Gradual Voluntary Participation: A Framework for Participatory AI Governance in Journalism

MATILDE BARBINI*, École Polytechnique Fédérale de Lausanne (EPFL), Switzerland
STEFANO SORRENTINO*, École Polytechnique Fédérale de Lausanne (EPFL), Switzerland
DANIEL GATICA-PEREZ, Idiap Research Institute, Switzerland and École Polytechnique Fédérale de Lausanne (EPFL), Switzerland

The integration of AI into journalism challenges participatory design (PD), particularly with respect to stakeholder influence, workplace perceptions, and organizational dynamics. Traditional PD assumes that users can shape technologies, yet AI systems resist influence due to opaque data, fixed architectures, and inaccessible objectives. Through interviews with 10 journalists, we identify the *perception gap*, showing that trust in AI depends on perceived agency within workplace participatory workflows. Informed by these findings, we introduce the Gradual Voluntary Participation (GVP) framework in journalism and its five core principles, reconceptualizing participation as a gradual and voluntary process that can be operationalized at the newsroom level, beyond fixed workshops or one-time preference-elicitation campaigns. Addressing epistemic burdens, participatory ceilings, and performative consultations, GVP treats gradualism and voluntariness as design dimensions that shape perception, legitimacy, and ownership. Moving beyond unidimensional ladder metaphors and adopting a bidimensional matrix structure, the framework maps stakeholders across depth and scope, offering a new model for local participatory AI governance that balances technological transformation with stakeholder empowerment in rapidly evolving hybrid workplaces.

CCS Concepts: • **Human-centered computing**; • **Computing methodologies** → **Artificial intelligence**; • **Social and professional topics** → **Professional topics**;

Additional Key Words and Phrases: AI, Large Language Models, journalism, participatory design, gradual voluntary participation, perception gap, stakeholder engagement, epistemic burden

ACM Reference Format:
Matilde Barbini, Stefano Sorrentino, and Daniel Gatica-Perez. 2027. Gradual Voluntary Participation: A Framework for Participatory AI Governance in Journalism. *Proc. ACM Hum.-Comput. Interact.* 11, 1, Article GROUP21 (January 2027), 48 pages. https://doi.org/10.1145/3859407

1 Introduction

Newsrooms have long been early adopters of technological change, and the integration of large language models (LLMs) into journalistic practice has opened new possibilities for experimentation [93]. Yet, AI adoption in journalism raises organizational questions about who can shape these systems, under what conditions, and with what consequences for editorial authority, professional accountability, and newsroom coordination [30]. As AI enters high-stakes professional settings, participation becomes not only a matter of consultation or usability, but of how groups negotiate

*Both authors contributed equally to the paper.

Authors' Contact Information: Matilde Barbini, matilde.barbini@epfl.ch, École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland; Stefano Sorrentino, stefano.sorrentino@epfl.ch, École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland; Daniel Gatica-Perez, gatica@idiap.ch, Idiap Research Institute, Martigny, Switzerland and École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland.



This work is licensed under a Creative Commons Attribution 4.0 International License.
© 2027 Copyright held by the owner/author(s).
ACM 2573-0142/2027/1-ARTGROUP21
https://doi.org/10.1145/3859407

influence over technologies that affect their work while remaining only partially open to local intervention. In journalism, however, participation is related to editorial choices, professional accountability, and public-facing responsibility. Unlike other professional domains, newsroom actors operate under norms of verification, attribution, and audience trust, which make the question of who can influence AI systems both epistemic and civic, beyond purely organizational.

This creates a challenge for participatory design (PD), which has historically emphasized stakeholders' ability to shape the technologies that affect them [6, 7, 52, 63, 98]. With contemporary AI systems, many consequential decisions remain upstream, with participation unevenly distributed across implementation layers: training data, model architectures, optimization objectives, and update cycles are often inaccessible to the professionals asked to use these tools [99].

Existing work has shown that participation in AI can become tokenistic, burdensome, or disconnected from actual power [9, 92], with journalism research likewise documenting concerns around trust, autonomy, and ownership in AI adoption [101, 102]. How newsroom actors evaluate the legitimacy of participation itself is less clear, i.e., how they distinguish between forms of involvement that support editorial agency and those that merely shift responsibility onto already constrained professionals. This question is especially pressing in resource-limited news organizations, where participation competes with deadlines, hierarchies, and dependence on external vendors [15, 88, 89].

In this paper, we examine how journalists and editors in an independent, Italian newsroom understand and evaluate participation in AI adoption, focusing specifically on generative AI (GenAI) and large language models (LLMs). We use 'AI' throughout to refer mainly to these generative technologies unless otherwise noted. Drawing on semi-structured interviews with 10 newsroom actors, we show that participants did not primarily evaluate AI integration through performance or efficiency claims, but through the conditions under which AI would enter newsroom practice: whether editorial sovereignty was preserved, responsibilities remained accountable, systems became intelligible in use, and participation was recognized as legitimate work rather than additional burden. We describe this empirical pattern as a *perception gap*: a mismatch between formal and nominal means of participation in AI adoption and the organizational conditions under which participation is experienced as meaningful.

From these findings, we develop *Gradual Voluntary Participation* (GVP), a framework for participatory AI governance in journalism. GVP conceptualizes participation across two dimensions: *depth*, the degree of influence actors have over AI-related decisions and arrangements, and *scope*, the arenas in which participation takes place, from local tool use to broader organizational and external governance contexts. It adopts a matrix form to make participation visually concrete while framing it as conditional, non-linear, and shaped by the practical realities of newsroom work. The framework is articulated through five principles derived from our data: recognition without coercion, continuous value alignment, contextual knowledge scaffolding, sovereignty preservation, and transparent accountability.

Our contribution is threefold. First, we provide an empirical account of how newsroom actors evaluate participation in AI adoption through questions of agency, accountability, and organizational credibility rather than technical performance alone. Second, we introduce GVP as a matrix-based framework that moves beyond static, unidimensional "ladder-like" models of participation [5] by conceptualizing participation across organizational depth and collective scope in AI-integrated newsroom settings. Third, we articulate a set of principles for supporting participatory AI governance under conditions of workload pressure, role asymmetry, and limited leverage over external AI infrastructures.

By reframing AI adoption as a problem of organizational participation and coordination, this paper contributes to ongoing discussions in participatory design and human-AI interaction about

how professional groups can exercise meaningful influence over opaque and externally governed systems.

2 Related Work: Participation, Power, and Organizational Reconfiguration in AI-Mediated Journalism

Participatory design (PD) has long treated the expansion of stakeholder *voice* and *power* as a democratic aspiration, especially where technologies reshape work practices and decision-making relations [6, 23, 24, 28]. Yet participatory outcomes are difficult to evaluate in heterogeneous organizational settings, particularly when what is being designed is not a bounded artifact but an evolving sociotechnical arrangement embedded in ongoing work [13, 28, 99]. These difficulties intensify in AI contexts, where retraining, adaptation, and changing data inputs make participation hard to locate at any single design moment [11, 13]. Participation therefore cannot be assessed only by whether stakeholders are consulted, but also by who can influence implementation, which decisions remain contestable, and how authority is distributed over time [28, 43]. Recent work on "participation in the age of foundation models" [99] conceptualizes AI systems as stratified across "foundation", "subfloor", and "surface" layers, showing how participatory interventions tend to cluster at the interface level, while excluding stakeholders from upstream decisions about data, optimization, and infrastructure. In journalism, this stratification maps directly onto newsroom-tool relationships, where journalists primarily engage with interfaces while relying on external actors for the underlying system design and development, reinforcing structural limits to meaningful participation [99].

This problem is especially acute in newsrooms, framed as hierarchical organizations in which editors and publishers set policies and control resources, while reporters exercise uneven and status-dependent autonomy [15, 81]. Editorial independence, while central to journalistic identity, is therefore conditional on funding arrangements, infrastructural access, organizational relationships, and technological dependencies [36, 79, 88, 90]. Participation in journalistic AI should not be framed simply as stakeholder inclusion in design, but as negotiating influence within settings already structured by unequal decision rights, professional norms, and external constraints [24, 25, 102].

Journalism as an organizational site of power. Both recent and foundational scholarship shows that in complex organizations such as newsrooms, power operates through direct control, agenda-setting, preference-shaping, and the normalization of certain assumptions and discourses over others [21, 26, 37, 59, 94]. This matters for AI because design and deployment do not introduce tools into neutral settings; they can also encode, formalize, and stabilize existing institutional logics [51, 74, 76]. Workflow technologies act not only as coordination mechanisms but also as organizational *accounting devices* that render work visible, allocable, and subject to managerial oversight [34]. In journalism, where editorial authority is already contested across managerial, commercial, technical, and professional lines [15, 36, 81], questions of participation are inseparable from questions of who defines the goals of AI adoption, whose expertise counts as legitimate, and which values become embedded in systems and workflows [43, 50, 75, 102]. Such power dynamics are often reinforced by a pervasive lack of technical knowledge or self-confidence, with journalists reportedly identifying as "not data people" [20]. This creates an internal power imbalance and epistemic resistance where journalists experience difficulties navigating technical tasks and delegating authority and institutional power to technologists [20, 107].

AI as a reconfiguration of intra-newsroom group dynamics. AI adoption redistributes power and editorial authority within the newsroom: responsibilities once associated with editors and reporters may shift toward managers, technical specialists, and algorithmic systems, while journalists become more tightly coupled to digitized workflows that can narrow their discretion [32, 49, 64, 94, 101].

Yet the effects of algorithmic governance are not uniformly coercive. Ethnographic work on platform-based algorithmic management suggests that workers may retain significant discretion even under opaque control, depending on organizational context and the specific configuration of human-algorithmic authority [65]. At the same time, separation between editorial and engineering units fosters knowledge silos and weak cross-departmental coordination, making it harder for non-technical actors to contest system behavior or locate accountability [22, 32, 33]. Research on groupware adoption in distributed organizations similarly show how conflicting views of collaboration, incompatible standards, and uneven access to resources shape which groups can appropriate shared tools and which remain excluded [72]. This primary technical barrier to collaboration can be understood as a "double black box" problem, representing a dual opacity in which editors cannot see a journalist's AI workflows (e.g. prompts), and journalists cannot understand AI's "reasoning" [108]. Individual, private AI use is often framed by journalists as a "backstage behavior", reportedly associated with a sense of "professional shame" [108]. Such a hidden use can create chains of mutual suspicion among teammates, where editors must check all drafts as if they were machine-generated to ensure the provenance of the work [108]. These dynamics also produce hybrid roles that may open new possibilities for collaboration while unsettling existing boundaries of expertise and control [2, 69, 101]. Importantly, local newsrooms may end up relying on a single tech-enthusiast journalist, creating a precarious environment in which when this individual leaves, the newsroom's capacity for AI innovation collapses [86]. Participation in AI-mediated journalism is thus fundamentally a group problem, unfolding across differently positioned actors whose roles, authority, and responsibilities are being actively reconfigured [69, 101, 104].

Participation under inter-organizational dependency. AI adoption in journalism is conditioned not only by newsroom dynamics but also by dependencies beyond the newsroom itself. AI adoption is uneven across organizations: large publishers are better able to build internal infrastructures and specialized teams, whereas smaller and local outlets are more often pushed toward commoditized tools, external funding schemes, and vendor-defined solutions [82, 88, 90, 94]. These asymmetries generate lock-in effects and risks of infrastructural capture as platform companies extend their influence from distribution into core stages of news production [42, 80, 88]. This scenario is further exacerbated by a *Matthew effect*, in which local newsrooms' cultures and path dependencies limit their capacity to innovate, alongside strict regulatory and funding structures, that in contrast tend to protect larger, established newsrooms, leaving local newsrooms "behind" and lacking the financial cushion to risk experimentation and to pivot to new solutions [38, 86]. Under such conditions, local participation does not necessarily translate into meaningful control over the most consequential layers of the AI stack [25, 41, 47, 88]. Research on large-scale scientific collaborations offers a useful analogue: actors from lesser-resourced institutions often faced structurally limited access to shared infrastructure and decision-making even when formally included [70]. This pattern is particularly relevant for journalism, where smaller organizations may adopt AI as a survival strategy even when doing so requires partly relinquishing editorial control, compromising transparency, or accepting limited bargaining power over external providers [79, 82, 89, 90]. Accordingly, participatory outcomes must be evaluated not only within the newsroom, but also in relation to broader platform and vendor ecosystems [36, 42, 80, 88, 90]. In journalism, this interdependence has been conceptualized by Tseng et al. [102] through participatory AI approaches attentive to ownership, control, and infrastructural dependency across micro, meso, and macro levels of the news ecosystem [102]. Similarly, participation in the age of foundation AI models, as per Suresh et al. [99], extends beyond internal newsroom inclusion toward ongoing coordination with external technology providers, where journalists contribute domain expertise, articulate professional constraints, and influence

how tools evolve in practice. Under these conditions, participation becomes inherently layered and cross-organizational, requiring alignment across editorial, technical, and commercial actors.

