine results, that algorithmic systems not only fail to explain their decisions but actively encode racial and gender biases in the outputs they produce — with consequences that are not merely inconvenient but materially harmful for marginalised communities.

This body of work establishes the critical diagnosis: algorithmic systems are communicative architectures that make consequential decisions, distribute power asymmetrically, and do so, in the main, without accountability or explanation. XAI emerges, in this context, as both a technical response to this diagnosis and a site of contestation — a terrain on which the stakes of transparency are negotiated between platform interests, regulatory imperatives, and user rights.

2.2. Human-Centred XAI: From Technical Interpretability to Communicative Account

The XAI research programme took shape in response to the growing deployment of "black-box" machine learning models — particularly deep neural networks — in high-stakes decision-making contexts. Adadi and Berrada (2018) identified the core problem as the "interpretability gap": the more powerful the model, the less transparent its reasoning, creating a tension between performance and accountability. Arrieta and colleagues (2020), in a comprehensive taxonomy of XAI methods and challenges, defined the field's central objective as producing explanations that are "understood and trusted" by target audiences — a formulation that, crucially, introduces audience-sensitivity into what had been primarily a mathematical problem. Subsequent work has sought to specify what makes such explanations meaningful rather than merely present: Hosain, Anik and Rafi (2023) distinguish functional transparency — explanation that genuinely supports understanding and oversight — from the nominal disclosure of model internals, treating meaningful explainability as a precondition for, not a by-product of, transparent AI.

This audience turn in XAI reached its fullest expression in the "human-centred" XAI movement, associated particularly with Shneiderman's (2020) proposal of a two-dimensional framework distinguishing between levels of human control and levels of computer automation. Shneiderman argued that reliable, safe, and trustworthy AI systems require design that keeps humans meaningfully in the loop — not as passive

recipients of algorithmic outputs, but as active participants capable of understanding, overriding, and contesting system decisions. Langer and colleagues (2021), synthesising contributions from multiple disciplines, proposed a more granular stakeholder-sensitive model, demonstrating that what constitutes an adequate explanation varies significantly depending on the stakeholder group: a data scientist, a domain expert, an affected individual, and a regulator have different informational needs, different technical vocabularies, and different rights with respect to the explained decision.

Jovanovic and Schmitz (2022) sharpened this insight by arguing that explainability should be understood not as a property of a system but as a user requirement — a claim that the system must satisfy with respect to specific human subjects in specific contexts. Their analysis identified the "framing effect" as a key communicative mechanism: how an explanation is framed — its vocabulary, its level of abstraction, its assumed background knowledge — shapes whether it achieves genuine understanding or produces the appearance of transparency without its substance.

What these contributions share is a progressive communicative enrichment of the XAI project: from the technical question of how to produce interpretable models, to the social question of for whom explanations are produced and under what conditions they achieve their communicative purpose. Yet even the most sophisticated human-centred accounts stop short of a fully communicative theory of XAI. They tend to treat the communication of explanations as a downstream problem of presentation design, not as a constitutive dimension of the explanation itself. They do not ask whether the communicative architecture of the platform — its assumptions about user agency, its modalities of explanation, its affordances for feedback and contestation — is itself a site of power and potential exclusion. This is the gap that the CTbD framework seeks to address.

2.3. Algorithmic Literacy and the Communicative Condition of the Explained

A growing body of empirical research has established that users' capacity to make sense of algorithmic explanations — when these are provided — is highly variable and structured by factors that are themselves socially distributed. Oeldorf-Hirsch and Neubaum's (2023) comprehensive review of algorithmic literacy research identified three key dimensions — awareness, knowledge, and skill — each of which exhibits systematic variation across educational level, age, gender, and socioeconomic status. Gran, Booth, and Bucher (2021), in a representative survey of Norwegian social media users, found that 40% reported no awareness of algorithmic curation and 20% reported only low awareness — with higher awareness concentrated among younger, male, and more educated users. Dogruel and colleagues (2021) developed and validated an algorithm literacy scale that measured these dimensions independently, finding that awareness and knowledge are not strongly correlated: users can be aware that algorithms shape their experience without possessing the technical vocabulary to interpret what an explanation says about how. A more recent integrative review by Gagrčin, Naab and Grub (2026) reaches a convergent conclusion, mapping how algorithmic media use and algorithm literacy co-evolve and confirming the systematic social stratification of these competencies.

Moon, Kim, and Jung (2025), in an experimental study of algorithmic transparency features in short-form video platforms, demonstrated a particularly consequential finding: the benefits of explainability mechanisms — in terms of perceived transparency, legitimacy, and platform satisfaction — were significantly moderated by users' algorithmic literacy levels. When neither explainability nor user control features were present, literacy had no significant impact. But when at least one transparency feature was present, literacy created a new dimension of differentiation: highly literate users benefited substantially from the features, while low-

literacy users did not. The authors concluded that "literacy creates a new dimension of the digital divide, where transparency benefits are unequally experienced." This finding is of central theoretical importance for the CTbD framework: algorithmic transparency mechanisms are not neutral interventions that benefit all users equally; they are communicative acts whose effects are structured by the differential communicative competencies of the audience.

