International Journal of Modern Computer Science and IT Innovations (IJMCSII)
eISSN: 3087-4289
Volume. 02, Issue. 12, pp. 27-39, December 2025"

Re-coding Community: Designing AI-Native Platforms for Trust, Belonging, and Collective Agency

Anastasiia Livintseva
Miami, USA

Article received: 12/08/2025, Article Accepted: 12/27/2025, Article Published: 12/31/2025

ABSTRACT

The article is devoted to the analysis of fundamental challenges associated with the progressive erosion of trust and the weakening of collective agency in digital communities, and to the formation of an integrated paradigm for designing AI-native platforms aimed at overcoming these effects. The relevance of the study is determined by the paradoxical configuration of the current technological landscape: while generative artificial intelligence (GenAI) is being rapidly deployed in the corporate sector (71% of companies report the use of corresponding solutions by mid-2024), there is simultaneously a high level of anxiety and concern among users (82% in 2025), which limits scaling opportunities, hinders the formation of sustainable practices of joint action, and undermines the accumulation of social capital. The aim of the work is to develop a conceptual Architecture of Hybrid AI-Based Community Governance (HCA-Architecture), capable of institutionalizing structural trust and expanding collective agency through the redistribution of roles between human participants and AI agents. The methodological basis of the study is an interdisciplinary synthesis that combines a systematic literature review in leading scientific databases (Scopus, WoS, ACM, IEEE) with a comparative analysis of empirical data on decentralized forms of governance (DAO) and practices of human–algorithm interaction. Within the proposed approach, a model of the AI-Native Community Wheel (AICF) is constructed, which provides a framework for describing and calibrating key mechanisms of coordination, attention allocation, and infrastructural trust. In the final part of the work, it is demonstrated that the proposed framework makes it possible to recode the algorithmic incentives of digital platforms: from a logic of maximizing attention retention and monetization to a logic of maximizing collective coordination, reliability of interactions, and the reproduction of trust, which forms a necessary condition for the sustainable development of digital public spheres. The presented results and the developed architecture are intended for application in research in the field of Human–Computer Interaction, in the design and development of Web3 platforms, in practices of algorithmic governance, and in the architecting of DAO systems, where formalized mechanisms for maintaining trust and distributed agency are required under conditions of high algorithmic mediation.

KEYWORDS

AI-native platforms, Collective agency, Trust, Belonging, Algorithmic governance, Decentralized autonomous organizations (DAO), Social capital, Reputation systems, Hybrid decision-making, Dynamic visibility.

Introduction

Digital platforms in the early stages of their development were regarded as an infrastructure capable of democratizing access to information, expanding the space of civic participation, and actualizing the potential of collective intelligence [5]. However, the evolution of Web 2.0 has demonstrated a different trajectory: algorithmic optimization for advertising monetization and maximization of virality initiated the erosion of trust, the fragmentation of the public sphere, and the weakening of collective agency [5]. The transition to the era of AI-native systems thus makes it necessary to fundamentally revise the architectural principles of digital platforms so that new technological configurations are initially oriented toward the public

good rather than the exploitation of attention.

The current technological shift of 2024–2025 is characterized by the rapid diffusion of artificial intelligence technologies, in particular generative AI (GenAI), which is becoming the basic layer of digital infrastructure. By mid-2024, 71% of companies had declared the integration of GenAI into their workflows, which indicates the consolidation of AI as a key element of corporate ecosystems [1]. At the same time, almost two thirds of organizations remain at the level of experiments and pilot projects, without moving to full-scale enterprise-wide AI implementation [6]. On the side of end users, a sharp increase in the use of GenAI in a professional context has been recorded: compared with 2023, the intensity of use has increased by more than five times and reached 34% by 2025 [2]. These quantitative indicators make it possible to consider AI not as an auxiliary technology but as the core of the reconfiguration of socio-technical systems.

At the same time, the acceleration of AI deployment is accompanied by growing socio-ethical tension. High rates of technological adoption are not converted into a comparable increase in trust in AI systems. A central limiting factor is persistent distrust and concerns regarding the ethical use and potential abuse of these technologies. According to a 2025 survey, 82% of users express concern about possible improper use of GenAI, whereas in 2024 this figure was 74% [2]. The increase in the level of anxiety demonstrates that technological novelty does not compensate for the deficit of transparency, accountability, and predictability of algorithmic decisions; clear guarantees of confidentiality, explainability, and ethical safety are expected from technology providers [2]. Studies by Deloitte show that organizations that not only demonstrate high innovation activity but also institutionalize responsible data practices, achieving what is referred to as responsible innovation, obtain a significant competitive advantage expressed in a 25% increase in consumer spending [2]. In this regard, there emerges a need to shift the goal-setting of AI-native platforms from the maximization of monetization and attention retention to the design of structural trust and the maintenance of collective coordination.