Participatory responses and their limits. Against these redistributions of authority, participatory approaches are often proposed as ways to reclaim agency, reduce knowledge silos, and align AI systems with journalistic values [24, 25, 28, 102]. Existing work argues that involving journalists and other affected actors in co-planning and co-design can support mutual learning, context-specific implementation, and collective ownership over newsroom change [24, 43, 45, 60, 102]. Participatory approaches have also been framed as ways of creating shared spaces among journalists, developers, managers, and communities through bridge-building roles and structured collaboration around governance, evaluation, and oversight [50, 62, 75, 107]. Such "AI news zones", as per Xiao et al. [108], can be understood as safe spaces for experimentation: dedicated, low-risk editorial spaces that allow for "just-in-time" learning and prototypes without triggering professional judgment. Similar shared spaces would help making the previously mentioned "backstage behaviour" more visible, converting it into legitimate and auditable labor [108]. Such visibility could ultimately be expanded to the public, addressing the rise in public vigilance and suspicion toward AI [108]. At the same time, such efforts remain limited in practice [28, 92, 99]. In journalism, many proposals are still exploratory, unevenly institutionalized, and vulnerable to tokenistic or performative adoption [57, 102, 107]. More critically, participatory AI scholarship warns of a *participatory ceiling*, in which stakeholders may shape downstream implementation while remaining excluded from foundational decisions about adoption, optimization, and infrastructure [99]. In such cases, participation may become an *epistemic burden* rather than an empowering process [9]. In journalism, this ceiling is particularly visible in the misalignment between newsroom expectations and vendor-driven tool development: while journalists seek influence over editorial standards, alignment, verification practices, and accountability mechanisms, their participation is often restricted to feedback on surface-level functionality. This reinforces a structural gap between newsroom governance needs and the layers of the AI development that remain externally controlled.

Bridge to our empirical contribution. Taken together, the existing literature suggests that participation in AI-mediated journalism is shaped by newsroom hierarchies, cross-functional coordination, shifting divisions of labor, and unequal infrastructural dependencies [15, 22, 24, 28, 32, 79, 88, 99, 101, 102]. Participatory outcomes in journalism therefore depend not only on inclusion but on how participation is *experienced* within newsrooms and organizational arrangements that unevenly distribute agency, accountability, and epistemic authority, and on whether system behavior, decisions, and accountability structures are sufficiently legible to support shared interpretation and contestation [12, 28, 29, 40, 56, 99]. These perspectives suggest that participatory AI in journalism must be understood as both layered (across normative, technical, and operational dimensions) and relational (unfolding through ongoing coordination between newsrooms, external technology providers, and the public) requiring frameworks that account for constrained influence within and beyond the newsroom. Building on these concerns empirically, our interviews show that newsroom stakeholders evaluate AI integration less through formal participation opportunities and more through the *experienced* conditions of agency, accountability, and epistemic legibility. We introduce a construct for this empirical pattern in our findings and design a framework to address it.

3 Overview of Our Approach

This paper offers an empirical and conceptual account of participation in AI adoption in journalism, organized as three contributions that stand in an analytic sequence: (1) an empirical account of how newsroom actors construe legitimate participation in AI adoption (Section 4); (2) a two-dimensional framework for reasoning about participation across organizational depth and collective scope

(Section 5); and (3) a set of organizational principles for implementing participatory AI governance under real workplace constraints (Section 6). The empirical findings were developed first, through thematic analysis of our interview corpus; the framework and its principles were then derived from them (the full analytic process is described in Appendix C.3). Concretely, an empirical pattern (*Contribution 1*, Section 4), motivates a conceptual reframing of participation (*Contribution 2*, Section 5), which in turn requires a set of organizational principles to remain credible in practice (*Contribution 3*, Section 6). We state these inferential moves explicitly below, articulating these three contributions in detail, and we provide a table (Table 1) to compare how our contributions build on the previously discussed foundational works (Section 2).

Starting point: the perception gap (Contribution 1). Drawing on interviews conducted in an independent local newsroom in continental Europe (Section 4), we find that stakeholders evaluate AI adoption not primarily through the formal availability of participation opportunities, but through the experienced conditions under which participation becomes meaningful: agency, epistemic legibility, and accountable governance. We describe this pattern as a *perception gap* between nominal inclusion and credible influence (Section 4.2). The gap is articulated along three dimensions, epistemic controllability, procedural agency, and the legitimacy of participation demands, and is accompanied by three patterns of conditional, role-dependent willingness (Patterns A-C, Section 4.4). These dimensions and patterns are analytic constructs we developed from the interview corpus (Appendix C.3).

From the perception gap to a participation matrix (Contribution 2). The derivation from the interview findings (Section 4) to the design space (Section 5) followed four explicit steps.

Step 1: extracting requirements. From the coded material we isolated two kinds of recurring statements about participation: statements about *what* participants wanted to influence (red lines and permissible uses; verification, disclosure, and escalation routines; prompts and tool behavior; shared resources that persist beyond individual use; Sections 4.2 and 4.3), and statements about *where* that influence should be exercised (desk and team practice, the newsroom as an organization, the profession, and relations with vendors; Section 4.3). We treated each recurring statement as a requirement that any model of participation in this setting must be able to represent.

Step 2: testing existing models against these requirements. A ladder [5] represents participation as a single ordered quantity, i.e., more or less power. Participants, however, did not describe more or less of the same thing. They distinguished different *objects* of influence by kind rather than amount, for example normative rule-setting versus optimizing a product (P7). They also placed the same object in different *arenas*: P4 wanted to shape guidelines "both here in our newsroom and for the profession more broadly." A single scale cannot hold both distinctions without collapsing one into the other.

Step 3: separating the two distinctions into axes. The object and the arena of influence varied independently across participants. P10 wanted consultative input on internal rules but no coordinating role; P3 sought deep, tool-level engagement and a direct line to vendors; field reporters wanted to receive guidelines rather than write them (Section 4.4). We therefore treated object and arena as orthogonal dimensions. Ordering the objects of influence by how far into the AI arrangement they reach (stating boundaries, shaping workflows, adapting tools, building durable shared arrangements) yields the *depth* axis. This ordering coincides with the surface-to-foundation layering of Suresh et al. [99], whose vocabulary we adopt. Ordering the arenas by how many actors a decision binds yields the *scope* axis, aligned with Tseng et al.'s micro/meso/macro levels [102]. The public level extends this axis from participants' concern with audience trust and traceable accountability and from prior work on reader-facing participation; it is therefore the level least directly derived from the interviews.

Step 4: rejecting a monotonic ordering of cells. Willingness was conditional and non-monotonic: 8 of 10 participants framed it in conditional terms, and several were willing to engage deeply in one area while declining what a ladder would treat as a "lower" rung (Section 4.4). Hence no cell is a prerequisite for another. This is what turns the crossed axes into a navigable matrix rather than a two-dimensional ladder, and it is the property the voluntary principle (Section **??**) formalizes. Each cell of the GVP matrix (Section 5) is then obtained by crossing one depth level with one scope level and asking which activity would realize that object of influence in that arena.

From the matrix to five organizational principles (Contribution 3). A matrix locates participation but does not, on its own, state what makes participation in any given position credible and not burdensome or performative. To answer this, we work backwards from the legitimacy conditions participants articulated to *five organizational principles* (Section 6): recognition without coercion, continuous value alignment, contextual knowledge scaffolding, sovereignty preservation, and transparent accountability. Each principle is grounded in specific perception-gap dimensions and willingness patterns: for instance, *recognition without coercion* derives from the legitimacy dimension and the recognition-sensitive, polarized responses to vendor-facing participation (Pattern C, Section 4.4), whereas *contextual knowledge scaffolding* derives from the epistemic-controllability dimension and the training conditions of Pattern B (Section 4.4). The complete mapping from findings to principles is provided in Appendix A.

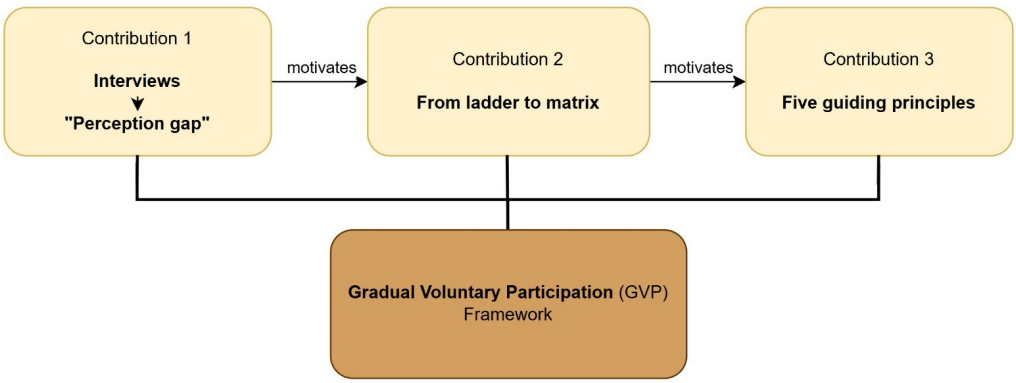


Fig. 1. Overview of paper contributions and their relationships. The three contributions build sequentially: empirical findings from newsroom interviews reveal the perception gap (Contribution 1), which motivates reconceptualizing participation as a navigable matrix rather than a unidirectional ladder (Contribution 2), which in turn motivates the derivation of five guiding principles for meaningful participation (Contribution 3). Together, these contributions inform the Gradual Voluntary Participation (GVP) framework for participatory AI integration in journalism described in Section 7.

Table 1. How GVP Contributions Build on Reference Papers

GVP Contribution	How it Builds on Related Work
Contribution 1: The "Perception Gap" (Empirical Account of Agency and Trust)	• **Operationalizes Opacity:** Draws on the "double black box" (dual opacity of prompts and reasoning) and "backstage behavior" (undisclosed AI use) to explain why formal rollouts fail to provide meaningful agency [108]. • **Addresses Confidence Gaps:** Responds to journalists identifying as "not a data person" and lack of technical self-confidence by framing these as barriers to experienced influence [20]. • **Protects Authority:** Integrates the finding that journalists view the "human element" as an "insurance policy" for professional authority, making human-led accountability a requirement for meaningful participation [4].
Contribution 2: The Two-Dimensional Matrix (Depth and Scope)	• **Adapts Depth Axes:** Moves beyond unidimensional ladders by adapting the "foundation, subfloor, and surface" layered blueprint into journalistic depths: Policy, Process, Co-design, and Infrastructuring [99]. • **Extends Organizational Scope:** Adopts and extends the "micro, meso, and macro" levels of participation, adding the Public scope to address documented audience disaffection [102]. • **Bypasses Path Dependency:** Addresses the "Matthew effect" and "innovation laggard" dynamics by creating a navigable matrix where local newsrooms can experiment without the risk of organizational failure [38].
Contribution 3: Five Organizational Principles (Implementation Framework)	• **Sovereignty Preservation:** Enforces the call for human-led "pluralistic governance" through unambiguous override mechanisms [4]. • **Contextual Knowledge Scaffolding:** Replaces the "training gap" with "just-in-time" modular learning linked to newsroom decision points [19]. • **Recognition Without Coercion:** Responds to the "extra labor" and "invisible labor" concerns of data workers and journalists in by making participation a credited professional path [102, 107]. • **Transparent Accountability:** Directly counters the "chain of mutual suspicion" and "hidden shame" of AI use through interactive audit trails and shared workspaces [108]. • **Continuous Value Alignment:** Operationalizes the Newsroom Tooling Alliance requirement for collective use-case voting and shared evaluation [102].

4 Contribution 1: Empirical Findings from a Local Independent Newsroom

This paper starts from semi-structured interviews with journalists and editors about *where* and *how* they would want to participate in AI/LLM integration. Rather than assuming participation as a uniform normative ideal, we examine how it is articulated in practice through professional roles, organizational constraints, editorial responsibilities, and varying degrees of willingness to engage. Across the interviews, participation emerged not simply as a question of access to tools, but as a question of governance, accountability, knowledge, and recognition within newsroom work. The framework developed in the following sections is derived analytically from these empirical patterns. We take these perspectives as the empirical problem-definition that the framework must satisfy: participation in AI adoption is heterogeneous, conditioned by workload and hierarchy, and experienced as legitimate only under specific organizational conditions. Importantly, participation and training were not the sole *a priori* focus of the interviews: they emerged as salient concerns as interviewees reflected on how generative AI intersects with professional identity, editorial standards, accountability, and the future organization of newsroom work.

4.1 Data and analytic approach

We conducted 10 semi-structured, face-to-face interviews in an independent local newsroom in continental Europe, with journalists occupying different roles and the newsroom director. Participants included seven journalists, one journalist with an additional technical role, one newsroom director, and one social media content manager. The sample comprised six male and four female participants. Ages ranged from 30 to 60 years. Five participants were in the 50–60 age range, two were in the 40–49 range and three were in the 30–39 range (a mapping of the participants' demographics and roles can be found in Table 2).