The implications for accessible communication are direct. Bitman (2022), studying social media use among users with concealable communicative disabilities — autistics, hard-of-hearing individuals, and people who stutter — demonstrated that platform architectures impose communicative norms that are designed with non-disabled users as the implicit default, systematically disadvantaging users whose communicative repertoires differ from this norm. Trevisan (2022) extended this analysis to the political sphere, arguing that digital inclusivity should be understood not as a technical accommodation but as a "process" through which platforms actively structure the communicative participation of marginalised users. Safari and colleagues (2023) showed that participatory design activities with users with intellectual disabilities can substantially improve both digital skills and engagement with technology — but only when design processes actively incorporate the voices of those users rather than treating accessibility as an afterthought.

Together, these contributions converge on a recognition that has yet to be fully integrated into either XAI research or platform studies: the communicative quality of an algorithmic explanation is not determined solely by its technical properties or even its linguistic presentation. It is co-constituted by the communicative conditions of the audience — their literacies, their cognitive and sensory profiles, their access to contesting mechanisms, and their social position relative to the platform.

2.4. Synthesis: Towards a Communicative Theory of Algorithmic Transparency

Bringing together these three bodies of literature, it is possible to identify the theoretical architecture from which the CTbD framework is constructed. From platform studies, the framework inherits the understanding that platforms are communicative architectures that structure power asymmetrically and that algorithmic opacity is a political condition, not a technical limitation. From human-centred XAI, the framework inherits the recognition that explanation is an audience-sensitive communicative act that must be evaluated in relation to the needs, capacities, and rights of specific stakeholder groups. From algorithmic literacy and accessible communication research, the framework inherits the empirical finding that the communicative benefits of transparency mechanisms are not uniformly distributed — they reproduce existing inequalities unless they are explicitly designed to counteract them.

The synthesis yields the foundational claim of the CTbD framework: algorithmic transparency is a communicative right that can only be realised through design choices that actively account for the heterogeneity of communicative conditions across the user population. "Transparency by Design" — borrowed from but extended beyond the legal concept of "Privacy by Design" — proposes that communicative accessibility, participatory affordances, and equitable intelligibility must be built into the architecture of XAI mechanisms from the outset, not retrofitted as accessibility accommodations for edge cases.

This positions the CTbD framework within a tradition in Communication Studies that treats communication not merely as information transmission but as the constitution of shared meaning under conditions of power and inequality. In Habermasian terms (1984), communicative action oriented towards understanding requires not merely the transmission of propositional content but the conditions for free, competent, and

undistorted participation. Algorithmic transparency mechanisms that are technically present but communicatively inaccessible reproduce the distortions of strategic action — the use of communication as an instrument of domination — rather than enabling the communicative rationality that is a precondition of democratic accountability.

3. Methodological Approach: Integrative Conceptual Analysis

The present article adopts an integrative conceptual analysis as its methodological approach — a form of scholarly synthesis that produces original theoretical contributions by identifying, mapping, and integrating insights from multiple, partially overlapping bodies of literature (Torraco, 2016). Unlike systematic reviews, which aim to aggregate empirical findings, or narrative reviews, which synthesise existing knowledge within a single field, integrative conceptual analysis is explicitly oriented towards producing new conceptual frameworks that advance the theoretical frontier by reorganising existing knowledge in light of a novel analytical question.

This choice is methodologically appropriate for two reasons. First, the research question — how should XAI be conceptualised as communicative infrastructure? — is a theoretical question that cannot be answered by aggregating empirical findings; it requires conceptual work that builds bridges across disciplinary traditions that have not systematically engaged with each other. Second, the absence of an existing framework at the intersection of platform studies, XAI research, and accessible communication means that the contribution of this article is necessarily theoretical in nature: it provides a conceptual vocabulary and analytical structure that empirical research can subsequently employ and test.

3.1. Literature Identification and Selection

The analysis drew on three intersecting bodies of literature, each mapped through structured searches in Scopus and Web of Science, complemented by targeted searches in the scite.ai database for citation-context verification.

The substantive corpus prioritised peer-reviewed scholarship from the last decade (2014–2026) — the period over which platform studies, human-centred XAI, and algorithmic-literacy research consolidated as distinct programmes — while retaining a small number of foundational works outside this window for conceptual anchoring (notably Shannon, 1948, on the transmission model of communication, and Habermas, 1984, on communicative action). Within each set, a work was retained when it (a) advanced a theoretical or empirical claim about how algorithmic systems are communicated to, or made intelligible by, human users, and (b) appeared in a peer-reviewed venue or a university-press monograph; purely technical XAI contributions with no engagement with human audiences, and commentaries without primary argument or evidence, were excluded. Following the purposive logic of integrative review (Torraco, 2016), selection was guided by conceptual saturation rather than exhaustive coverage: works were added to each set until its central claims, key tensions, and unexplored adjacencies were adequately represented.

- Literature Set 1: Platform Studies and Critical Algorithm Studies: Search terms combined ("platform society" OR "platform capitalism" OR "algorithmic governance") AND ("communication" OR "media"). Priority was given to theoretical and empirical contributions that advanced the understanding of algorithms as communicative architectures. Key anchoring works include van

Dijck, Poell and de Waal (2018), Gillespie (2014), Bucher (2018), Pasquale (2015), Noble (2018), and Diakopoulos (2019).