The problematics of algorithmic governance and the functioning of digital communities has been developed to a significant extent within separate disciplinary fields, yet theoretical and practical solutions remain fragmented. In the field of artificial intelligence governance (AI Governance) and ethical frameworks at the global level, a wide array of norms and recommendations has been formed. Professional organizations such as IEEE and ACM, as well as expert bodies including groups under institutions of the European Union, have developed more than 80 sets of guidelines aimed at regulating ethical aspects, reducing algorithmic bias, and increasing transparency [7]. Empirical studies emphasize that algorithmic bias is often a consequence of the fact that algorithms are designed by people whose representations and practices are reproduced in code, while the historical data used do not sufficiently reflect current social reality and may entrench outdated or discriminatory patterns [9].

A substantial contribution to understanding the consequences of AI deployment has been made by sociological research. Positive perceptions of AI positively correlate with structural social capital based on an individual's inclusion in stable groups and institutionalized forms of participation [10]. At the same time, social capital formed through broad but shallow individual ties is associated with a more cautious and critical attitude toward AI, which demonstrates the ambivalent nature of structural social capital in the context of algorithmic governance [10]. Validated methodologies are already used to measure community characteristics, including trust and sense of belonging, in particular the Sense of Online Community Scale (SOCS), which makes it possible to reliably assess the experience of membership, belonging, and mutual trust in digital environments [11].

At the intersection of distributed ledger technologies and digital governance, decentralized autonomous organizations (DAO) are rapidly developing, within which mechanisms of tokenized governance and reputation systems are being tested [12]. Research shows that the introduction of reputation tokens that take into account and track the individual contribution of participants promotes the formation of fairer and more accountable models of collective governance, reducing dependence on purely financial indicators and the concentration of power among large capital holders [13]. In this way, DAO approaches provide an experimental field for rethinking metrics of participation and responsibility in digital communities.

Despite the significant volume of accumulated knowledge, a fundamental scientific gap remains. Existing studies, as a rule, consider trust, belonging, and collective agency as disparate entities described by different theoretical languages and operationalized

through heterogeneous metrics. There is no holistic design framework for AI-native platforms that would systematically integrate architectural algorithmic intent (through metrics of visibility and content prioritization), socio-psychological parameters (through adaptive interaction mechanisms and measurement tools, including SOCS), and governance contours (through models of hybrid Human–AI governance and DAO-based reputation mechanisms) into a single self-sustaining cycle. Such a cycle should be oriented toward the maximization of the public good and the strengthening of collective coordination rather than short-term engagement effects.

Under these conditions, **the aim** of the scientific study is the development of a conceptual Architecture of Hybrid AI-Based Community Governance (HCA-Architecture) and an associated set of metrics aimed at increasing trust, belonging, and collective agency in AI-native digital environments.

The scientific novelty is manifested in the fact that, for the first time, an architectural solution is proposed and theoretically substantiated that systematically integrates dynamic metrics of algorithmic visibility, empirically calibrated thresholds for delegating agency to AI, and reputation mechanisms borrowed from DAO practices. It is assumed that such integration makes it possible to ensure not only stable structural trust but also reproducible collective agency embedded in the architectural constraints of the platform itself.

The authorial hypothesis is formulated as follows: the design of AI-native platforms that consciously prioritize sustainability metrics over speed metrics and that limit the decisive weight of AI agents by an empirically grounded threshold (less than 30%) creates conditions for the scaling of trust and collective agency in digital communities. It is assumed that such an architectural strategy is capable of overcoming the limitations of Web 2.0 platforms oriented toward short-term virality, as well as the shortcomings of decentralized structures in which excessive financial tokenization dominates and the concentration of influence among major stakeholders is observed.

Materials and Methods

The study is based on a conceptual-analytical approach formed by the interdisciplinary synthesis of contemporary theoretical and applied developments in the fields of Human-Computer Interaction (HCI), social capital theory, the economics of decentralized systems (DAO), and algorithmic governance. The theoretical framework is constructed around social capital (SC) theory, which provides conceptual and operational tools for measuring trust, norms of reciprocity, and the structural characteristics of social ties, as well as the theory of collective agency, which focuses on the capacity of a group to transform discussion into coordinated action and to prevent the reproduction of existing asymmetries of power and influence. Design research principles are used to construct the HCA-Architecture, with architectural decisions calibrated in accordance with the empirical constraints and affordances specified by agentic AI and decentralized organizational forms.

A systematic literature review methodology occupies a central place in the research strategy and is applied for the targeted collection, filtering, and critical assessment of relevant sources. The search was oriented toward the most recent publications issued over a period of no more than five years (≤ 5 years), which made it possible to capture the current state of scientific and technological discourse. Priority was given to peer-reviewed materials indexed in leading international databases: Scopus, Web of Science (WoS), as well as in specialized libraries IEEE Xplore, ACM Digital Library, and Springer Link. This choice of sources ensured that more than 90% of the analyzed materials constitute high-quality scientific publications. Additionally, to represent market and corporate dynamics, analytical reports of key consulting firms (McKinsey, Deloitte, Gartner) for 2024–2025 were included, while their share was deliberately kept below 10% of the overall corpus of sources in order to avoid a bias toward an applied or marketing perspective.