Table 2. Interview Participant Demographics and Roles

Participant	Age	Gender	Role
P1	50-60	Male	Journalist (sport and local news)
P2	40-49	Female	Journalist (sport)
P3	40-49	Male	Tech editor and journalist
P4	30-39	Female	Social media manager
P5	50-60	Male	Editor and journalist
P6	40-49	Female	Journalist (local and legal news)
P7	50-60	Female	Newsroom director
P8	50-60	Male	Journalist (local news)
P9	50-60	Male	Journalist (local news)
P10	30-39	Male	Journalist (sport and local news)

Each interview was conducted individually, in person, by two PhD researchers and lasted approximately one hour[1]. The interview protocol elicited preferences about (i) *internal participation* (collaboration with newsroom management on policies, permissible uses, and rollout rules), (ii) *knowledge-building participation* (internal/external training and peer learning to reduce knowledge silos), and (iii) *external participation* (feedback loops with vendors or providers about tool limitations,

[1]All participants provided informed consent. The study was approved by the Human Research Ethics Committee (HREC) of École Polytechnique Fédérale de Lausanne (EPFL) under protocol HREC000677, as part of the project "Value-ALIGNed socio-technical systems using large-language models (LLMs)" (decision issued 2 October 2025). Interviews were recorded, transcribed, anonymized, and the recordings were then deleted.

professional needs, and workflow constraints). More broadly, interviews explored how newsroom professionals perceived and approached generative AI, including patterns of use, expectations, concerns, values, professional identity, editorial standards, trust, authorship, perceived alignment, and AI's prospective role in journalism. Participation and training surfaced repeatedly within these discussions.

We analyzed the material through thematic synthesis focused on participation targets, willingness conditions, and resource constraints. Coding proceeded iteratively across transcripts, with themes refined through comparison across newsroom roles and positions. For a complete description of the analysis of the interview materials see Appendix C.3. See Appendix C for the full interview protocols.

The newsroom is small and editorially independent (not part of a larger media group) and operates across multiple platforms, producing both an online newspaper and television output; it accordingly brings together professionals with different editorial focuses and expertise. It operates within a national regulatory context in which journalists are bound by a statutory code of professional conduct, enforced by a professional body with mandatory registration and continuing-education requirements. At the time of our study, the newsroom had not adopted any formal internal policy governing AI use: journalists experimented individually with general-purpose generative AI tools (e.g., large language models and AI-based image and voice generation), accessed at their own discretion rather than through a centrally provisioned or sanctioned system, provided their practice remained consistent with the professional code. We return to how these characteristics may shape our findings in Section 7.2.

We deliberately focused on a single newsroom community to retain a shared organizational context, meaning constraints, norms, workflows, and governance structures, while still capturing heterogeneity in roles, demographics, experience, and attitudes toward AI. This sampling strategy supports a contextually grounded analysis of how participation is understood and negotiated within a specific organizational environment and aligns with participatory design research emphasizing that meaningful design knowledge is situated [35, 39, 91, 97]. We also selected a local independent newsroom because such settings remain comparatively under-examined in human-AI interaction and newsroom AI research [73, 103], despite often facing sharper resource constraints and distinct accountability dynamics than large national outlets [1, 83]. Following qualitative design research, we prioritize *transferability* over statistical generalizability [46, 68]: thick description enables readers to judge relevance to their own settings, and intensive observation of critical cases can yield insights that broad surveys cannot [96]. The framework is therefore grounded in this setting, while future work should test its transferability across multiple newsrooms, including larger organizations and different national contexts (Section 7.2.)

4.2 An emergent finding: the "perception gap"

Across interviews, journalists did not primarily evaluate generative AI adoption through model-centric criteria (e.g., output quality or efficiency), but through the *experienced conditions* under which AI would enter newsroom practice. In coding, we observed a recurring mismatch between how AI integration is often framed as a technical rollout and how participants assessed it through (i) perceived epistemic controllability, (ii) perceived procedural agency, and (iii) perceived legitimacy of participation demands. We group this mismatch under the **perception gap** theme: *a divergence between technical deployment and journalists' perceived ability to understand, influence, and remain accountable for AI-mediated work.*

Epistemic controllability: trust depends on perceived limits, provenance, and "red lines". Participants repeatedly emphasized that meaningful interaction requires profession-specific knowledge about

what AI can and cannot be relied upon for, where it gets information from, and how to recognize failure modes. Rather than requesting technical literacy for its own sake, journalists asked for normative and practical boundaries (e.g., "what I must not do with AI as a journalist"), as well as guidance on trust and tool selection**[P1, P2]**.

Procedural agency: participation was framed as influence over standards and accountability, not interface co-design. When participants expressed interest in participation, they emphasized contributing to internal policies and broader professional regulation, and some extended this to vendor-facing channels. Notably, participation was repeatedly framed as normative work (standards, rules, accountability), explicitly distinguished from merely optimizing a product or service**[P4, P7, P2, P3]**.

Legitimacy of participation demands: willingness depends on recognition, time, and role boundaries. Even participants favorable to participation described it as conditional on workload protections and recognition as work (e.g., compensated time, explicit role credit), while others resisted any expectation of taking on coordinator-like responsibilities outside journalistic duties. Time scarcity and asymmetric burdens made "participation" risk being perceived as unpaid extra labor rather than empowerment**[P6, P9, P10]**.

Why this matters for the framework. These patterns show that participation in AI adoption is not simply a matter of improving technical capability or offering formal opportunities for input. Rather, it is an organizational and group-work problem: journalists assess AI adoption through whether they have meaningful voice in decisions, whether systems are sufficiently legible to support professional judgment, and whether participation is recognized as part of newsroom work rather than added burden. The Gradual Voluntary Participation (GVP) framework is intended to address this problem by structuring participation across different levels of depth and scope and by making visible the organizational conditions under which participation becomes credible, coordinated, and sustainable.

4.3 What "participation" meant to participants

Participants consistently framed participation less as "co-designing AI" in an abstract sense and more as institutional involvement in decisions that shape everyday work, especially rules, boundaries, and accountability arrangements. Across interviews, participation clustered around three concrete themes:

(1) **Internal governance participation (newsroom-level):** contributing to policies and rules for AI use, defining acceptable use cases, and shaping implementation pathways with management. Participants referenced decisions about what AI uses should be permitted, disclosure requirements when AI is used in published work, verification standards for AI-generated content, and accountability escalation when AI outputs raise editorial concerns. Several emphasized wanting clarity on "what I must not do with AI as a journalist" and "what happens when something goes wrong" rather than technical configuration details. These forms of participation reflect journalism's strong tradition of internal norm-setting around editorial standards, where legitimacy derives from collectively defined professional, beyond purely technical optimization.
(2) **Capacity-building participation (learning & diffusion):** access to training and courses (internal and external), peer exchange, and mechanisms to prevent knowledge concentration in small groups. Participants distinguished between generic AI literacy courses (often offered

by professional organizations like the Order of Journalists for continuing education credits[2]) and journalism-specific training covering practical use cases, regulatory/legal constraints, and professional standards. Many cited time scarcity as the primary barrier, expressing preference for internal newsroom training (integrated into work hours and "on-the-job" learning through experimentation) over external theoretical courses. Several noted existing knowledge silos where only certain staff members understand AI capabilities, creating uneven power dynamics.

(3) **External feedback participation (vendor-facing):** communicating newsroom needs, reporting limitations discovered in practice, and influencing tool evolution through feedback channels, ranging from bug reporting to structured consultations. Participants envisioned vendor-facing participation primarily as individual journalists providing feedback on tool limitations discovered through daily use, communicating professional constraints around sourcing verification, legal risk, editorial tone, and audience harm considerations. Some referenced structured consultation channels (pilot programs, user feedback sessions), while others imagined less formal bug reporting or feature requests. Several noted that tools often felt too generic or multi-purpose rather than tailored to journalism standards, creating frustration that could be addressed through direct professional input to developers. This external orientation reflects the ties between newsrooms and technology providers, where journalistic practice depends on tools developed outside the profession, making cross-organizational participation a necessary condition for maintaining editorial standards.

4.4 Patterns of willingness: participation as conditional, not binary

Interviewees rarely expressed a simple "yes/no" stance toward participation. Instead, they articulated *conditional willingness*: participation was desirable under certain organizational conditions and undesirable under others. We summarize these conditions as patterns that recur across the three arenas above. The majority of participants (8 out of 10) explicitly framed willingness in conditional terms, using phrases like "I would like to participate, but..." or "it depends on whether...", rather than expressing unconditional enthusiasm or outright refusal. Three patterns emerged:

Pattern A: Governance-first participation is widely desired, but with role-specific boundaries. Many participants expressed strong interest in contributing to internal rules and policies, often because these decisions affect accountability, autonomy, and workload distribution. However, the preferred form of involvement varied by role and professional focus, and participants differed in how much time they could dedicate. Field reporters emphasized needing clear guidelines more than involvement in creating them, while the newsroom director emphasized collaborative discussion without formal documentation. Several journalists wanted active roles in both internal standards and broader professional regulation, while others preferred consultative input without coordinator-level responsibility.

Illustrative quotes: Journalists wanted a say in "what is allowed," "who decides," and "what happens when something goes wrong," but not necessarily in the technical configuration of the tool. One participant working in social media management for the newsroom, who had engaged in multiple conversations with journalists about AI use, reported wanting "a more active role in giving feedback on regulations for AI in journalism, both here in our newsroom and for the profession more broadly. We need clear guidelines about what's acceptable and how we keep our standards when using these tools" [P4]. Another emphasized that participation must extend beyond product development: "For me, participation can't just be about the product or making it work better from a technical point of view. It has to be normative, regulatory. Journalists should be involved in setting the rules and must

[2]https://www.odg.it/

figure this out all together" [P7]. Conversely, one participant expressed interest in contributing while maintaining clear boundaries: "I'm interested in these discussions, but I don't want an overly active role in this, like being a coordinator or taking on extra responsibilities. I'm a journalist. Everything else is kind of an appendage to my actual job. I want to participate, but I need to protect my time for the work I was actually hired to do" [P10].

How this motivates our framework. The prominence of internal governance suggests that participation must be representable as policy and process involvement, not only tool co-design. This motivates treating policy/process participation as a first-class participation mode in GVP, distinct from technical co-design, with varying depth levels that accommodate both active regulatory roles and consultative input, without mandatory coordination responsibilities.

Pattern B: Training is seen as participation infrastructure, but its value depends on access and institutional support. Participants described training courses as a mechanism to avoid knowledge silos and uneven power. Interest was high when training was (i) integrated into working hours, (ii) recognized as part of the job rather than extracurricular effort, and (iii) connected to concrete newsroom policies (not generic AI literacy). All 10 participants mentioned training or courses. Journalists across different experience levels wanted journalism-specific content covering both practical functionality and regulatory/legal constraints, distinguishing this from generic courses that "did not give me guidelines" and served primarily for earning professional credits [P1]. Several participants also emphasized preferring practical, on-the-job experimentation.

Illustrative quotes:

Interviewees framed training as necessary to "speak the same language" and obtain meaningful participation across the newsroom, but resisted it when it became an additional burden on top of deadlines. One journalist explained their training needs: "I don't need training about AI from a technical point of view. What I need are clear guidelines about what I must not do with AI specifically as a journalist. And management needs to commit to training employees" [P1]. Another emphasized the need for profession-specific education: "I want specific courses on AI in journalism, not generic AI courses. I would like to learn about the regulatory and legal side, applied to our work and profession, specifically" [P2]. Several participants also stressed that effective learning should be practical and embedded in newsroom routines, preferring "trying things out with AI, understanding where it fails" [P7] over predominantly theoretical instruction and favoring internally organized newsroom training over external courses. However, time constraints were pervasive: "I would like to do these courses, but it's very hard to find time. I'm already stretched thin, I never disconnect, I'm always on call" [P9]. Knowledge silos were explicitly noted, with several participants describing uneven AI literacy creating barriers to collective participation.

How this motivates our framework. Training functions as participation scaffolding: it enables meaningful engagement with governance and tool use, and it shapes who can participate. This motivates GVP's contextual knowledge scaffolding principle, emphasizing that training must be integrated into work time, tailored to professional contexts (not generic AI literacy), and provided through multiple modalities (courses, peer learning, on-the-job experimentation) to accommodate diverse learning preferences and time constraints.

Pattern C: Vendor-facing participation is polarizing and reveals workload/recognition conditions. Responses diverged most strongly on direct participation with tech companies and providers. Some interviewees welcomed the opportunity to influence tools using professional expertise; others saw vendor feedback as extra labor that should be compensated or formally recognized. Several framed vendor-facing participation as only feasible if the organization created protected time and clear incentives. Divergence appeared linked to workload intensity, prior experiences with vendors not

meeting professional standards, and frustration with tools being too generic rather than journalism-specific. Those most enthusiastic about vendor participation emphasized normative/regulatory influence, not just product optimization.