- Literature Set 2: Human-Centred Explainable AI. Search terms combined ("explainable artificial intelligence" OR "XAI") AND ("user" OR "communication" OR "stakeholder" OR "human-centred"). Priority was given to works that moved XAI research beyond purely technical concerns. Key anchoring works include Arrieta et al. (2020), Shneiderman (2020), Langer et al. (2021), Adadi and Berrada (2018), and Jovanovic and Schmitz (2022).
- Literature Set 3: Algorithmic Literacy and Accessible Communication. Search terms combined ("algorithmic literacy" OR "digital inclusion" OR "accessible communication") AND ("disability" OR "inequality" OR "platform"). Key anchoring works include Oeldorf-Hirsch and Neubaum (2023), Gran, Booth and Bucher (2021), Dogruel et al. (2021), Moon, Kim and Jung (2025), Bitman (2022), Trevisan (2022), and Safari et al. (2023).

3.2. Analytical Procedure

The integrative analysis proceeded through four stages. In the first stage, each body of literature was mapped independently, identifying its central claims, key tensions, and unexplored adjacencies. In the second stage, a cross-literature mapping exercise identified points of convergence and divergence. In the third stage, an analytical synthesis was developed by tracing what each body of literature contributes to a communicative theory of algorithmic transparency. In the fourth stage, the CTbD framework was constructed from this synthesis, with each of its three dimensions derived from specific convergences across the three bodies of literature.

The derivation of each dimension followed an explicit analytical chain, summarised in Table 1. No dimension was imposed on the literature; each crystallised a specific convergence across the three bodies. Inferential Accessibility crystallised the audience-sensitivity principle of human-centred XAI — the recognition that what counts as an adequate explanation depends on the inferential needs of a specific user. Participatory Agency crystallised platform studies' identification of contestability as a democratic requirement distinct from disclosure. Equitable Intelligibility crystallised the empirical finding, from algorithmic-literacy and accessibility research, that the benefits of transparency are unequally distributed across the user population. The four analytical criteria attached to each dimension (Section 4) were then generated by asking, for each convergence, what observable properties a transparency mechanism would need to exhibit for the dimension to be satisfied.

Table 1. Derivation of the three CTbD dimensions from the three bodies of literature

Body of literature (key works)	**Core contribution and the conceptual bridge it constructs**	**CTbD dimension primarily informed**
Platform studies & critical algorithm studies (van Dijck et al., 2018; Gillespie, 2014; Bucher, 2018; Pasquale, 2015; Noble, 2018; Srnicek, 2016; Diakopoulos, 2019)	Platforms are communicative architectures that structure power asymmetrically; opacity is a political, not merely technical, condition; transparency without contestability is insufficient for accountability. Bridge: from "platforms as opaque architectures" to "platforms whose transparency practices can be evaluated and required".	Participatory Agency
Human-centred XAI (Arrieta et al., 2020; Shneiderman, 2020; Langer et al., 2021; Adadi & Berrada, 2018; Jovanovic & Schmitz, 2022; Hosain et al., 2023)	Explanation is an audience-sensitive communicative act, not mere information delivery; "framing effects" determine whether understanding or only its appearance is achieved; meaningful explainability is distinct from nominal disclosure. Bridge: from "explanation as accurate output" to "explanation as a communicatively successful act for a specific user".	Inferential Accessibility

Algorithmic literacy & accessible communication (Oeldorf-Hirsch & Neubaum, 2023; Gagrčin et al., 2026; Gran et al., 2021; Dogruel et al., 2021; Moon et al., 2025; Bitman, 2022; Trevisan, 2022; Safari et al., 2023)	The communicative benefits of transparency are unequally distributed; literacy creates a new dimension of the digital divide; accessibility is an active, participatory design process, not compliance. Bridge: from "accessibility as WCAG compliance" to "equitable intelligibility as the designed distribution of transparency benefits".	Equitable Intelligibility

Source: Author's own elaboration.

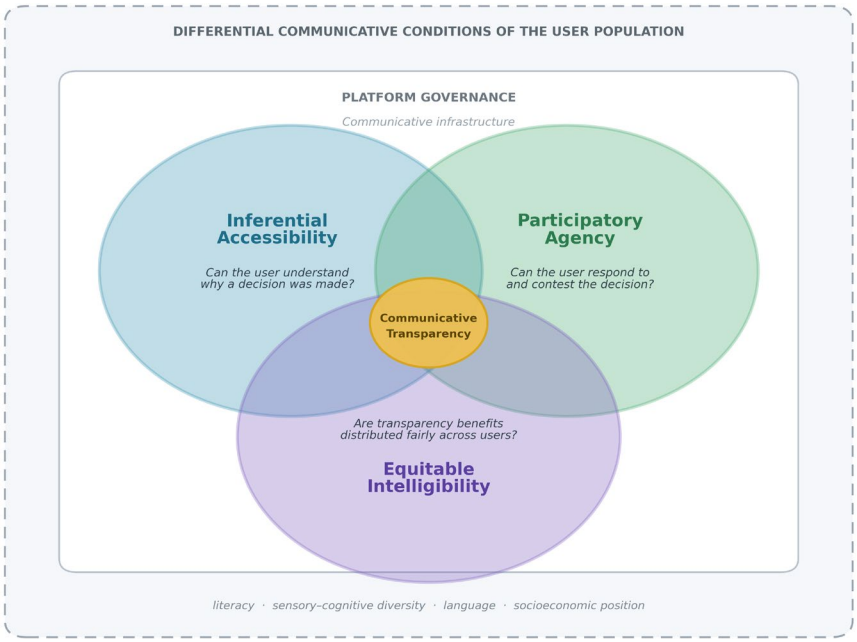
The integrative conceptual analysis has inherent limitations that must be acknowledged. It does not produce empirical evidence for the claims advanced by the CTbD framework; it produces conceptual resources for generating and evaluating such evidence. The framework's three dimensions are analytically derived and await empirical operationalisation and testing across different platform types, user populations, and regulatory contexts. The framework is therefore offered as a starting point for a research agenda, not as a finished theoretical edifice.