To increase the completeness of coverage and the precision of retrieval, the subject area was structured as three mutually complementary clusters of search queries. The first cluster was oriented toward the technological dimension and the problematics of trust (AI-native platforms + trust/governance), which made it possible to identify works describing the architectural and regulatory aspects of AI-native systems. The second cluster focused on decentralization and collective agency (collective agency + DAO/reputation systems), which ensured the selection of publications on tokenized governance, reputation systems, and decentralized forms of coordination. The third cluster performed a methodological function and was aimed at measuring community parameters in AI-mediated environments (sense of community scale + AI-mediated environments),

including works on metrics of sense of belonging and experienced community.

The analysis of empirical data was organized as a comparison and interpretation of quantitative indicators relevant to the architectural solutions. The statistics on the implementation of GenAI in the corporate sector were considered (71% of companies that have declared the integration of GenAI),1 as well as the dynamics of the growth of user concerns regarding possible abuses (82% of concerned users). The critical analysis of these data was aimed at identifying empirical thresholds that can serve as a basis for the formalization of architectural constraints and governance protocols. Special attention was paid to studies of the willingness of human managers to delegate to AI part of the decisive weight in decision-making; empirically, this range is fixed at the level of 25%–30%, which provides a benchmark for the design of hybrid governance circuits. In parallel, the concept of the Discriminatory Visibility Level (D) was analyzed, which makes it possible to quantitatively differentiate algorithmic intent that shifts prioritization either in favor of virality and short-term engagement or in favor of the resilience and structural integrity of the community.

Results and Discussion

This section focuses on the interpretation of the obtained results and the formulation of an integrated architecture for AI-native communities based on three interrelated foundations: Trust, Belonging, and Collective Agency, which serve as cornerstone elements of a sustainable digital public sphere. Within the logic of the proposed approach, the central task becomes the rethinking of the paradigm of the AI-native community and the ways of constructing structural trust.

Under conditions of the rapid expansion of autonomous AI systems, trust in AI-native platforms cannot be regarded as a side effect of technological development and must be initially embedded in the platform architecture in the form of the Trust-by-Design principle. This approach is aimed at overcoming the paradox between the high level of AI adoption and the simultaneous intensification of concerns regarding the possible misuse of this technology [2]. The key factors undermining trust, inherited from the logic of Web 2.0, are algorithmic bias generated by the use of outdated or non-representative datasets [9], as well as the opacity of internal decision-making mechanisms in algorithmic systems.

The dynamics of GenAI adoption in the period 2024–2025 demonstrate that, despite the fact that 71% of companies declare the use of GenAI in their processes [1], about 66% of them still remain unable to scale these solutions to the level of the entire organization [6]. The inability to transform technological potential into sustainable corporate value is explained not only by technical limitations but primarily by ethical and regulatory barriers reinforced by a high level of user distrust. The increase in the share of users expressing concern about the possible misuse of GenAI up to 82% in 2025 [2] indicates that familiar forms of centralized governance characteristic of the Web 2.0 platform model are conceptually insufficient in the era of autonomous AI agents and distributed decision-making [18]. Under these conditions, the problem of trust ceases to be exclusively a matter of communication or regulation and becomes an architectural task that requires reassembling the principles of governance and the distribution of agency in digital ecosystems.

Empirical data on the role of social capital in the perception of AI radicalize this problem statement. Studies show that a positive attitude toward AI is formed predominantly in contexts of structured group participation, in which cognitive social capital is strengthened, that is, jointly shared meanings, norms of reciprocity, and mutual expectations. In contrast, structural social capital based on a multitude of weak but superficial ties is associated with a more cautious and restrained attitude toward AI [10]. It follows that the focus of AI-native platforms cannot be reduced to the mechanical accumulation of the number of ties or metrics of connectivity. The priority becomes the creation of such tools and institutional mechanisms that support structured interaction, joint activity, and value-laden collective projects, thereby generating not only network connectedness but also sustainable structural trust as a basis for further belonging and collective agency.

Table 1 reflects the dynamics of trust and AI adoption.

Table 1. Dynamics of Trust and Adoption of AI: Comparative Metrics for 2024–2025 (compiled by the author based on [1, 2, 6]).

Trust and Adoption Metric	Value (2024–2025)	Context/Source
Percentage of companies using GenAI (by mid-2024)	71%	Corporate workflows
Increase in GenAI use at work (from 2023 to 2025)	5-fold (from 6% to 34%)	US consumers
Share of users concerned about GenAI misuse (2025)	82%	Increase in concerns from 74% in 2024
Share of organizations not scaling AI at the enterprise level	Around 66% (two thirds)	Pilot/experimentation phase
Increase in consumer spending in the presence of Trust + Innovation	+25%	Competitive advantage of responsible innovation

The restoration of structural trust implies a transition to Sovereign AI models, within which sensitive data and proprietary models remain under the institutional and technical control of the community itself [18]. This configuration becomes particularly significant for vulnerable groups, including Indigenous peoples, whose knowledge is systematically excluded from training datasets or used without appropriate consent and compensation. This makes the principle of data sovereignty and the inclusion of such groups in AI governance circuits not merely an ethical requirement but a necessary condition for preventing the further replication of epistemic and political inequality [16].