Illustrative quotes (pro): Some participants valued the chance to communicate professional constraints and newsroom standards directly to tool-makers, particularly those with direct technical experience. One journalist with hands-on experience using both LLMs and AI-based image and voice generation tools explained: "I've been experimenting quite a bit with these technologies and I think I understand how they work in general terms and I was able to notice some limitations as well. I'd really like to be involved in discussions with tech companies to provide direct feedback about what actually happens when you try to use these tools for journalism work" [P3]. One journalist expressed enthusiasm for direct engagement: "I have strong interest in having a more participatory role, both internally in how we use and regulate these systems in the newsroom, and in having a more direct line with producers and tech companies to communicate my feedback on the tools and my needs as a professional. About errors I may find as well" [P2]. Another emphasized that participation should extend to regulatory dimensions: "It is really interesting when companies propose that their employees actively participate in the development and building of the tools they will then use, thinking about the professional standards and constraints we need from the start" [P7]. Professional constraints emphasized included sourcing verification requirements, legal risk boundaries, editorial tone standards, and preventing audience harm.

Illustrative quotes (conditional): Others were willing only if it counted as work: paid time, explicit role credit, or workload reductions elsewhere. One journalist stated: "it would also be interesting to participate, to have a more active role, but it is extra work and should be compensated/considered working time. I would not want it to be imposed on me as another job" [P6]. Recognition forms proposed included formal working time allocation, compensation for feedback contributions, and acknowledgment in role descriptions or performance evaluations.

Illustrative quotes (resistant): Some rejected it as "not my job," preferring vendor interaction to be delegated to dedicated roles. One participant stated: "we have to consider that I do another job; I wouldn't want it imposed. I am a journalist, everything else I consider an appendage to my job, so I would not want to have an important and overly active role in this aspect" [P10]. Suggested delegation targets included product managers, innovation teams, newsroom standards leads, or professional organization representatives who could aggregate and communicate feedback on behalf of multiple journalists.

How this motivates our framework: This divergence highlights the need to model participation as voluntary and resource-sensitive, with explicit handling of recognition, compensation, and role boundaries. This motivates GVP's recognition-without-coercion principle and its emphasis on voluntary participation with transparent time/resource requirements. The framework must accommodate both individual direct engagement and mediated participation through dedicated roles or professional organizations, allowing journalists to opt into vendor-facing participation only when adequately supported and recognized, rather than imposing it as a default expectation.

5 Contribution 2: The Design Space - Rethinking Participation in Journalism, from Ladder to Matrix

The perception gap identified in Section 4.2 requires reconceptualizing participation in two specific ways. First, the *procedural-agency* finding, meaning that journalists differentiate influence over standards, rules, and accountability from influence over tool functionality, shows that participation varies in *depth*, i.e., in the kind and degree of influence at stake, not only in how much of it one has. Second, the finding that legitimate participation is distributed across newsroom, professional, and vendor-facing dimensions (Section 4.3) shows that participation also varies in *scope*, i.e., in the

dimension where it takes place. Taken together with the conditional, non-linear willingness patterns of Section 4.4, these findings reveal why a single ascending scale is inadequate, and motivate the move from a ladder [5] to a matrix developed in this section.

5.1 Beyond the Ladder: Depth and Scope

Arnstein's influential ladder of participation [5] conceptualizes participation as a vertical progression from non-participation, through tokenism, to citizen power. While foundational, this model assumes a largely unidirectional movement toward increased power and does not readily account for the heterogeneous, overlapping, and role-dependent forms of participation that characterize AI adoption in organizational settings. In newsrooms, participation is not simply a matter of moving upward along a single scale, but is instead shaped by professional hierarchy, workload, domain expertise, and the distribution of decision-making authority across actors within and beyond the organization. As a result, ladder-based metaphors are limited in their ability to capture how participation is organized, constrained, and negotiated across multiple dimensions of collective work.

Other frameworks, such as the one proposed by Suresh et al. [99], tackle the need to conceptualize "participation in the age of foundation models", hence visualizing participation as potentially happening at different institutional and design layers on top of the chosen "foundation" model, adding a sense of *depth* rather than only evolving gradually like across the rungs of a ladder. Similarly, in the context of journalism, a mapping of the evolving dynamics across stakeholders dealing with AI implementation was proposed by Tseng et al. [102], who highlight different *scopes* of participation at "micro", "meso", and "macro" levels for newsrooms.

We advocate for PD frameworks for AI in journalism that are built on top of these concepts across two main dimensions of a participatory matrix (Figure 2):

- **Depth** (X axis): the kind and degree of influence over AI-related decisions, from *policy definition* (articulating values, red lines, and acceptable-use boundaries), through *process shaping* (influencing how AI enters workflows, review, and escalation), to *co-design* (iterating on tool use, prompts, and evaluation criteria) and *infrastructuring* (shaping durable sociotechnical arrangements that persist beyond a single tool), mirroring Suresh et al.'s layered blueprint for participation in foundation-model ecosystems [99].
- **Scope** (Y axis): the dimension in which participation takes place, from *micro* (individual and team practice), through *meso* (organization-wide or cross-newsroom coordination) and *macro* (sectoral and professional governance), to *public* (reader- and civic-facing accountability). The first three levels follow Tseng et al.'s micro/meso/macro tensions in journalism [102]; we add the public level to capture reader-facing participation.

To clarify how these levels can be interpreted in practice, Appendix B (Table 4) provides working definitions, observable indicators, and example governance artifacts for each.

We adopt the checkerboard as a visual metaphor for its capacity to render participation legible as a spatial configuration, making the framework visually concrete much as the ladder [5] did for earlier models of participation. The board supports orientation by making positions, trajectories, and co-existing forms of engagement visible at a glance, enabling stakeholders to situate themselves relative to others without prescribing sequences or outcomes. In this sense, the checkerboard emphasizes stakeholder agency in choosing pathways, discovering alternatives, and coordinating with peers across the participation landscape, beyond simply "climbing a ladder" of participation.

This model reflects the dynamic and socio-technical nature of digital journalism and newsrooms, as stakeholders do not necessarily move through these dimensions in a linear way. A journalist

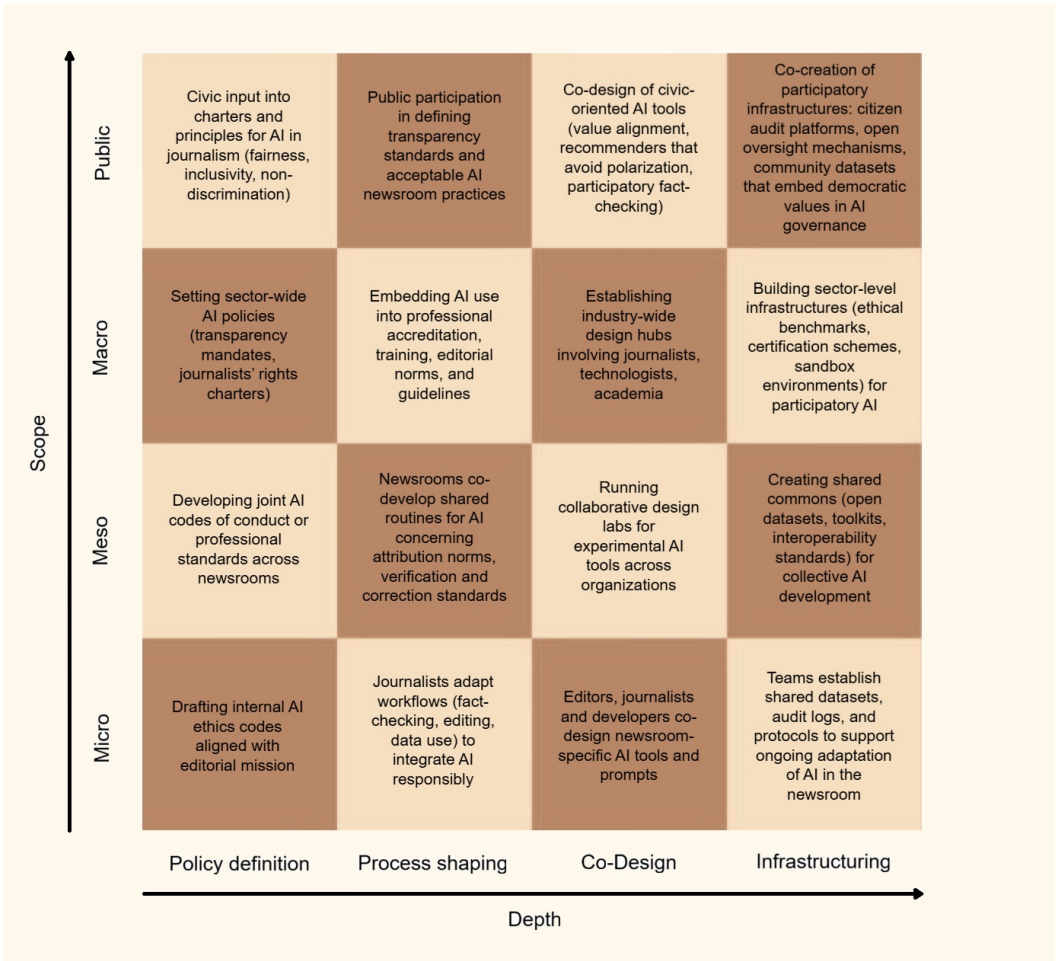


Fig. 2. The GVP checkerboard, where each quadrant represents a distinct degree of participation across various scopes and depths of AI integration in journalism.

might deeply engage with prompt engineering (high depth, narrow scope) while minimally participating in industry-wide policy discussions (low depth, broad scope). Additionally, this variation creates natural lightweight "experimental" conditions for understanding how different participation configurations influence perception and workflow adaptation, addressing recently surfaced challenges and opportunities for AI in local journalism [20, 108]. The journalist with deep technical engagement but minimal policy involvement likely perceives AI differently than another one with inverse participation patterns. These differences become analyzable, rather than problematic, when the participation framework itself supports such multidirectional movement as an expected design feature, instead of treating it as deviance from a prescribed path. Unlike hierarchical models that assume unidirectional progression, this matrix creates a navigable participation space: stakeholders can explore multiple pathways simultaneously, moving vertically (increasing depth within current scope), horizontally (expanding scope at current depth), or diagonally (combined shifts across both dimensions). This multidirectional movement reflects how participation unfolds in practice, as

actors engage across different roles, decision arenas, and levels of influence rather than following a single prescribed sequence.

The matrix thus functions as both a conceptual model and an organizational tool: it makes visible where participation is located, what forms of involvement are available under different conditions, and how participation is distributed across actors, roles, and arenas of coordination, making peer participation patterns visible across the landscape. Framed this way, the matrix supports what Niemantsverdriet et al. [77] call "awareness in shared systems", making participation, interdependence, and accountability more visible across roles and sites of coordination.

Prior work suggests that structured spatial representations can help users understand complex multidimensional relationships more effectively than fragmented alternatives [85]. Relatedly, feedforward mechanisms can support exploration by previewing possible consequences before commitment, especially in open-ended tasks where actors may consider multiple participation pathways [31, 100, 105].

The proposed matrix structure responds to what Bratteteig and Verne [11] identify as AI's core challenge to traditional PD: systems that change continuously through learning require participation models that accommodate ongoing, multi-directional engagement rather than fixed, unidirectional progression.

5.2 The Voluntary Principle and Self-Determination Theory

Mandatory participation creates what Birhane et al. call "epistemic burden" [9], requiring stakeholders to invest cognitive and temporal resources without compensation. In practice, participation in AI design, adoption and governance is often presented as voluntary while remaining structurally compulsory through organizational expectations, role responsibilities, and the coordination demands of collective work. Non-voluntary participation may take the form of mandatory tool use without meaningful avenues for feedback, implicit expectations to engage in consultation processes as part of professional duties, or alignment activities that are formally voluntary but tied to performance evaluation, workload, or organizational compliance [54]. An alternative is to center participation on voluntary choice while still recognizing contributions, thereby avoiding coercion while maintaining engagement.

This approach draws from self-determination theory, which identifies three basic psychological needs driving intrinsic motivation: autonomy (the need to volitionally endorse one's actions), competence (the need to feel effective), and relatedness (the need to connect with others) [27]. By preserving choice while providing recognition, voluntary participation sustains engagement without exhausting stakeholders. The voluntary principle suggests that chosen participation produces fundamentally different perceptions than mandated engagement, a distinction critical for understanding why similar participation activities produce varying perceptual outcomes across different organizational contexts.

6 Contribution 3: Five Principles for Gradual and Voluntary Participation in Newsrooms

To center participants' lived experience within a framework for AI governance, a set of organizational principles is needed to ensure that even gradual and voluntary participation remains meaningful within collective work, rather than tokenistic or performative. In this section, we present the principles as well as corresponding operational steps (concrete recommendations) that help translate theory into practice. A detailed mapping of how each principle is grounded in the empirical findings is provided in Appendix A, Table 3.