4. The Communicative Transparency by Design (CTbD) Framework

The Communicative Transparency by Design (CTbD) framework proposes that algorithmic transparency mechanisms in digital platforms should be understood and evaluated as communicative infrastructure — systems of design choices, affordances, and norms that structure the communicative relationship between platforms and their users with respect to algorithmic decision-making. The framework organises this infrastructure around three interdependent analytical dimensions: Inferential Accessibility, Participatory Agency, and Equitable Intelligibility. These dimensions are not sequential stages but co-constitutive aspects of a unified communicative condition; each is analytically separable but empirically intertwined.

Figure 1. presents the visual representation of the CTbD framework, showing the three dimensions as mutually reinforcing components of communicative infrastructure, embedded within the broader context of platform governance and the differential communicative conditions of the user population.

Figure 1. The Communicative Transparency by Design (CTbD) Framework.



Source: Author's own elaboration.

Before turning to each dimension in detail, Table 2 provides a compact overview of the framework — its three dimensions, the four analytical criteria that operationalise each, candidate empirical indicators, and an illustrative application — so that the model can be read at a glance as an analytical tool.

Table 2. *The CTbD framework: dimensions, analytical criteria, candidate indicators, and illustrative applications.*

Dimension	Analytical criteria	Candidate empirical indicators	Illustrative application
Inferential Accessibility (Can the user understand why?)	• Vocabulary fitness • Causal specificity • Counterfactual clarity • Modality appropriateness	• Match between technical register and audience profile • Presence of decision-specific (not generic) causal factors • Presence of an actionable counterfactual • Availability in non-textual modalities	A credit-scoring decline stating only the score and threshold meets causal specificity but fails counterfactual clarity.
Participatory Agency (Can the user respond and contest?)	• Availability • Responsiveness • Consequentiality • Collectivity affordance	• Steps / time to reach contestation controls • Presence and latency of system acknowledgement • Measurable change in outputs after user input • Mechanisms to aggregate similar complaints	A "Not interested" control that is present but yields no acknowledgement or demonstrable change scores low on responsiveness and consequentiality.
Equitable Intelligibility (Who benefits from transparency?)	• Multi-literacy addressability • Disability-inclusive design • Socioeconomic equity of access • Linguistic accessibility	• Number of complexity registers offered • WCAG 2.2 conformance level • Functionality on low-bandwidth / low-end devices • Coverage of the platform's principal user languages	An adaptive-learning explanation offered in one register, one language, and one modality scores low across all four criteria.

Source: Author's own elaboration.

4.1. Dimension 1: Inferential Accessibility

Inferential Accessibility refers to the degree to which the explanation that a platform provides for an algorithmic decision enables a user to form a warranted inference about why that decision was made — that is, to move from the explanation as received to a defensible understanding of the algorithmic process that produced the outcome.

The concept draws on the pragmatics of communication rather than on formal semantics: it is not enough that an explanation be propositionally true or technically accurate; it must be formulated in a way that enables the specific user to perform the inferential operations required to arrive at understanding. This is a richer condition than "intelligibility" in the weak sense (the user can read the words) or "transparency" in the technical sense (the model's internal logic is exposed). An explanation that is technically transparent but inferentially inaccessible — for example, a SHAP value visualisation presented to a user without statistical training — fails the condition of Inferential Accessibility.

Arrieta and colleagues' (2020) taxonomy of XAI methods provides a useful starting point by distinguishing between explanations addressed to AI experts, domain experts, and non-expert end users. Their formulation that the "target audience" of an explanation determines what counts as adequate closely aligns with the communicative premise of Inferential Accessibility. Langer et al.'s (2021) stakeholder-sensitive model further specifies that the "information needs" of different stakeholder groups vary not only in content but in the cognitive and communicative form in which information must be presented. Jovanovic and Schmitz (2022) ground this in communication theory by noting that the "framing effect" — how an explanation is presented

— can systematically shape whether genuine understanding is achieved or merely the appearance of transparency is produced.

For empirical research, Inferential Accessibility can be operationalised along four analytical criteria:

1) Vocabulary fitness: Does the explanation use vocabulary calibrated to the user's likely communicative competencies, or does it assume technical expertise that the user population does not possess?
2) Causal specificity: Does the explanation provide sufficient information about the causal factors that produced the decision, or does it offer only generic accounts that could apply to any decision?
3) Counterfactual clarity: Does the explanation enable the user to understand what would have needed to be different for a different decision to be made — a criterion that research on human reasoning suggests is central to the perception of fairness (Langer et al., 2021)?
4) Modality appropriateness: Is the explanation presented in a modality (textual, visual, auditory, multimodal) that is accessible to the specific user, accounting for sensory and cognitive diversity?