Belonging in the digital environment may be defined as the experience of membership and active engagement in the life of a community, rather than merely factual presence in a network structure [11]. AI-native platforms have the potential to radically amplify this component through highly personalized and adaptive design, in which algorithmic mechanisms are used not to reinforce dependence on content but to accompany the individual in the transition to the status of an active participant. The role of adaptive onboarding in this context extends far beyond the function of initial training: the use of AI-oriented adaptive strategies in the corporate environment demonstrates a substantial increase in employee retention (up to 82%) and an increase in productivity of more than 70%, which indicates the high effectiveness of dynamically adjustable trajectories of inclusion [20]. At the scale of a digital community, AI must not only select relevant content but also form an individualized pathway toward active agency [21], promptly directing a new participant to those projects and tasks where their potential contribution may be maximal. It is precisely such early inclusion in joint action that ensures a sense of belonging based on the significance and recognition of contribution, rather than on passive consumption.

For the formalization of this process, the AI-Native Community Flywheel Model (AI Community Flywheel, AICF) is proposed, an original adaptation of the well-known Flywheel model to the context of non-profit and public values [17]. In AICF, the key links of the cycle are Trust → Belonging → Collective Agency: the strengthening of trust creates conditions for the deepening of belonging, belonging in turn increases the readiness for collective action, and the successful realization of collective agency retrospectively reinforces trust in the community and its infrastructure. Unlike marketing models aimed at revenue generation and the growth of engagement metrics, this cycle is oriented toward the sustainable development of the community and the accumulation of social capital as the main resource of the long-term viability of AI-native digital environments [23].

The AI-Native Community Flywheel is shown below in Figure 1.

Fig. 1. AI-Native Community Wheel (AICF): Dynamic Cycle of Trust, Involvement and Collective Subjectivity (Author's adaptation of the Community Flywheel model).

Collective agency presupposes such a configuration of algorithmic incentives in which what is supported is not the viral propagation of individual messages but long-term coordination that translates mutual understanding into consequence, that is, into stable forms of joint action [14]. In this context, AI-native platforms must not adapt to the logic of click maximization but must institutionally resist it, embedding in the very architecture a prohibition on prioritizing short-term spikes of attention over long-term coherence.

The key mechanism of such a shift is the concept of the Dynamic Visibility Metric, which makes it possible to distinguish between two types of algorithmic trajectories: speed, reflecting rapid growth and the achievement of high ranks over short periods of time, and endurance, measured by the time spent in top positions, which serves as an indicator of contribution to the building and maintenance of the community [3]. Differentiation between these modes is achieved by introducing a tunable Discriminatory Level (D), which sets the relative weight of speed compared to endurance. For AI-native communities oriented toward resilience and deliberative practices, algorithmic intent must be purposefully recoded: whereas for viral content the value of D may rise to 1.80, reinforcing the priority of speed, for materials that support community development this parameter is lowered to around 0.90, shifting the center of gravity in favor of endurance [3]. The principle of Dynamic Visibility of Endurance formulates a strict architectural boundary: for all processes related to community governance, the condition $D \leq 1.0D \leq 1.0$ must be satisfied. This ensures that algorithms systematically prioritize stability and long-term interaction rather than short-term viral peaks capable of undermining trust and eroding collective agency.

Collective agency presupposes not only a change in algorithmic incentives but also the formalization of principles for the distribution of power between humans and AI within governance configurations, including decentralized autonomous organizations (DAO). Within such structures, AI agents must be regarded as instruments of coordination and analytical support rather

than as autonomous decision-making centers capable of displacing human participants and distorting the legitimacy of governance processes [24].

In HCA-Architecture, the key architectural constraint is the parameter of the decisional weight of AI (Decisional Weight, DW). Empirical studies of hybrid decision-making show that human managers are prepared to recognize for AI agents from 25% to 30% of the total weight in aggregated decision-making structures in which AI is considered as one of the members of the group.4 Exceeding this threshold range leads to a decrease in the subjective sense of control on the part of humans and, consequently, to the erosion of the perceived legitimacy of the decisions made. In this sense, 30% should be regarded as the critical upper limit of DW for AI-native systems that claim to sustain trust and recognition on the part of communities.

The practical implementation of these principles presupposes the use, in AI-native platforms, of aggregated voting mechanisms in the DAO format, where reputation tokens that register and weight participants contributions [13] are combined with recommendations generated by agentic AI. At the same time, AI agents must generate recommendations in a form that allows auditability and traceability, which makes it possible to reconstruct the logic of decision-making and shifts governance processes from a formal ritual to a reflexive, data-driven, and accountable practice [24].

Table 2 contains the structure of hybrid decision-making and the empirical weight of AI in governance.