The five principles below are derived directly from the empirical findings of Section 4: each translates a recurring concern from the perception gap (Section 4.2) and the conditional-willingness

patterns (Patterns A-C, Section 4.4) into a design-level commitment that keeps gradual and voluntary participation meaningful (rather than tokenistic or performative) within collective work. For each principle we first state the perception-gap dimension and pattern(s) it primarily responds to, then give concrete operational steps that translate it into practice. Appendix A (Table 3) provides the full mapping, including the analytic move from each situated finding to its corresponding principle.

6.1 Recognition Without Coercion

Grounded in: the legitimacy of participation demands (Section 4.2), especially Pattern C and Pattern A (Section 4.4).

We call the first interaction design principle *Recognition Without Coercion*, reflecting what Delgado et al.'s participation spectrum suggests [28]: multiple valid approaches rather than hierarchical progression. Transparent certification systems can document participation levels, providing visible feedback about engagement status without creating prerequisites.

Concrete Recommendations:

- Develop transparent and shared certification systems and interfaces that document participation levels without creating professional prerequisites.
- Establish multiple navigable pathways to professional development, with AI engagement as an option, not an obligation.
- Conduct stakeholder mapping and power analysis to surface hidden asymmetries and contextualize recognition processes.

We hypothesize that visibly recognized but non-mandatory participation can support collective sense-making by making AI-governance labor legible as work, rather than informal extra effort. This directly addresses the legitimacy dimension of the perception gap (Section 4.2): participation is experienced as meaningful only when its time, burden, and role implications are acknowledged. Recognition also shapes how the newsroom distributes interpretive and coordination work, allowing journalists to see how colleagues contribute to policy definition, value alignment, or vendor-facing feedback without being pressured to follow the same path. Since responsible technology use emerges through the coordination of different roles, concerns, and interpretations [56], recognition without coercion can prevent AI governance from concentrating among a few technically confident actors while preserving multiple legitimate forms of participation.

6.2 Structures for Continuous Value Alignment

Grounded in: procedural agency (Section 4.2) and Pattern A, reinforced by Pattern B (Section 4.4).

Continuous Value Alignment aims to show system responsiveness through regular value collection cycles and threshold-triggered policy reviews, where different stakeholders can engage directly through interactive touchpoints to propose values, and have their opinions collected and visualized by the system. Drawing on participatory action research [55, 84], values must emerge from ongoing negotiation among affected parties through responsive interaction mechanisms rather than static imposition.

Concrete Recommendations:

- Implement regular value collection cycles capturing different stakeholder priorities.
- Create threshold feedback mechanisms triggering policy reviews when values shift significantly.
- Enable bottom-up value proposition through accessible input affordances, while maintaining institutional coherence.
- Incorporate value collection logs to reveal implicit tensions and inform alignment cycles.

Supporting alignment also requires mechanisms that make the effects of participation traceable over time. Prior work on history flow visualizations [106] and knowledge transfer graphs [110] shows how contributions can be followed as they propagate through collaborative processes, directly applicable to showing journalists how their value propositions influence policy evolution, workflows, or governance decisions.

We hypothesize that continuous value-alignment cycles can serve as deliberative infrastructure for the newsroom and not only as mechanisms for collecting individual preferences. This responds to the procedural-agency dimension of the perception gap (Section 4.2): journalists need influence over standards, rules, and accountability arrangements, not only opportunities to react to already-defined tools. By repeatedly surfacing and documenting stakeholder priorities, these cycles help journalists, editors, managers, and technical intermediaries coordinate concerns about acceptable AI use, editorial standards, and responsibility. Value alignment is therefore treated as collective sense-making: negotiation becomes visible, revisable, and open to contestation over time [56, 65]. This responds to Pattern A (Section 4.4), where participants prioritized governance and rule-setting over product optimization, and to work showing that responsible data and algorithmic practices require collective interpretation rather than one-time technical settlement.

6.3 Contextual Knowledge Scaffolding

Grounded in: epistemic controllability (Section 4.2) and Pattern B (Section 4.4).

Contextual Knowledge Scaffolding addresses information asymmetries through modular learning resources linked to specific interaction and decision points in the AI-augmented content generation pipeline. Drawing on situated learning theory [66], this interaction principle recognizes that perception emerges from legitimate peripheral participation, with knowledge acquisition occurring through graduated involvement rather than front-loaded prerequisites.

Concrete Recommendations:

- Develop modular knowledge resources linked to specific interaction and decision points.
- Create multidirectional transparency where all stakeholders can access the same learning information.
- Design peer-learning networks supplementing formal training.
- Integrate scaffolding activities and interfaces with stakeholder mapping to ensure accessibility across differing expertise levels.

This scaffolding approach builds on Yildirim et al.'s finding that providing capability information at ideation-time helps stakeholders generate a broader range of concepts [109], though ensuring technical feasibility requires active collaboration with data scientists. In journalism contexts, just-in-time learning resources linked to specific decision points (e.g., when selecting AI summarization parameters) could reduce epistemic burden while enabling informed participation.

Empirical work supports this just-in-time approach to knowledge support. Grossman and Fitzmaurice found that contextual video assistance helped users complete substantially more unfamiliar tasks than traditional online help, with better retention over time [53]. Similarly, Chilana et al.'s *LemonAid* system showed that contextual, selection-based help could surface relevant guidance for many user queries [17]. These findings reinforce GVP's emphasis on embedding knowledge support within ongoing work practices, rather than treating competence as something that must be fully established in advance.

We hypothesize that contextual knowledge scaffolding can support collective sense-making by reducing role-based knowledge asymmetries. This addresses the epistemic-controllability dimension of the perception gap (Section 4.2): journalists need shared resources for understanding AI limits, provenance, and failure modes before they can meaningfully participate in governance. In our case,

training was not only an individual up-skilling demand, but a condition for avoiding knowledge silos and enabling different actors to deliberate with a shared vocabulary. Modular, situated, and peer-supported learning can make AI behavior discussable across roles, allowing journalists, editors, managers, and technical intermediaries to compare interpretations of risks and appropriate uses. Training thus becomes coordination infrastructure, in line with work on collective sense-making and cross-functional collaboration around AI in newswork [56, 107].

6.4 Sovereignty Preservation

Grounded in: procedural agency, with epistemic controllability as a precondition (Section 4.2); Pattern A (Section 4.4).

Sovereignty Preservation maintains clear interaction hierarchies between human judgment and AI output through unambiguous override mechanisms. Clear interface hierarchies, where human judgment supersedes algorithmic output, contribute to human agency preservation.

Concrete Recommendations:

- Establish unambiguous override affordances, mechanisms and standards at every interaction and decision point.
- Create documentation systems recording when and why human judgment diverges from AI suggestions.
- Build accountability structures, such as certification seals visible in user interfaces, ensuring humans remain responsible for outcomes.
- Adopt graduated implementation pathways that begin with voluntary pilots and expand incrementally based on demonstrated trust and value.

These override mechanisms align with foundational human-AI interaction principles. Horvitz [58] and later Amershi et al. [3] proposed that systems must support efficient invocation, termination, dismissal, and correction of AI services, minimizing costs when users reject automated outputs. Buçinca et al. further demonstrated that cognitive forcing functions, which require analytical engagement before revealing AI recommendations, reduce overreliance, and support appropriate human authority [16].

These patterns also align with Berge et al.'s findings on human-AI collaboration in professional settings [8]: preserving professional autonomy in high-stakes domains requires workflows that sustain human judgment throughout ongoing work, rather than relying on post-hoc override options that arrive too late in the workflow.

We hypothesize that documenting human divergence from AI recommendations can support collective rule formation, not only individual autonomy. This responds primarily to the procedural-agency dimension of the perception gap , while also making accountability tensions visible where editorial judgment conflicts with algorithmic suggestions. Override events can become organizational evidence for revising guidelines, escalation paths, verification routines, or restrictions on use. Sovereignty preservation is therefore not just a human-in-the-loop safeguard, but a way for the newsroom to learn where AI assistance misaligns with professional practice. This connects GVP to work on group-level interactions with algorithmic systems and on moving individual GenAI use into collaborative newswork [65, 108].

6.5 Transparent Accountability

Grounded in: procedural agency (Section 4.2), with provenance/traceability concerns from epistemic controllability; Pattern A (Section 4.4).

Transparent Accountability requires more than reporting outcomes; it involves making the processes behind those outcomes visible and explorable to different stakeholders through interactive

mechanisms and interfaces. This can be achieved through multi-level disclosure tailored to the needs of each group, creating what the algorithmic accountability literature [67] demands: process visibility and legibility through interaction beyond outcome transparency.

Concrete Recommendations:

- Create interactive audit trails that document how inputs shape outputs.
- Establish clear attribution linking decisions to responsible parties.
- Develop accessible documentation and interfaces translating technical details into understandable terms.

This multi-level approach operationalizes Dhanorkar et al.'s framework [29], which originally demonstrated that data scientists, product managers, and business owners require different explanations at various AI lifecycle stages, to the newsroom context. This application recognizes that developers, editors, journalists, and readers similarly possess distinct transparency needs, confirming that a single explanation mechanism cannot serve all stakeholder groups simultaneously.

We hypothesize that transparent accountability mechanisms can support retrospective newsroom sense-making by helping actors reconstruct how AI-assisted decisions emerged, where human intervention occurred, and how responsibility should be allocated. This addresses the accountability problem at the core of the perception gap: participation is not credible if actors cannot later understand, reconstruct, or contest how AI-mediated work was produced. Audit trails should therefore be treated as coordination devices, not only as evidence that participation matters. This is especially important where AI use remains private or backstage, creating uncertainty about provenance, authorship, and responsibility. By documenting prompts, revisions, overrides, value inputs, and decision points, accountability infrastructures can move the newsroom from suspicion to shared reconstruction, making AI-assisted work auditable and collectively governable [56, 65, 108].

The five principles are *transversal* to the matrix: rather than belonging to particular cells, each applies at every position a stakeholder may occupy. What changes from cell to cell is not *whether* a principle applies but *how* it is operationalized, since the *depth* and *scope* of participation shift the concrete form each principle takes. The newsroom scenario in Section 7.1 traces these instantiations in context: for example, *Recognition Without Coercion* appears as optional policy drafting at *micro/policy-definition* and as mediated representation at *meso/policy-definition*, while *Transparent Accountability* appears as shared prompt histories at *micro/co-design* and as collective audit logs at *micro/infrastructuring*.

7 GVP: The Gradual Voluntary Participation Framework

Building on these contributions, the Gradual Voluntary Participation (GVP) framework integrates the depth-scope matrix of Section 5 with the five principles of Section 6 into a single model for navigating stakeholders' perception of participatory AI across voluntary and gradual forms of engagement. The remainder of this section illustrates the framework through a newsroom scenario (Section 7.1) and highlights the challenges and limitations of our approach (Section 7.2).

7.1 Sketching a possible GVP application: addressing tensions in local newsrooms

The following scenario illustrates GVP's interaction patterns in practice, showing how stakeholders might engage with participation through specific interface touchpoints and feedback mechanisms, addressing both the structural tensions documented in local newsrooms and the empirical patterns identified in our interviews. We argue that when a hybrid professional environment such as a hypothetical newsroom adopts the Gradual Voluntary Participation (GVP) framework to guide their AI governance and implementation processes, participation can unfold as a lightweight, fluid, multi-level process that evolves voluntarily rather than through prescription and one-time

campaigns. We present here just one of the possible examples of this evolution; for reasons of space, we offer only a brief sketch, describing only the operationalization of the most relevant principles at different coordinates (importantly, each of them remains valid at each coordinate) and leaving deeper scenario exploration for future work.