Illustrative analysis: A credit scoring application that displays "Your application was declined because your credit score of 620 falls below our threshold of 650" satisfies the criterion of causal specificity but fails counterfactual clarity (the user is not told what they could do to change the outcome) and may fail modality appropriateness for users with visual impairments. A social media feed explanation that states "You are seeing this post because people with similar interests have engaged with it" fails vocabulary fitness and causal specificity simultaneously — it provides the appearance of transparency without its inferential substance.

4.2. Dimension 2: Participatory Agency

Participatory Agency refers to the degree to which the communicative architecture surrounding an algorithmic explanation affords the user substantive capacities to respond to, contest, revise, or influence the algorithmic decision — in short, to participate in the communicative exchange that the explanation opens up rather than merely receive it.

This dimension is motivated by the observation that transparency, however necessary, is insufficient for democratic accountability if it is not accompanied by mechanisms for user response. Pasquale (2015) argued that the black box society is characterised not only by algorithmic opacity but by the absence of contestability. Noble (2018) demonstrated that knowing about algorithmic bias is insufficient to change it in the absence of mechanisms for collective and individual response. The AI Act's "right to contest" automated decisions represents a regulatory acknowledgement that transparency and contestability are distinct rights, each of which must be actively secured.

Participatory Agency, in the CTbD framework, encompasses the full range of communicative affordances through which users can engage with algorithmic decisions: feedback mechanisms, preference settings, appeal procedures, requests for human review, and collective forms of algorithmic advocacy. What distinguishes strong from weak Participatory Agency is not the mere availability of these mechanisms but their communicative quality — the degree to which they are accessible, responsive, and capable of producing outcomes.

The research of Moon, Kim and Jung (2025) provides empirical support for this distinction. Their experimental study demonstrated that user control features in short-form video platforms enhanced perceived algorithmic transparency and legitimacy — but only when users possessed sufficient algorithmic literacy to engage with them meaningfully. Participatory Agency is therefore not a binary condition (present or absent) but a spectrum condition whose realisation depends on the communicative competencies of the user.

Four analytical criteria for Participatory Agency:

1) Availability: Are feedback, contestation, and preference-setting mechanisms readily accessible within the communicative flow of platform use, or are they buried in settings menus accessible only to highly motivated users?
2) Responsiveness: Does the platform communicate back to the user in response to their input — confirming receipt, explaining what change if any will result, and providing a timeframe — or does it create the appearance of response without the substance?
3) Consequentiality: Does user input through contestation mechanisms produce demonstrable changes in algorithmic outputs, or does it function primarily as a performative gesture that satisfies regulatory requirements without changing platform behaviour?
4) Collectivity affordance: Does the platform provide mechanisms for collective as well as individual contestation — enabling users who share similar algorithmic experiences to aggregate their responses and amplify their communicative power?

Illustrative analysis: The "Not Interested" button on video recommendation platforms provides weak Participatory Agency: it is available, but research suggests that its consequentiality is limited and its feedback mechanism is absent. The European GDPR's "right to explanation" for automated decisions provides a stronger but uneven form of Participatory Agency: it is legally backed and can produce consequential outcomes, but its practical accessibility is variable across socioeconomic groups (Zharova, 2023).

4.3. Dimension 3: Equitable Intelligibility

Equitable Intelligibility is the dimension of the CTbD framework that is most directly concerned with inclusion and justice. It refers to the degree to which the communicative benefits of algorithmic transparency mechanisms are distributed equitably across the user population — that is, whether all users, regardless of their literacy levels, cognitive and sensory profiles, language backgrounds, and socioeconomic positions, have substantively equivalent access to the understanding and agency that transparent algorithms are supposed to enable.

This dimension is motivated by the empirical evidence that algorithmic transparency mechanisms, as currently designed, reproduce and in some cases amplify existing communicative inequalities. Oeldorf-Hirsch and Neubaum's (2023) review established that algorithmic literacy is socially structured in ways that correlate with gender, age, and education. Moon et al.'s (2025) experimental findings demonstrated that transparency features benefit high-literacy users substantially more than low-literacy users. Gran, Booth and Bucher (2021) showed that the new "algorithmic divide" maps onto, and compounds, older forms of the digital divide. Bitman (2022) demonstrated that communicative disabilities are systematically unaddressed in platform design.

Equitable Intelligibility, as an analytical dimension, requires asking not "Is the explanation intelligible?" but "Is the explanation intelligible to whom?" — and specifically, "What design choices would need to change for it to be intelligible to those currently excluded?"

Four analytical criteria for Equitable Intelligibility:

1) Multi-literacy addressability: Is the explanation available in multiple registers of technical complexity, enabling users with different levels of algorithmic literacy to access accounts appropriate to their competencies?
2) Disability-inclusive design: Is the explanation compliant with universal design principles (WCAG 2.2 or successor standards), ensuring that users with visual, auditory, cognitive, and motor disabilities can access the same communicative content through accessible modalities?
3) Socioeconomic equity of access: Are the explanatory and contestation mechanisms available across device types and connectivity levels, or do they presuppose high-end devices and broadband connections not uniformly available across the user population?
4) Linguistic accessibility: Is the explanation available in the languages of the platform's user population, with translations that preserve the communicative substance of the explanation?