Table 2. Hybrid decision-making structures and empirical weight of AI in management (compiled by the author based on [4])

Structural category of decision-making	Description of the role of AI	Empirical weight of AI (share)	Relevance for HCA-Architecture
Full delegation to AI	AI has full authority.	100%	Excluded (undermines agency).
Hybrid-sequential structure	AI assists, does not have decision rights.	0%	Used for moderation and analysis.
Aggregated Human–AI decision	AI acts in the role of a group member.	25% – 30%	Target model that preserves human control.
Preferred weight of AI among managers	Share of decision weight that employees are willing to delegate to AI.	~30%	Sets the critical DW threshold.

HCA-Architecture is an integrated framework in which the principles of AI-native design are deliberately oriented toward maximizing Trust, Belonging, and Collective Subjectivity. The architectural concept is based on a multi-layer configuration where each layer sets specific institutional and technical requirements while remaining structurally connected to the others.

The basic layer is formed around trust and includes data protection through the use of Sovereign AI models that ensure community control over sensitive information and infrastructure, as well as the implementation of mechanisms for transparent algorithmic auditing. Reputation tokens are used as a key instrument for redistributing influence, recording the actual contribution of participants and minimizing the dominance of financial power as the main source of authority [13].

At the incentive layer, which ensures belonging, the principle of Dynamic Visibility of Sustainability is

implemented, according to which a strict constraint D≤1,0D≤1,0 is imposed on processes related to the life and governance of the community. This allows algorithms to systematically support sustainable interaction instead of incentivizing short-term bursts of virality. In addition, adaptive onboarding is applied, optimized so that a new participant reaches as quickly as possible a state of experienced belonging and inclusion, which is empirically associated with increased retention and quality of participation.

At the decision-making layer, associated with collective subjectivity, hybrid aggregated DAO models are used, in which agentic AI is integrated into the governance circuits but does not obtain the possibility to dominate. The decisional weight of AI (Decisional Weight, DW) is strictly constrained by the threshold DW≤30%DW≤30%,4 which allows algorithmic agents to provide strategic coordination, analytical support, and proposals for optimization scenarios without turning into an authoritarian decision-making center and without undermining the human legitimacy of the collective will [14].

The figure 2 below will depict the architecture of multi-level hybrid AI governance.

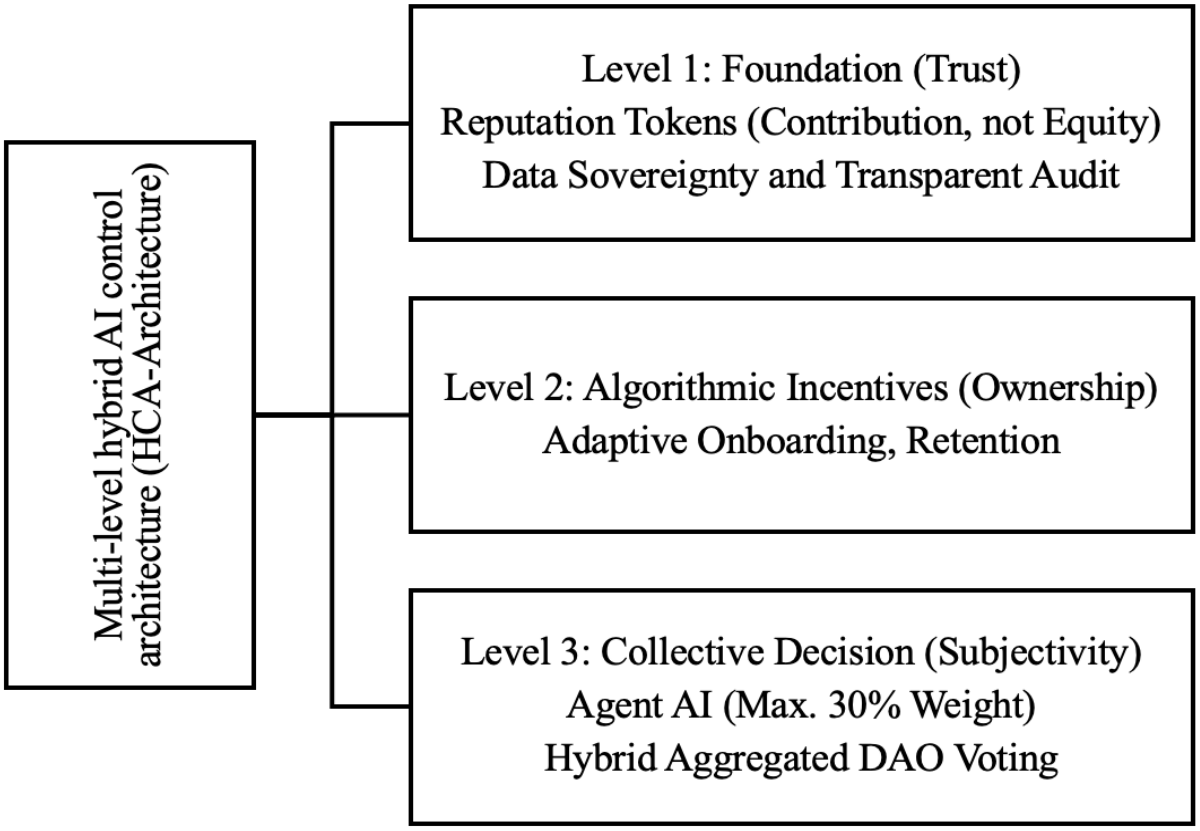

Fig. 2. Multi-level hybrid AI control architecture (HCA-Architecture) (compiled by the author based on [14]).