7.1.1 Micro level. At the micro level, participation can begin at the lowest depth, where the shared concern of the interviewees can be addressed directly during policy definition. Across interviews, participants asked for normative clarity: "what I must not do with AI as a journalist" and "what happens when something goes wrong" (P1). This governance-first orientation (Pattern A, Section 4.4) maps directly onto the “micro/policy-definition” cell of the checkerboard (Section 5), where drafting internal AI ethics codes aligned with the newsroom’s editorial mission is the primary activity. In a small local newsroom, this means a small group encompassing diverse roles voluntarily contributes to drafting a shared policy document covering permissible AI uses, disclosure requirements for published content, and accountability escalation procedures when AI outputs raise editorial concerns. Maintaining a valid *Recognition Without Coercion* (Principle 1, Section 6.1), the policy-drafting process is explicitly framed as credited, optional participation, and it’s not intended as a default professional obligation. This reflects what expressed by interviewees reluctant to take on coordinator-level responsibilities (P10: "I’m a journalist. Everything else is kind of an appendage to my actual job"), who are not required to join. From policy definition, participation can move horizontally to the “micro/process-shaping” cell, where the agreed policy is translated into concrete workflow adjustments: deciding at which editorial steps AI assistance is permitted, how AI-generated outputs must be flagged in the content management system, and how journalists should document their prompts and edits. This cell is where the “innovation enthusiast” dynamic becomes both a resource and a risk: research on local newsrooms identifies a small number of tech-confident journalists who drive AI experimentation while the majority of colleagues remain peripheral [86], a pattern visible in our own interviews, where P3 had already been "experimenting quite a bit" with LLMs and AI-based image and voice tools, while others described finding such technologies "intimidating." In the process-shaping cell, such “innovation enthusiasts” can contribute by prototyping workflow integrations in a safe "AI news zone" [107, 108] and documenting what actually happens when journalism-specific constraints (e.g. sourcing verification, legal risk, editorial tone) encounter the tool’s limitations. Crucially, *Contextual Knowledge Scaffolding* (Principle 3, Section 6.3) governs how this knowledge is made available to others: instead of requiring colleagues to attend external theoretical courses they have no time for (P9: "I’m already stretched thin, I never disconnect, I’m always on call"), the newsroom makes available these modular, “just-in-time” documentation resources attached to specific decision points in the agreed workflow [20]. A journalist who wants to understand only how AI summarization should be used in the sports section can access exactly that guidance, without being required to engage with the full technical landscape. This targeted approach directly addresses interviewees’ consistent request for profession-specific, rather than generic, AI literacy (P2: "I want specific courses on AI in journalism, not generic AI courses"). Moving further along the depth axis to the “micro/co-design cell”, editors, journalists, and, where available, in-house developers and data experts collaborate to build newsroom-specific AI configurations: boundary objects such as prompt templates calibrated to the outlet’s editorial voice, verification checklists adapted to its beat coverage, and integrations that make AI assistance visible within existing editorial flows [107]. This is also the cell where the previously mentioned "double black box" problem (i.e. the dual opacity in which editors cannot see how a journalist prompted an AI system, and journalists cannot understand the model’s underlying reasoning [108]) is directly addressed. *Transparent Accountability* (Principle 5, Section 6.5) is operationalized here through shared GenAI workspaces embedded in shared documents, displaying prompt histories, revision

paths, and human override events alongside the published output. This makes the journalist's authorial decisions auditable and legible to colleagues, directly countering the chain of mutual suspicion that arises when AI use remains "backstage" [108]. *Sovereignty Preservation* (Principle 4, Section 6.4) is simultaneously enforced at this depth through unambiguous override affordances built into every co-designed workflow step, ensuring that human editorial judgment is always the terminal decision point. This concretely addresses what our interviewees understood by procedural agency, meaning the assurance that accountability for editorial outcomes remains traceable and human (Section 4.2). The deepest micro-level cell, "micro/infrastructuring", is where teams can establish the shared datasets, audit logs, and adaptation protocols that sustain ongoing AI integration in the newsroom. Reaching this cell relies on accumulated trust and technical literacy, and not all journalists need to contribute here. GVP accommodates this through its voluntary principle: infrastructuring participation can be concentrated in a small group of interested journalists, editors and technically engaged staff, provided their contributions are governed through *Transparent Accountability* (Principle 5, Section 6.5) mechanisms that keep the broader newsroom informed and able to contest decisions. This cell also anticipates and partially addresses the "innovation driver" fragility documented in local news research [38]: by institutionalizing accumulated knowledge such as prompt libraries, audit logs and protocol documentation into shared infrastructure rather than individual expertise, the newsroom ensures that AI integration capacity does not collapse when a technically confident journalist leaves [86].

7.1.2 Meso level. As experiences grow, collaboration can expand to the meso level, where several newsrooms join forces across the scope axis. The "meso/policy-definition" cell is where collective negotiation power first becomes visible: multiple outlets can collaboratively develop joint codes of conduct and professional standards for AI use that no single resource-constrained newsroom could produce alone. This addresses both the "Matthew effect" in local news innovation, where only larger incumbents have the organizational cushion to invest in governance infrastructure [38], and the vendor-facing participation concern that was the most contested dimension in our interviews (Section 4.4). Interviewees who wanted a "more direct line with producers and tech companies to communicate feedback" (P2) and those who framed vendor-facing participation as only viable if protected time and compensation were provided (P6: "it is extra work and should be compensated") can both be accommodated at this cell: designated representatives can aggregate and communicate multi-newsroom feedback to external developers, reducing the epistemic burden on reluctant individuals while ensuring professional constraints around sourcing, legal risk, and editorial tone reach those responsible for tool development. Moving to "meso/process shaping", newsrooms co-develop shared workflow routines for AI-assisted content: attribution norms, verification standards, and correction procedures that apply consistently across outlets. The *Continuous Value Alignment* principle (Principle 2, Section 6.2) is operationalized here through periodic consortium meetings where earlier decisions are reviewed against experience gathered from practice, directly addressing what our interviewees described as the need for ongoing normative negotiation rather than static rule-imposition (P7: "Journalists should be involved in setting the rules and must figure this out all together"). At the "meso/co-design cell", collaborative design labs can bring together journalists, developers, and editors across organizations to build experimental AI tools adapted to shared local journalism needs: verification pipelines tuned to regional source networks, summarization tools calibrated to local audience registers, or interview transcription systems aligned with specific beat workflows. This cross-organizational co-design addresses the tool-specificity frustration that surfaced repeatedly in interviews, where participants noted that available tools "felt too generic or multi-purpose rather than tailored to journalism standards." Finally, the "meso/infrastructuring" cell is where shared commons are built: open repositories

of prompt templates, annotated datasets, and interoperability standards, governed as collective infrastructure. Similarly to what Tseng et al. conceptualize as a News Tooling Alliance [102], this ensures that knowledge and tooling investments accumulate across the network without remaining siloed within individual outlets, directly combating the innovation driver fragility that leaves small newsrooms perpetually rebuilding from scratch when staff turns over.

7.1.3 Macro level. A useful contrast to this trajectory is offered by recent conflicts around the use of journalistic content in foundation model training. As illustrated by Suresh et al., individual actors such as news organizations or journalists often lack the leverage to influence data practices at the foundation layer when acting alone, as exemplified by lawsuits filed by individual news organizations against major AI developers [78, 99]. In the absence of coordinated structures, such efforts risk remaining fragmented, costly, and ultimately ineffective in shifting industry norms. From a GVP perspective, this scenario highlights what happens when participation does not scale beyond isolated micro or meso engagements. In contrast, the expansion toward macro-level participation creates the conditions for collective stakeholder power. At the "macro/policy-definition" cell, newsrooms, unions, and professional organizations collaboratively set sector-wide AI policies: transparency mandates, journalists' rights charters, and data governance standards that no single outlet or consortium could impose unilaterally. This aligns with the "subfloor" dynamics described by Suresh et al., where coordinated group participation enables stakeholders to exert influence on the foundation layer itself [99], reframing participation as a proactive, infrastructural process through which journalistic actors shape the conditions of both AI development and integration. The "macro/co-design" cell brings together journalists, researchers, and academics in industry-wide design hubs to develop accreditation systems, training curricula, and editorial guidelines that embed professional norms into AI tool development from the outset, addressing at scale the governance-first orientation that our interviewees consistently prioritized over technical co-design. In parallel, at the "macro/infrastructuring" cell, sector-level ethical benchmarks, certification schemes, and sandbox environments for participatory AI are established as durable professional infrastructure.

7.1.4 Public level. From there, participation in GVP naturally extends to the public scope, where a wider range of stakeholders, including journalists, readers, researchers, and civic groups, can engage in open dialogues about AI's role in news production. The "public/policy-definition" cell involves civic input into charters and principles for AI in journalism, addressing questions of fairness, inclusivity, and non-discrimination, while the "public/co-design" cell opens the design of civic-oriented AI tools (value alignment recommenders, bias detection interfaces, participatory fact-checking systems) to reader participation. GVP addresses the documented growth of public suspicion toward AI-generated content [78, 108] by implementing *Sovereignty Preservation* (Principle 4) at the public scope: at the public/infrastructuring cell, citizen audit platforms and open oversight mechanisms provide tangible public evidence that human judgment remains the final authority over AI-assisted outputs. This connects back directly to the legitimacy concern our interviews identified about editorial sovereignty and publicly traceable accountability (Section 4.2), extending that internal requirement outward to audience trust. Readers may give feedback on AI-assisted articles or be consulted for value alignment through interactive touchpoints, while visible provenance markers and open audit trails make the process intelligible to audiences. At this level, we argue that the participatory effort resembles what Bratteteig et al. describe as infrastructuring [13], a sustained, collective process where journalists, organizations, and readers co-evolve practices through voluntary, aligned, and transparent engagement. Across these moves throughout the checkerboard, the back-and-forth exchange of journalists refining tools, organizations collecting insights, and readers evaluating outcomes is at the core of GVP: such dynamics can potentially

transform resistance into curiosity and position AI as a collectively and voluntarily governed innovation built on gradual and shared legitimacy, influencing perception.

7.2 Challenges and Limitations

Adopting a framework like GVP, while proactive and user-oriented, introduces several practical and conceptual challenges. A primary difficulty lies in operationalizing the five core principles across different depth–scope contexts, which may encounter resistance depending on where and how stakeholders choose to engage. For example, recurrent alignment mechanisms may be difficult to sustain in resource-constrained settings, while contextual knowledge scaffolding could create tensions between participants with differing expertise or priorities. A further tension emerges between GVP's emphasis on gradual, voluntary participation on one hand and organizational pressures for rapid AI adoption on the other. Newsrooms and other institutions often face external urgency driven by competitive dynamics, technological change, or managerial timelines, which may conflict with the slower, exploratory rhythms required for reflective participation. In such contexts, gradualism may be perceived by some stakeholders as delaying action rather than enabling legitimacy, particularly when participation is framed as optional rather than mandatory. This creates a risk that GVP is either bypassed in favor of fast deployment or selectively applied in ways that undermine its core principles. These pressures may also blur the boundary between voluntary and non-voluntary participation. For instance, stakeholders may be formally invited to engage while lacking a meaningful option to decline, or participation may be implicitly required to ensure compliance with new AI systems. In such cases, participation risks becoming performative or coercive, undermining the perceptual and legitimacy-building goals that GVP seeks to support. The dynamic, non-linear nature of participation trajectories also presents challenges. Stakeholders may instantiate unexpected movement patterns, contest roles, or experience conflicts when voluntary engagement intersects with organizational obligations. Finally, evaluating perception remains a persistent challenge [10, 13]. As GVP emphasizes voluntary and gradual engagement rather than prescriptive outcomes, traditional metrics of "success" or completion may be inadequate. Observed shifts in understanding, negotiation, and alignment can be subtle, uneven, or temporary, further complicating assessment. A further conceptual limitation concerns the meaning of governance itself. GVP is designed to support gradual, voluntary, and perception-sensitive participation, but it does not in itself specify enforcement mechanisms or consequences for actors who disregard agreed processes or norms. In this sense, GVP may be better understood as a governance-support framework useful to understand, plan and navigate participatory AI practices in multi-stakeholder domains such as journalism, rather than a fully institutionalized governance regime. This creates a risk that, in the absence of accountability structures, participation remains weak, symbolic, or performative rather than substantively shaping organizational AI practices. Acknowledging this limitation is important, as voluntary engagement alone does not guarantee compliance, responsibility, or institutional follow-through. Overall, while GVP offers mechanisms for shaping positive perceptions and supporting reflective engagement, these same features can introduce unpredictability and context dependence. Competitive, high-pressure, or resource-limited environments may amplify tensions, potentially limiting the framework's effectiveness. Careful facilitation, contextual adaptation, and iterative reflection are therefore essential to mitigate these challenges without undermining GVP's core principles. Finally, our empirical base is a single, specific case, a small, editorially independent newsroom in Italy. The perception gap and willingness patterns we report may be shaped by this newsroom's small size, editorial independence, and national legal, labor-relations, and press-freedom context. We therefore present GVP as an analytically grounded account proportionate to this case rather than a statistically generalizable one, and treat

its transferability to larger or differently regulated newsrooms as an empirical question for future work (Section 8).

8 Conclusions

Building on the empirical identification of a *perception gap* between nominal inclusion and credible influence, this paper introduced the Gradual Voluntary Participation (GVP) framework as a way of organizing participation across *depth* and *scope*, and articulated five principles for making such participation credible in newsroom AI adoption. Taken together, these contributions frame participation in AI adoption in journalism not as a matter of tool acceptance alone, but as an organizational process shaped by gradualism, voluntariness, and the distribution of voice, legitimacy, and influence. In line with PD evaluation debates that emphasize situated user gains [10] over universal metrics, GVP treats negotiated participation and shifts in how participation is experienced as meaningful outcomes in their own right.

From this perspective, participation in AI should be understood as a multidimensional organizational phenomenon rather than a binary of trust versus distrust. By making layers of influence visible, recognizing voluntary engagement, and documenting outcomes, GVP reframes familiar participatory limitations such as ceilings, burdens, and performative consultation as conditions that must be explicitly addressed in organizing AI adoption. Grounded in self-determination theory, it further suggests that autonomy-supportive participation may foster more constructive orientations toward AI than mandated engagement. GVP also extends discussion of participation in complex organizational settings by treating AI governance as a problem of coordination across differently positioned actors rather than only dyadic human-AI interaction. In doing so, it brings together concerns such as navigation [85], provenance [106], awareness [40], scaffolding [17], and override [3] within a single framework for reasoning about participation where professional roles, situated expertise, and decision-making authority are unevenly distributed, extending feedforward principles [31] from tool use to collective participation.