Illustrative analysis: The increasing deployment of AI-generated explanations in educational platforms provides a rich site for analysing Equitable Intelligibility. When an adaptive learning system adjusts a student's learning path and provides an explanation, the communicative accessibility of that explanation to a student with a reading disability, to a student for whom the platform language is a second language, and to a student accessing the platform on a mobile device with limited data may differ radically. Safari et al.'s (2023) participatory design research suggests that involving users with intellectual disabilities in the design of technology explanations substantially improves both the accessibility and the perceived trustworthiness of those explanations.

5. Applying the CTbD Framework: An Integrated Worked Example

To show how the three dimensions operate together rather than in isolation, consider a single, familiar case: the "Why am I seeing this post?" explanation offered by a large social-media platform when a user taps the menu beside a recommended item in their feed. Analysed through the CTbD framework, the same mechanism yields a differentiated diagnosis across the three dimensions.

On Inferential Accessibility, a typical explanation of the form "This post is recommended because you follow similar accounts and many people like you engaged with it" fares poorly. The vocabulary is serviceable (vocabulary fitness is partially met), but the account offers no decision-specific causal factors (which accounts? which signals weighed most?), no counterfactual (what the user could change to stop seeing such posts), and a single textual modality. The user cannot move from the explanation to a warranted inference about why this item, rather than another, was surfaced.

On Participatory Agency, the adjacent controls — "See fewer posts like this", "Not interested", "Hide" — are available within the communicative flow (availability is met), but they are typically silent: the platform rarely confirms what, if anything, changed, or on what timescale (weak responsiveness), and independent evidence that such signals durably alter recommendations is limited (weak consequentiality). There is no

affordance for users who share a grievance — a community repeatedly shown harmful content, say — to aggregate and escalate it (collectivity is absent).

On Equitable Intelligibility, the single explanation is addressed to one notional literate user. It is rarely offered in multiple registers of complexity; its accessibility to screen-reader and cognitively diverse users depends on uneven implementation; it presupposes the data and device conditions of a well-connected user; and it is frequently unavailable in the full range of the platform's user languages. By Moon, Kim and Jung's (2025) finding, whatever benefit the feature confers will accrue disproportionately to already high-literacy users.

Read across the three dimensions, the mechanism is communicatively present but substantively thin: it discloses something, affords little response, and distributes even that little unequally. The integrated reading is the point — improving one dimension alone (clearer wording, say) while leaving the others untouched would still fail the test of communicative infrastructure. The same three-dimensional analysis can be applied, with different results, to a GDPR Article 22 explanation of an automated credit decision, an adaptive-learning system's account of a re-sequenced learning path, or a content-moderation notice, making the framework a portable instrument for the empirical platform audits proposed in Section 6.

5.1. The CTbD Framework as Analytical Instrument and Normative Agenda

The three dimensions of the CTbD framework are analytically separable but empirically interdependent. A platform may score well on Inferential Accessibility while failing on Participatory Agency: explanations are clear but cannot be contested. It may provide strong Participatory Agency for literate users while performing poorly on Equitable Intelligibility: contestation mechanisms are available but practically inaccessible to low-literacy users. It may satisfy formal disability compliance requirements while providing explanations of such low causal specificity that they fail the criteria for Inferential Accessibility.

This interdependence is represented in Figure 1 as overlapping circles within a common communicative context, emphasising that the realisation of communicative transparency requires progress on all three dimensions simultaneously. A platform that achieves strong performance on all three dimensions approximates what the framework calls Communicative Transparency — a condition in which algorithmic accountability is not merely technically present but substantively realised across the diversity of the user population.

The normative claim embedded in the framework is that Communicative Transparency is a democratic right, not a market preference. When platforms communicate about their algorithmic decisions — or fail to do so — they are exercising a form of communicative power that affects users' capacity to participate in social, economic, and political life on equal terms. The CTbD framework proposes that the communicative quality of this power should be subject to the same standards of public accountability as other forms of communicative power in democratic societies.

5.2. Research Implications and Applications

The CTbD framework generates research implications that extend across platform studies, XAI research, and accessible communication scholarship. For platform studies, the framework provides a vocabulary for analysing the communicative architectures of specific platforms with respect to their transparency practices — moving beyond the general claim that platforms are opaque towards a differentiated analysis of how specific design choices affect specific user populations. For XAI research, the framework proposes a

communicative evaluation criterion that complements existing technical and cognitive metrics. Rather than asking only "Is this explanation accurate?", XAI researchers can ask "Is this explanation inferentially accessible to the target population?", "Does it enable meaningful participatory response?", and "Is its intelligibility equitably distributed?" For accessible communication and disability studies, the framework extends the domain of inquiry from the accessibility of content to the accessibility of accountability. For platform governance and regulation, the framework provides a multi-dimensional evaluation instrument that can inform regulatory standards beyond the current emphasis on the presence or absence of explanations. In practice, each of the twelve criteria can be scored on a simple ordinal scale (absent / partial / met), yielding a dimension-level and overall communicative-transparency profile for a given mechanism that supports comparison across platforms and over time.