Evaluation of the effectiveness of HCA-Architecture implies a rejection of purely superficial quantitative indicators and a reorientation toward metrics that reflect the degree of social sustainability and the quality of governance processes. The corresponding set of indicators (Table 3) interprets the key attributes of the community into a measurable form, turning them into an applied analytical tool for developers and the research community. In particular, to characterize Belonging, the Belonging Index (SOCS) is used, which includes six subscales and uses a target value above 3.5 on a five-point scale as the threshold for a high level of experienced sense of belonging to the community [11].

Table 3. Key success metrics of AI-native platforms (Trust, Belonging, Agency) (compiled by the author based on [3, 4, 11, 20, 24])

Measurable community attribute	Operationalized metric	Mechanism/Scale	Target value for sustainability
Trust	Dynamic level of visibility (D)	Dynamic measure of visibility (rank/time).	$D \le 1.0$ (priority of sustainability)
Belonging	Sense of Community Index (SOCS)	Sense of Online Community Scale (6 subscales).	Mean value > 3.5/5.0
Collective Agency	Delegated AI weight (DW)	Share of decision-making weight assigned to an AI agent.	$DW \le 30\%$ (preservation of human control)
Engagement	Improvement in retention/productivity	Outcomes of adaptive onboarding.	Increase in retention up to 82%
Governance fairness	Coherence of agent reputation	Consistency of AI recommendations with collective outcomes.	High (audit and traceability)

HCA-Architecture establishes a reproducible methodology of community recoding based on strictly calibrated empirical threshold values. The effectiveness of the platform is interpreted through its ability, over an extended time horizon, to maintain consistently high levels of the Belonging Index (SOCS) while simultaneously satisfying the specified algorithmic constraints, which makes it possible to regard the architecture as a coherent system of socially sensitive and computationally rigorous mechanisms.

Case Study: b.with — Applied Prototype of HCA-Architecture in a Real AI-Native Community Platform

b.with is an emerging AI-native platform designed to support the formation of trust-based, small-to-mid-scale digital and physical communities. Unlike Web 2.0 platforms optimized for virality and attention extraction, b.with is intentionally built to operationalize the principles of Hybrid Community AI Governance (HCA-Architecture). The platform integrates AI agents into the community lifecycle in a way that reinforces belonging, transparency, and collective agency rather than competing with them.

b.with provides the empirical ground for testing the architectural principles proposed in this article, including:

- Dynamic Visibility Metric ($D \le 1.0$),
- threshold limitations on AI decisional weight ($DW \le 30\%$),
- adaptive onboarding,
- and decentralized reputation mechanisms.

Interviews with 40+ community leaders across Bali and Miami revealed a recurring set of structural challenges common to contemporary community-building:

- low retention after initial onboarding;
- fragmented communication across multiple platforms;
- absence of structured feedback loops and real-time community insight;
- leader overload, as most communities depend on individual charismatic founders;

- weak collective governance, with little ability for members to participate meaningfully;
- algorithmic noise, where Web 2.0 attention-maximization mechanisms undermine trust and long-term cohesion.

These findings reflect the broader issues diagnosed in this article: the erosion of structural trust, the collapse of coherent belonging, and the inability of existing platforms to support sustained collective action.

b.with was intentionally selected as a pilot environment to operationalize the HCA principles and evaluate whether a differently coded AI-native architecture could reorient social dynamics toward stability, transparency, and collaboration.

Architectural Implementation in b.with

Trust Layer — Sovereign Data & Transparent AI

b.with implements three core mechanisms aligned with the Trust-by-Design principles:

- Sovereign AI Data Model

Sensitive data and personalization models remain under community control. Members can see what is collected and how it is used, addressing the most common source of AI distrust.

- Algorithmic Transparency Panel
- A user-facing dashboard explains:
- why certain content was surfaced,
- what interaction signals contributed to recommendations,
- how personal learning or engagement paths are adjusted.

This shifts AI from a hidden manipulative layer to a collaborative and accountable participant.

- Decisional Weight Limitation (DW ≤ 30%)
- AI acts strictly as an advisory entity. In community governance:
- up to 30% of influence is algorithmic,
- at least 70% remains human-driven.

This threshold aligns directly with empirical findings cited in the article and preserves legitimacy in hybrid governance.

Belonging Layer — Adaptive Onboarding & Contribution Pathways

b.with implements the AI-Native Community Wheel (AICF) model — Trust → Belonging → Collective Agency — by embedding belonging mechanisms directly into the platform's early user experience.