Methodologically, we derive a conceptual framework from qualitative data rather than presenting a validated intervention [46, 68]. As in prior work that developed conceptual frameworks from situated qualitative inquiry [18, 71, 95], we prioritize transferability through thick description over statistical generalizability, treating GVP as an analytically grounded account of participation under specific organizational conditions.

Future work should extend this contribution in two directions. First, empirical *in-situ* studies in newsroom settings can examine how gradual and voluntary participation shapes journalists' perceptions, experiences of AI, and longer-term adoption practices. Second, while developed for journalism, GVP may provide a useful basis for studying participatory AI governance in other professional domains adopting large language models, especially where influence is distributed across multiple actors, levels of work, and institutional constraints.

Acknowledgments

This work was funded by the EU Horizon Europe program, Marie Skłodowska-Curie Actions (MSCA), alignAI project (grant 101169473). We thank the reviewers for their time and valuable feedback. We also thank our colleagues at the Social Computing Group for their feedback and support throughout this work.

Generative AI Disclosure Statement

In the preparation of this manuscript, large language models were exclusively used for formatting assistance and language support, including designing images, refining paragraph layout and spacing,

and performing proofreading, spelling, grammar, and lexical checks. All substantive content, analysis, and interpretations were conceived and written by the authors.

References

[1] Penelope Muse Abernathy. 2020. Will local news survive? *News deserts and ghost newspapers* (2020).

[2] Iban Albizu-Rivas, Sonia Parratt-Fernández, and Montse Mera-Fernández. 2024. Artificial Intelligence in Slow Journalism: Journalists' Uses, Perceptions, and Attitudes. *Journalism and Media* 5, 4 (Dec. 2024), 1836–1850. doi:10.3390/journalmedia5040111 Publisher: Multidisciplinary Digital Publishing Institute.

[3] Saleema Amershi, Dan Weld, Mihaela Vorvoreanu, Adam Fourney, Besmira Nushi, Penny Collisson, Jina Suh, Shamsi Iqbal, Paul N. Bennett, Kori Inkpen, Jaime Teevan, Ruth Kikin-Gil, and Eric Horvitz. 2019. Guidelines for Human-AI Interaction. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–13. doi:10.1145/3290605.3300233

[4] Laura Amigo and Colin Porlezza. 2025. "Journalism Will Always Need Journalists." The Perceived Impact of AI on Journalism Authority in Switzerland. *Journalism Practice* 0, 0 (2025), 1–19. doi:10.1080/17512786.2025.2487534

[5] Sherry R. Arnstein. 1969. A Ladder Of Citizen Participation. *Journal of the American Institute of Planners* 35, 4 (July 1969), 216–224. doi:10.1080/01944366908977225

[6] Liam Bannon, Jeffrey Bardzell, and Susanne Bødker. 2018. Reimagining participatory design. *Interactions* 26, 1 (Dec. 2018), 26–32. doi:10.1145/3292015

[7] Aleks Berditchevskaia, Erini Malliaraki, and Kathy Peach. 2021. *Participatory AI for humanitarian innovation*. Technical Report. Nesta. https://media.nesta.org.uk/documents/Nesta_Participatory_AI_for_humanitarian_innovation_Final.pdf

[8] Arngeir Berge, Frode Guribye, Siri-Linn Schmidt Fotland, Gro Fonnes, Ingrid H Johansen, and Christoph Trattner. 2023. Designing for control in nurse-ai collaboration during emergency medical calls. In *Proceedings of the 2023 ACM Designing Interactive Systems Conference*. 1339–1352.

[9] Abeba Birhane, William Isaac, Vinodkumar Prabhakaran, Mark Diaz, Madeleine Clare Elish, Iason Gabriel, and Shakir Mohamed. 2022. Power to the People? Opportunities and Challenges for Participatory AI. In *Proceedings of the 2nd ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization* (Arlington, VA, USA) *(EAAMO '22)*. Association for Computing Machinery, New York, NY, USA, Article 6, 8 pages. doi:10.1145/3551624.3555290

[10] Claus Bossen, Christian Dindler, and Ole Sejer Iversen. 2010. User gains and PD aims: assessment from a participatory design project. In *Proceedings of the 11th Biennial Participatory Design Conference* (Sydney, Australia) *(PDC '10)*. Association for Computing Machinery, New York, NY, USA, 141–150. doi:10.1145/1900441.1900461

[11] Tone Bratteteig and Guri Verne. 2018. Does AI make PD obsolete? exploring challenges from artificial intelligence to participatory design. In *Proceedings of the 15th Participatory Design Conference: Short Papers, Situated Actions, Workshops and Tutorial - Volume 2 (PDC '18)*. Association for Computing Machinery, New York, NY, USA, 1–5. doi:10.1145/3210604.3210646

[12] Tone Bratteteig and Ina Wagner. 2012. Disentangling power and decision-making in participatory design. In *Proceedings of the 12th Participatory Design Conference: Research Papers - Volume 1 (PDC '12)*. Association for Computing Machinery, New York, NY, USA, 41–50. doi:10.1145/2347635.2347642

[13] Tone Bratteteig and Ina Wagner. 2016. What is a participatory design result?. In *Proceedings of the 14th Participatory Design Conference: Full papers - Volume 1 (PDC '16)*. Association for Computing Machinery, New York, NY, USA, 141–150. doi:10.1145/2940299.2940316

[14] Virginia Braun and Victoria Clarke. 2019. Reflecting on reflexive thematic analysis. *Qualitative research in sport, exercise and health* 11, 4 (2019), 589–597.

[15] Warren Breed. 1955. Social Control in the Newsroom: A Functional Analysis. *Social Forces* 33, 4 (May 1955), 326–335. doi:10.2307/2573002 Number: 4.

[16] Zana Buçinca, Maja Barbara Malaya, and Krzysztof Z. Gajos. 2021. To Trust or to Think: Cognitive Forcing Functions Can Reduce Overreliance on AI in AI-assisted Decision-making. *Proc. ACM Hum.-Comput. Interact.* 5, CSCW1, Article 188, 21 pages. doi:10.1145/3449287

[17] Parmit K. Chilana, Amy J. Ko, and Jacob O. Wobbrock. 2012. LemonAid: selection-based crowdsourced contextual help for web applications. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Austin, Texas, USA) *(CHI '12)*. Association for Computing Machinery, New York, NY, USA, 1549–1558. doi:10.1145/2207676.2208620

[18] John Joon Young Chung, Shiqing He, and Eytan Adar. 2022. Artist Support Networks: Implications for Future Creativity Support Tools. In *Proceedings of the 2022 ACM Designing Interactive Systems Conference* (Virtual Event, Australia) *(DIS '22)*. Association for Computing Machinery, New York, NY, USA, 232–246. doi:10.1145/3532106.3533505

[19] Besjon Cifliku and Hendrik Heuer. 2025. "This could save us months of work" - Use Cases of AI and Automation Support in Investigative Journalism. In *Proceedings of the Extended Abstracts of the CHI Conference on Human Factors*

in Computing Systems (CHI EA '25). Association for Computing Machinery, New York, NY, USA, Article 29, 8 pages. doi:10.1145/3706599.3719856

[20] Besjon Cifliku and Hendrik Heuer. 2026. They Think AI Can Do More Than It Actually Can: Practices, Challenges, & Opportunities of AI-Supported Reporting In Local Journalism. In *Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems (CHI '26)*. Association for Computing Machinery, New York, NY, USA, Article 968, 20 pages. doi:10.1145/3772318.3791130

[21] Stewart Clegg, Torgeir Skyttermoen, and Anne Live Vaagaasar. 2024. Project leadership and representation: powering purposive social value. *International Journal of Managing Projects in Business* (Dec. 2024). doi:10.1108/IJMPB-05-2024-0119

[22] Hannes Cools and Michael Koliska. 2024. News Automation and Algorithmic Transparency in the Newsroom: The Case of the Washington Post. *Journalism Studies* 25, 6 (April 2024), 662–680. doi:10.1080/1461670X.2024.2326636

[23] Ned Cooper, Tiffanie Horne, Gillian R Hayes, Courtney Heldreth, Michal Lahav, Jess Holbrook, and Lauren Wilcox. 2022. A Systematic Review and Thematic Analysis of Community-Collaborative Approaches to Computing Research. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 73, 18 pages. doi:10.1145/3491102.3517716

[24] Ned Cooper and Alexandra Zafiroglu. 2024. From Fitting Participation to Forging Relationships: The Art of Participatory ML. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 746, 9 pages. doi:10.1145/3613904.3642775

[25] Eric Corbett, Remi Denton, and Sheena Erete. 2023. Power and Public Participation in AI. In *Proceedings of the 3rd ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization* (Boston, MA, USA) *(EAAMO '23)*. Association for Computing Machinery, New York, NY, USA, Article 8, 13 pages. doi:10.1145/3617694.3623228

[26] Robert A. Dahl. 1957. The concept of power. *Behavioral Science* 2, 3 (1957), 201–215. doi:10.1002/bs.3830020303

[27] Edward L Deci and Richard M Ryan. 2012. Self-determination theory. *Handbook of theories of social psychology* 1, 20 (2012), 416–436.

[28] Fernando Delgado, Stephen Yang, Michael Madaio, and Qian Yang. 2023. The Participatory Turn in AI Design: Theoretical Foundations and the Current State of Practice. In *Proceedings of the 3rd ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization* (Boston, MA, USA) *(EAAMO '23)*. Association for Computing Machinery, New York, NY, USA, Article 37, 23 pages. doi:10.1145/3617694.3623261

[29] Shipi Dhanorkar, Christine T Wolf, Kun Qian, Anbang Xu, Lucian Popa, and Yunyao Li. 2021. Who needs to know what, when?: Broadening the Explainable AI (XAI) Design Space by Looking at Explanations Across the AI Lifecycle. In *Proceedings of the 2021 ACM Designing Interactive Systems Conference*. 1591–1602.

[30] Nicholas Diakopoulos and Michael Koliska. 2017. Algorithmic transparency in the news media. *Digital journalism* 5, 7 (2017), 809–828.

[31] Tom Djajadiningrat, Kees Overbeeke, and Stephan Wensveen. 2002. But how, Donald, tell us how? On the creation of meaning in interaction design through feedforward and inherent feedback. In *Proceedings of the 4th conference on Designing interactive systems: processes, practices, methods, and techniques*. 285–291.

[32] Tomás Dodds, Astrid Vandendaele, Felix M Simon, Natali Helberger, Valeria Resendez, and Wang Ngai Yeung. 2024. The Impact of Knowledge Silos on Responsible AI Practices in Journalism. (2024). doi:10.48550/arXiv.2410.01138

[33] Konstantin Nicholas Dörr and Katharina Hollnbuchner. 2017. Ethical challenges of algorithmic journalism. *Digital journalism* 5, 4 (2017), 404–419.

[34] Paul Dourish. 2001. Process descriptions as organisational accounting devices: the dual use of workflow technologies. In *Proceedings of the 2001 ACM International Conference on Supporting Group Work*. 52–60.

[35] Paul Dourish. 2001. *Where the action is: the foundations of embodied interaction*. MIT press.

[36] M. Z. van Drunen and D. Fechner and. 2023. Safeguarding Editorial Independence in an Automated Media System: The Relationship Between Law and Journalistic Perspectives. *Digital Journalism* 11, 9 (2023), 1723–1750. doi:10.1080/21670811.2022.2108868

[37] Tony Dundon, Miguel Martinez Lucio, Debra Howcroft, Emma Hughes, Arjan Keizer, and Roger Walden. 2017. Power dynamics in work and employment relationships: The capacity for employee influence. (2017). https://www.cipd.org/globalassets/media/knowledge/knowledge-hub/reports/power-dynamics-in-work-and-employment-relationships_2017-the-capacity-for-employee-influence_tcm18-33089.pdf

[38] Maximilian Eder and Helle Sjøvaag. 2025. Falling behind the adoption curve: Local journalism's struggle for innovation in the AI transformation. *Journal of Media Business Studies* (2025), 1–19.

[39] Pelle Ehn. 2017. Scandinavian design: On participation and skill. In *Participatory design*. CRC press, 41–77.

[40] Thomas Erickson and Wendy A Kellogg. 2000. Social translucence: an approach to designing systems that support social processes. *ACM transactions on computer-human interaction (TOCHI)* 7, 1 (2000), 59–83.