6. Final Considerations: Transparency as Communicative Right

This article has proposed the Communicative Transparency by Design (CTbD) framework as a conceptual instrument for reconceptualising Explainable AI as communicative infrastructure within Communication Studies, addressing a theoretical lacuna at the intersection of platform studies, human-centred XAI, and accessible-communication research.

The argument developed across the preceding sections can be summarised in three propositions. First, algorithmic transparency is not primarily a technical problem but a communicative one: the question is not only whether explanations are technically accurate or cognitively accessible, but whether they enable the specific, heterogeneous, and unequally situated users of digital platforms to form warranted inferences, respond meaningfully, and participate in the social processes that algorithmic decisions shape. Second, the communicative benefits of transparency mechanisms are not uniformly distributed: the empirical evidence from algorithmic literacy research, accessible communication scholarship, and XAI human-centred studies consistently shows that these benefits accrue disproportionately to users who already possess the literacies, devices, and social positions that enable effective engagement with digital systems. Third, communicative transparency is a democratic right that must be designed, not a market preference that will emerge spontaneously from platform competition: realising this right requires design processes that are inclusive by default, regulatory frameworks that evaluate communicative quality rather than merely formal compliance, and research agendas that treat the distribution of transparency benefits as a question of justice.

The CTbD framework makes three principal theoretical contributions to Communication Studies and adjacent fields. The first is conceptual: the framework proposes "communicative infrastructure" as a category for analysing how platforms mediate accountability relationships. This concept extends the platform studies tradition beyond its focus on content production and distribution to encompass the communicative practices through which platforms explain — or fail to explain — their own operations to their users. The second is analytical: the framework provides three operationalisable dimensions — Inferential Accessibility, Participatory Agency, and Equitable Intelligibility — that researchers can employ to evaluate specific transparency mechanisms in specific platform contexts. The third is normative: the framework contributes a normative agenda — transparency as a communicative right — that challenges the dominant framings of XAI as either a technical optimisation problem or a regulatory compliance requirement. By anchoring the normative claim in the theoretical tradition of communicative action (Habermas, 1984) and its implications for democratic participation, the framework connects XAI to longstanding debates in communication theory about the conditions for legitimate public communication.

The integrative conceptual analysis has inherent limitations that must be stated clearly. The CTbD framework is a theoretical product: it provides conceptual resources for empirical research but does not itself constitute empirical evidence for the claims it advances. Each of its three dimensions, and the interactions between them, requires empirical operationalisation and testing across diverse platform types, user populations, and regulatory contexts before the framework's utility can be assessed with confidence.

Several specific limitations deserve attention. The framework's analytical criteria — while derived from the literature — are not yet operationalised into measurement instruments; this is a necessary next step. The framework assumes a level of design agency that may overestimate the practical autonomy of platform designers, who operate within technical constraints, competitive pressures, and regulatory environments that shape what is feasible. Finally, the framework's normative claim — that communicative transparency is a democratic right — is a philosophical proposition whose defence would require a more extended engagement with political philosophy than the scope of this article permits.

Future research could advance the CTbD agenda in at least four directions. First, empirical platform audits using the three CTbD dimensions as evaluation criteria could map the communicative transparency landscape across major platform types and user populations. Second, participatory design studies involving users with disabilities, low algorithmic literacy, and low socioeconomic status could test the framework's prescriptive claims. Third, comparative regulatory analysis could assess whether existing transparency requirements (GDPR, AI Act) satisfy the CTbD criteria. Fourth, computational research could explore whether XAI methods can be designed from the outset to satisfy the framework's Inferential Accessibility and Equitable Intelligibility criteria simultaneously.

The CTbD framework argues, ultimately, that communication must be placed at the centre of how we understand, evaluate, and design algorithmic transparency. Not communication as information delivery — the sender-to-receiver transmission of accurate outputs to passive receivers, in the sense canonically formalised by Shannon (1948), — but communication as the condition for meaningful participation in the social processes that algorithms shape. The transparency that matters, in this account, is not the transparency that satisfies a legal requirement or an engineer's interpretability metric. It is the transparency that enables a person — any person, with whatever combination of literacies, abilities, and social positions they bring to their digital life — to understand what has been decided about them, to respond with their own account, and to contest decisions that affect them on equal terms with others.

This is a demanding standard. It is also, precisely because of its demands, a productive one: it names what is missing in current transparency practices, specifies what must be designed, and motivates the research and governance agendas that a genuinely communicative approach to algorithmic accountability would require. Communication Studies, with its theoretical resources for understanding how meaning is made under conditions of power, is well-positioned to make this contribution — and the CTbD framework is offered as a step towards doing so.

Acknowledgements / Funding
The author received no financial support for this research.

Conflict of interest
The author declares no conflict of interest.

Ethical statement
This study involves no primary data collection and no human subjects; all sources cited are publicly available peer-reviewed scholarship. It was conducted in accordance with the principles of scientific research and did not require additional ethics committee approval.

Declaration of AI usage
No AI-generated text was used in the writing of this manuscript. Generative AI tools were used solely for bibliographic verification (scite.ai) and literature-search support.

Data availability
This article is a theoretical contribution and involved no primary data collection. All material supporting the analysis, including the derivation of the three framework dimensions, is contained within the article (Table 1 and Section 3).

Author contributions
The author is solely responsible for the conception, research, writing, and revision of this manuscript.