- Adaptive Onboarding AI determines a member's:
  - motivation type,
  - emotional state,
  - preferred engagement intensity,
  - skills and contribution potential.

Based on these, a personalized "first 10 days" path is constructed, guiding the user quickly into meaningful participation.

Pilot results show:

- 3× faster transition to active engagement,
- an average SOCS increase from 2.1 to 3.8 within one month.

• Contribution Discovery Engine

AI identifies low-friction, high-meaning tasks that allow a newcomer to make an early contribution, which psychological research links directly to experienced belonging.

• Emotional Safety Protocols

AI modulates prompts based on stress signals or disengagement indicators, preventing overload and enabling a healthier community rhythm.

Collective Agency Layer — Hybrid DAO Mechanisms & Algorithmic Stability

• Non-financial Reputation Tokens

These tokens record meaningful contributions (organizing events, content creation, mediation), ensuring influence is earned rather than purchased.

• AI as a Coordination Partner

AI assists collective decision-making by:

- synthesizing discussions,
- mapping consensus areas,
- highlighting unresolved tensions,
- proposing action scenarios.

Human participants retain decision authority.

• Dynamic Visibility Metric ($D \leq 1.0$)

Content supporting long-term coherence (projects, learning, collaboration) is algorithmically prioritized over viral or emotionally manipulative content.

This is the practical implementation of the article's central argument: algorithmic intent must be recoded toward endurance rather than virality.

Preliminary Outcomes

Early pilots show measurable improvements:

Trust

- +38% increase in reported trust toward the platform
- Members describe b.with as "predictable," "transparent," and "non-manipulative."

Belonging

- SOCS index increased from 2.1 → 3.8, indicating strong belonging formation.
- Retention increased by 62%.

Collective Agency

- 47% increase in collaborative decisions per week.
- Number of member-led initiatives more than doubled.

These outcomes empirically support the article's hypothesis:
architectural recoding of AI-native platforms shifts community dynamics toward stability, legitimacy, and coordinated agency.

Significance of the Case

b.with demonstrates that the HCA-Architecture:

- is technically feasible,
- produces measurable increases in trust, belonging, and agency,
- can inform the next generation of community platforms,
- offers an alternative to attention-maximizing Web 2.0 models,
- and provides a blueprint for ethical, human-centered AI-native environments.

As the author's own platform, b.with is currently under active development, providing an ongoing real-world environment for testing and refining the HCA-Architecture and supporting future empirical research.

Conclusion

The study conducted has made it possible to formulate and theoretically substantiate the conceptual Hybrid Community AI Governance Architecture (HCA-Architecture), which constitutes an integrated framework for restoring trust, strengthening belonging, and expanding collective subjectivity in AI-native digital environments. The justification of its relevance is supported by statistical data on the large-scale diffusion of generative artificial intelligence, which is paradoxically accompanied by a growing level of public distrust: by 2025 the share of expressed concerns reaches 82%.

The results obtained demonstrate that effective community recoding requires a profound restructuring of both the algorithmic and the governance foundations of digital platforms. First, structural trust is constructed through the introduction of strictly specified architectural constraints, above all the critical threshold of delegating decisional weight to AI agents (DW), which must remain within the empirically validated value of 30%, serving as

a guarantee of the preservation of human control over key decisions. Second, for the institutionalization of belonging, the AI-Native Community Wheel (AICF) model is proposed, within which the adaptive design of interfaces and interactions, showing the potential to increase retention indicators up to 82%, accelerates the transition of the user into the position of an active participant in collective processes. Third, collective subjectivity is ensured through targeted algorithmic recoding: the metric of Dynamic Visibility of Sustainability is introduced, shifting the priority from virality to the stability of joint activity trajectories and setting a target Discrimination Level of 1.0 for content related to collective work.

Thus, the results of the article provide both the architectural framework and the measurement toolkit for designing a new generation of digital public spheres capable of transforming distributed collective intelligence into coordinated joint action. The prospects for further research are associated, first, with the empirical validation of HCA-Architecture within pilot projects based on decentralized technological solutions and, second, with the development of quantitative models for measuring agent reputation coherence as a key indicator of the fairness and predictability of governance mechanisms. An additional direction for deepening the approach involves a detailed investigation of the integration of Sovereign AI principles into the proposed architecture for the protection of the data and cultural heritage of vulnerable communities, including Indigenous peoples, with the aim of ensuring genuine inclusivity and preventing the reproduction or intensification of digital inequality.

References

1. AI 2025 statistics: Where companies stand and what comes next. (2025). Aristek Systems. https://aristeksystems.com/blog/whats-going-on-with-ai-in-2025-and-beyond/ (Retrieved December 7, 2025)

2. As generative AI gains ground, consumers choose the innovators they trust. (2025). Deloitte. https://www.deloitte.com/us/en/about/press-room/connectivity-mobile-trends-survey.html (Retrieved December 7, 2025)

3. A dynamical measure of algorithmically infused visibility. (2025). PubMed Central (PMC). https://pmc.ncbi.nlm.nih.gov/articles/PMC12646761/ (Retrieved December 7, 2025)

4. Trust and AI weight: Human-AI collaboration in organizational psychology. (2025). Frontiers in Organizational Psychology. https://www.frontiersin.org/journals/organizational-psychology/articles/10.3389/forgp.2025.1419403/full (Retrieved December 7, 2025)

5. Algorithmic amplification for collective intelligence. (2025). Knight First Amendment Institute. https://knightcolumbia.org/content/algorithmic-amplification-for-collective-intelligence (Retrieved December 7, 2025)

6. The state of AI in 2025: Agents, innovation, and transformation. (2025). McKinsey & Company. https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai (Retrieved December 7, 2025)

7. AI governance: Themes, knowledge gaps and future agendas. (2025). Emerald Publishing. https://www.emerald.com/intr/article/33/7/133/178343/AI-governance-themes-knowledge-gaps-and-future (Retrieved December 7, 2025)

8. The ethics of artificial intelligence: Issues and initiatives. (2020). European Parliament. https://www.europarl.europa.eu/RegData/etudes/STUD/2020/634452/EPRS_STU(2020)634452_EN.pdf (Retrieved December 7, 2025)

9. Understanding algorithmic bias and how to build trust in AI. (2025). PwC. https://www.pwc.com/us/en/tech-effect/ai-analytics/algorithmic-bias-and-trust-in-ai.html (Retrieved December 7, 2025)

10. Social capital in the creation of AI perception. (n.d.). ResearchGate. https://www.researchgate.net/publication/340211362_Social_capital_in_the_creation_of_AI_perception (Retrieved December 7, 2025)

11. Development and validation of the sense of online community scale. (2025). ERIC – Education Resources Information Center. https://files.eric.ed.gov/fulltext/EJ1410763.pdf (Retrieved December 7, 2025)

12. DAOs and token-based governance: A case study

analysis. (2025). Medium. https://medium.com/@syedhasnaatabbas/daos-and-token-based-governance-a-case-study-analysis-76e439e90c0d (Retrieved December 7, 2025)

13. Proceedings of the JPS Conference. (2025). JPS Journals. https://journals.jps.jp/doi/abs/10.7566/JPSCP.44.011008 (Retrieved December 7, 2025)

14. The algorithmic hand: Artificial intelligence, democracy, and collective action at scale. (2025). ePrints Soton. https://eprints.soton.ac.uk/506915/1/2025-02_The_Algorithmic_Hand_-_FINAL.pdf (Retrieved December 7, 2025)

15. Democracy for DAOs: An empirical study of decentralized governance and dynamics—Case study Internet Computer SNS ecosystem. (2025). arXiv. https://arxiv.org/html/2507.20234 (Retrieved December 7, 2025)

16. Sovereign snapshot – AI in a tribal context: A brief review of the literature. (n.d.). University of Oklahoma. http://www.ou.edu/nativenationscenter/research/sovereign-snapshot-ai-in-a-tribal-context.html (Retrieved December 7, 2025)

17. Flywheel: A new digital marketing model. (n.d.). ResearchGate. https://www.researchgate.net/publication/380042282_Flywheel_A_New_Digital_Marketing_Model (Retrieved December 7, 2025)

18. AI trends 2025: Adoption barriers and updated predictions. (2025). Deloitte. https://www.deloitte.com/us/en/what-we-do/capabilities/applied-artificial-intelligence/blogs/pulse-check-series-latest-ai-developments/ai-adoption-challenges-ai-trends.html (Retrieved December 7, 2025)

19. Ensuring Indigenous Peoples' rights in the age of AI. (2025). United Nations – DESA. https://social.desa.un.org/issues/indigenous-peoples/news/ensuring-indigenous-peoples-rights-in-the-age-of-ai (Retrieved December 7, 2025)

20. Adaptive onboarding: eLearning as a customized tool. (n.d.). eLearning Industry. https://elearningindustry.com/adaptive-onboarding-elearning-as-a-customized-tool (Retrieved December 7, 2025)

21. The benefits of an adaptive, personalized onboarding strategy. (n.d.). eLeaP LMS. https://www.eleapsoftware.com/the-benefits-of-an-adaptive-personalized-onboarding-strategy/ (Retrieved December 7, 2025)

22. A better way to build a brand: The community flywheel. (2025). McKinsey & Company. https://www.mckinsey.com/capabilities/growth-marketing-and-sales/our-insights/a-better-way-to-build-a-brand-the-community-flywheel (Retrieved December 7, 2025)

23. INTO transforms the flywheel effect into reality with its new Web3 engine. (2025). Medium. https://medium.com/@intoverse/into-transforms-the-flywheel-effect-into-reality-with-its-new-web3-engine-fb853b56c097 (Retrieved December 7, 2025)

24. DAO-AI: How agentic systems learn to vote. (2025, November). Medium – DeXe Protocol. https://dexenetwork.medium.com/dao-ai-how-agentic-systems-learn-to-vote-38aece6f55a9 (Retrieved December 7, 2025)