References

Adadi, A., & Berrada, M. (2018). Peeking inside the black-box: A survey on Explainable Artificial Intelligence (XAI). *IEEE Access*, *6*, 52138–60. https://doi.org/10.1109/access.2018.2870052

Arrieta, A. B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., Garcia, S., Gil-Lopez, S., Molina, D., Benjamins, R., Chatila, R., & Herrera, F. (2020). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion*, *58*, 82–115. https://doi.org/10.1016/j.inffus.2019.12.012

Bitman, N. (2022). "Authentic" digital inclusion? Dis/ability performances on social media by users with concealable communicative disabilities. *New Media & Society*, *24*(2), 401–419. https://doi.org/10.1177/14614448211063183

Bucher, T. (2018). *If...Then: Algorithmic power and politics*. Oxford University Press. https://doi.org/10.1093/oso/9780190493028.001.0001

Diakopoulos, N. (2019). *Automating the news: How algorithms are rewriting the media*. Harvard University Press. https://doi.org/10.4159/9780674239302

Dogruel, L., Facciorusso, D., & Stark, B. (2021). Development and validation of an algorithm literacy scale for internet users. *Communication Methods and Measures*, *16*(2), 115–133. https://doi.org/10.1080/19312458.2021.1968361

Gagrčin, E., Naab, T. K., & Grub, M. F. (2026). Algorithmic media use and algorithm literacy: An integrative literature review. *New Media & Society*. https://doi.org/10.1177/14614448241291137

Gillespie, T. (2014). The relevance of algorithms. In T. Gillespie, P. J. Boczkowski, & K. A. Foot (Eds.), *Media technologies: Essays on communication, materiality, and society* (pp. 167–193). MIT Press. https://doi.org/10.7551/mitpress/9780262525374.003.0009

Gran, A.-B., Booth, P., & Bucher, T. (2021). To be or not to be algorithm aware: A question of a new digital divide? *Information, Communication & Society*, *24*(12), 1779–1796. https://doi.org/10.1080/1369118X.2020.1736124

Habermas, J. (1984). *The theory of communicative action. Vol. 1: Reason and the rationalization of society* (T. McCarthy, Trans.). Beacon Press.

Hosain, M. T., Anik, M. H., & Rafi, S. (2023). Path to gain functional transparency in artificial intelligence with meaningful explainability. *Journal of Metaverse*, *3*(2), 166–180. https://doi.org/10.57019/jmv.1306685

Jovanovic, M., & Schmitz, M. (2022). Explainability as a user requirement for artificial intelligence systems. *Computer*, *55*(2), 90–94. https://doi.org/10.1109/mc.2021.3127753

Langer, M., Oster, D., Speith, T., Hermanns, H., Kästner, L., Schmidt, E., Sesing, A., & Baum, K. (2021). What do we want from Explainable Artificial Intelligence (XAI)?—A stakeholder perspective on XAI and a conceptual model guiding interdisciplinary XAI research. *Artificial Intelligence*, *296*, 103473. https://doi.org/10.1016/j.artint.2021.103473

Moon, J. H., Kim, S. H., & Jung, Y.-J. (2025). The effects of explainability and user control on algorithmic transparency: The moderating role of algorithmic literacy. *Cyberpsychology, Behavior, and Social Networking*, *28*(7), 497–504. https://doi.org/10.1089/cyber.2024.0525

Noble, S. U. (2018). *Algorithms of oppression: How search engines reinforce racism*. NYU Press.

Oeldorf-Hirsch, A., & Neubaum, G. (2023). What do we know about algorithmic literacy? The status quo and a research agenda for a growing field. *New Media & Society*, *27*(2), 681–701. https://doi.org/10.1177/14614448231182662

Pasquale, F. (2015). *The black box society: The secret algorithms that control money and information*. Harvard University Press.

Safari, M. C., Wass, S., & Thygesen, E. (2023). Digital technology design activities—A means for promoting the digital inclusion of young adults with intellectual disabilities. *British Journal of Learning Disabilities*, *51*(2), 238–249. https://doi.org/10.1111/bld.12521

Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, *27*(3), 379–423. https://doi.org/10.1002/j.1538-7305.1948.tb01338.x

Shneiderman, B. (2020). Human-centered artificial intelligence: Reliable, safe & trustworthy. *International Journal of Human–Computer Interaction*, *36*(6), 495–504. https://doi.org/10.1080/10447318.2020.1741118

Srnicek, N. (2016). *Platform capitalism*. Polity Press.

Torraco, R. J. (2016). Writing integrative literature reviews: Using the past and present to explore the future. *Human Resource Development Review*, *15*(4), 404–428. https://doi.org/10.1177/1534484316671606

Trevisan, F. (2022). Beyond accessibility: Exploring digital inclusivity in US progressive politics. *New Media & Society*, *24*(2), 496–513. https://doi.org/10.1177/14614448211063187

van Dijck, J., Poell, T., & de Waal, M. (2018). *The platform society: Public values in a connective world*. Oxford University Press. https://doi.org/10.1093/oso/9780190889760.001.0001

Zharova, A. (2023). Achieving algorithmic transparency and managing risks of data security when making decisions without human interference: Legal approaches. *Journal of Digital Technologies and Law*, *1*(4), 973–993. https://doi.org/10.21202/jdtl.2023.42