[41] Hansen et al. 2023. Initial white paper on the social, economic, and political impact of media AI technologies (D2.2). https://www.ai4media.eu/reports/initial-white-paper-on-the-social-economic-and-political-impact-of-

media-ai-technologies-2/
[42] Alexander Fanta and Ingo Dachwitz. 2020. *Google, the media patron: How the digital giant ensnares journalism.* OBS-Arbeitsheft 103. Otto Brenner Foundation (OBS), Frankfurt am Main. https://www.otto-brenner-stiftung.de/fileadmin/user_data/stiftung/02_Wissenschaftsportal/03_Publikationen/AH103_Google_EN.pdf
[43] Michael Feffer, Michael Skirpan, Zachary Lipton, and Hoda Heidari. 2023. From Preference Elicitation to Participatory ML: A Critical Survey & Guidelines for Future Research. In *Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society*. ACM, Montréal QC Canada, 38–48. doi:10.1145/3600211.3604661
[44] Jennifer Fereday and Eimear Muir-Cochrane. 2006. Demonstrating rigor using thematic analysis: A hybrid approach of inductive and deductive coding and theme development. *International journal of qualitative methods* 5, 1 (2006), 80–92.
[45] Gerhard Fischer. 2011. Understanding, fostering, and supporting cultures of participation. *Interactions* 18, 3 (May 2011), 42–53. doi:10.1145/1962438.1962450
[46] Bent Flyvbjerg. 2006. Five misunderstandings about case-study research. *Qualitative inquiry* 12, 2 (2006), 219–245.
[47] Robert Schuman Centre for Advanced Studies. 2020. *Monitoring media pluralism in the digital era – Application of the media pluralism monitor in the European Union, Albania and Turkey in the years 2018-2019 – Report 2020.* European University Institute. doi:doi/10.2870/21728
[48] Nicola K Gale, Gemma Heath, Elaine Cameron, Sabina Rashid, and Sabi Redwood. 2013. Using the framework method for the analysis of qualitative data in multi-disciplinary health research. *BMC medical research methodology* 13, 1 (2013), 117.
[49] José Alberto García Avilés, León , Bienvenido, Sanders , Karen, , and Jackie Harrison. 2004. Journalists at digital television newsrooms in Britain and Spain: workflow and multi-skilling in a competitive environment. *Journalism Studies* 5, 1 (Feb. 2004), 87–100. doi:10.1080/1461670032000174765
[50] Anne Gerdes. 2022. A participatory data-centric approach to AI Ethics by Design. *Applied Artificial Intelligence* 36, 1 (2022). doi:10.1080/08839514.2021.2009222
[51] Lisa Gitelman. 2013. *Raw data is an oxymoron.* MIT press. https://mitpress.mit.edu/9780262518284/raw-data-is-an-oxymoron/
[52] Judith Gregory. 2003. Scandinavian Approaches to Participatory Design. *International Journal of Engineering Education* 19, 1 (2003), 62–74.
[53] Tovi Grossman and George Fitzmaurice. 2010. ToolClips: an investigation of contextual video assistance for functionality understanding. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Atlanta, Georgia, USA) *(CHI '10)*. Association for Computing Machinery, New York, NY, USA, 1515–1524. doi:10.1145/1753326.1753552
[54] Lara Groves, Aidan Peppin, Andrew Strait, and Jenny Brennan. 2023. Going public: the role of public participation approaches in commercial AI labs. In *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency (FAccT '23)*. Association for Computing Machinery, New York, NY, USA, 1162–1173. doi:10.1145/3593013.3594071
[55] Gillian R. Hayes. 2014. Knowing by Doing: Action Research as an Approach to HCI. In *Ways of Knowing in HCI*, Judith S. Olson and Wendy A. Kellogg (Eds.). Springer, New York, NY, 49–68. doi:10.1007/978-1-4939-0378-8_3
[56] Naja Holten Møller, Gina Neff, Jakob Grue Simonsen, Jonas Christoffer Villumsen, and Pernille Bjørn. 2021. Can workplace tracking ever empower? Collective sensemaking for the responsible use of sensor data at work. *Proceedings of the ACM on human-computer interaction* 5, GROUP (2021), 1–21.
[57] Md Naimul Hoque, Ayman A Mahfuz, Mayukha Sridhatri Kindi, and Naeemul Hassan. 2024. Towards Designing a Question-Answering Chatbot for Online News: Understanding Questions and Perspectives. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 154, 17 pages. doi:10.1145/3613904.3642007
[58] Eric Horvitz. 1999. Principles of mixed-initiative user interfaces. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*. 159–166.
[59] Andreas Jungherr, Oliver Posegga, and Jisun An. 2019. Discursive Power in Contemporary Media Systems: A Comparative Framework. *The International Journal of Press/Politics* 24, 4 (2019), 404–425. doi:10.1177/1940161219841543
[60] Christopher M. Kelty. 2017. The Participatory Development Toolkit. *Limn* 9 (Nov. 2017), 110–110. https://limn.press/article/the-participatory-development-toolkit/ Publisher: Limn Press.
[61] Nigel King. 2012. Doing template analysis. *Qualitative organizational research: Core methods and current challenges* 426 (2012), 426–450.
[62] Andrew Konya, Lisa Schirch, Colin Irwin, and Aviv Ovadya. 2023. Democratic Policy Development using Collective Dialogues and AI. doi:10.48550/arXiv.2311.02242
[63] Philip Kraft and Jørgen Bansler. 1994. The collective resource approach: the Scandinavian experience. *Scandinavian Journal of Information Systems* 6, 1 (Jan. 1994). https://aisel.aisnet.org/sjis/vol6/iss1/4

[64] Arne Krumsvik, Tanja Storsul, Lucy Kung, Leyla Dogruel, Gunnar Liestøl, Sabine Baumann, Eli Skogerbø, Gillian Doyle, Sven Lindmark, Heritiana Ranaivoson, Karen Donders, Pieter Ballon, Stine Lomborg, Rasmus Helles, Piet Bakker, Charles Davis, Iris Jennes, Jo Pierson, and Kjartan Müller. 2013. *Media Innovations: A Multidisciplinary Study of Change (2013).* doi:10.13140/2.1.1328.9284

[65] Kalle Kusk and Claus Bossen. 2022. Working with Wolt: an ethnographic study of lenient algorithmic management on a food delivery platform. *Proceedings of the ACM on Human-Computer Interaction* 6, GROUP (2022), 1–22.

[66] Jean Lave and Etienne Wenger. 1991. *Situated Learning: Legitimate Peripheral Participation.* Cambridge University Press.

[67] Leo Leppänen, Myriam Munezero, Mark Granroth-Wilding, and Hannu Toivonen. 2017. Data-Driven News Generation for Automated Journalism. In *Proceedings of the 10th International Conference on Natural Language Generation*, Jose M. Alonso, Alberto Bugarín, and Ehud Reiter (Eds.). Association for Computational Linguistics, Santiago de Compostela, Spain, 188–197. doi:10.18653/v1/W17-3528

[68] Yvonna S Lincoln and Egon G Guba. 1985. *Naturalistic inquiry.* sage.

[69] Juliane A Lischka, Nadja Schaetz, and Anna-Lena Oltersdorf. 2023. Editorial technologists as engineers of journalism's future: exploring the professional community of computational journalism. *Digital Journalism* 11, 6 (2023), 1026–1044. doi:10.1080/21670811.2021.1995456

[70] Airong Luo, Margaret Ann Murphy, and Ted Hanss. 2012. Human infrastructure as process and effect: its impact on individual scientists' participation in international collaboration. In *Proceedings of the 2012 ACM International Conference on Supporting Group Work*. 45–54.

[71] Caitlin Lustig, Maya A Kaneko, Meghna Gupta, Kavita Dattani, Audrey Desjardins, and Daniela Rosner. 2024. Porous by Design: How Childcare Platforms Impact Worker Personhood, Safety, and Connection. In *Proceedings of the 2024 ACM Designing Interactive Systems Conference* (Copenhagen, Denmark) *(DIS '24)*. Association for Computing Machinery, New York, NY, USA, 1336–1349. doi:10.1145/3643834.3661552

[72] Gloria Mark and Steven Poltrock. 2003. Shaping technology across social worlds: groupware adoption in a distributed organization. In *Proceedings of the 2003 ACM International Conference on Supporting Group Work*. 284–293.

[73] Philip M Napoli, Sarah Stonbely, Kathleen McCollough, and Bryce Renninger. 2017. Local journalism and the information needs of local communities: Toward a scalable assessment approach. *Journalism Practice* 11, 4 (2017), 373–395.

[74] Roberto Navigli, Simone Conia, and Björn Ross. 2023. Biases in Large Language Models: Origins, Inventory, and Discussion. *Journal of Data and Information Quality* 15, 2 (June 2023), 1–21. doi:10.1145/3597307 Number: 2.

[75] Elie Neghawi, Zerui Wang, Jun Huang, and Yan Liu. 2023. Linking Team-level and Organization-level Governance in Machine Learning Operations through Explainable AI and Responsible AI Connector. In *2023 IEEE 47th Annual Computers, Software, and Applications Conference (COMPSAC)*. 1223–1230. doi:10.1109/COMPSAC57700.2023.00185

[76] Anna Neumann, Elisabeth Kirsten, Muhammad Bilal Zafar, and Jatinder Singh. 2025. Position is Power: System Prompts as a Mechanism of Bias in Large Language Models (LLMs). In *Proceedings of the 2025 ACM Conference on Fairness, Accountability, and Transparency (FAccT '25)*. Association for Computing Machinery, New York, NY, USA, 573–598. doi:10.1145/3715275.3732038

[77] Karin Niemantsverdriet, Harm van Essen, Minna Pakanen, and Berry Eggen. 2019. Designing for Awareness in Interactions with Shared Systems: The DASS Framework. *ACM Transactions on Computer-Human Interaction* 26 (11 2019), 1–41. doi:10.1145/3338845

[78] Medill School of Journalism and Konrad-Adenauer-Stiftung. 2024. *Impact of AI on local news models.* Technical Report. https://localnewsinitiative.northwestern.edu/assets/research/medill_ai_local_news_report_2024.pdf

[79] Sejin Paik. 2023. Journalism Ethics for the Algorithmic Era. *Digital Journalism* 13, 4 (April 2023), 696–722. doi:10.1080/21670811.2023.2200195

[80] Charis Papaevangelou. 2024. Funding Intermediaries: Google and Facebook's Strategy to Capture Journalism. *Digital Journalism* 12, 2 (Feb. 2024), 234–255. doi:10.1080/21670811.2022.2155206

[81] John Pavlik. 2000. The Impact of Technology on Journalism. *Journalism Studies* 1, 2 (May 2000), 229–237. doi:10.1080/14616700050028226

[82] Marianne Aubin Le Quéré and Maurice Jakesch. 2022. Trust in AI in Under-resourced Environments: Lessons from Local Journalism. In *CHI'22: Workshop on Trust and Reliance in AI-Human Teams*.

[83] Damian Radcliffe and Christopher Ali. 2017. Local news in a digital world: Small-market newspapers in the digital age. *Damian Radcliffe, Christopher Ali* (2017).

[84] Lauge Baungaard Rasmussen. 2004. Action research—Scandinavian experiences. *AI & SOCIETY* 18, 1 (Jan. 2004), 21–43. doi:10.1007/s00146-003-0269-7

[85] Robert W Reeder, Lujo Bauer, Lorrie Faith Cranor, Michael K Reiter, Kelli Bacon, Keisha How, and Heather Strong. 2008. Expandable grids for visualizing and authoring computer security policies. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 1473–1482.

[86] Aimee Rinehart and Ernest Kung. 2022. *Artificial Intelligence in Local News. A survey of US newsrooms' AI readiness.* Technical Report. The Assosciated Press. https://www.amic.media/media/files/file_352_3673.pdf

[87] Jane Ritchie and Liz Spencer. 2002. Qualitative data analysis for applied policy research. In *Analyzing qualitative data.* routledge, 173–194.

[88] Felix M. Simon. 2022. Uneasy Bedfellows: AI in the News, Platform Companies and the Issue of Journalistic Autonomy. *Digital Journalism* 10, 10 (Nov. 2022), 1832–1854. doi:10.1080/21670811.2022.2063150

[89] Felix M. Simon. 2024. *Artificial Intelligence in the News: How AI Retools, Rationalizes, and Reshapes Journalism and the Public Arena.* Technical Report. Tow Center for Digital Journalism, Columbia University. doi:10.7916/ncm5-3v06

[90] Felix M. Simon. 2024. Escape Me If You Can: How AI Reshapes News Organisations' Dependency on Platform Companies. *Digital Journalism* 12, 2 (Feb. 2024), 149–170. doi:10.1080/21670811.2023.2287464

[91] Jesper Simonsen, Connie Svabo, Sara Malou Strandvad, Kristine Samson, Morten Hertzum, and Ole Erik Hansen. 2014. Situated methods in design. *Situated design methods* (2014), 1–21.

[92] Mona Sloane, Emanuel Moss, Olaitan Awomolo, and Laura Forlano. 2022. Participation Is not a Design Fix for Machine Learning. In *Proceedings of the 2nd ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization* (Arlington, VA, USA) *(EAAMO '22)*. Association for Computing Machinery, New York, NY, USA, Article 1, 6 pages. doi:10.1145/3551624.3555285

[93] Alem Febri Sonni, Hasdiyanto Hafied, Irwanto Irwanto, and Rido Latuheru. 2024. Digital Newsroom Transformation: A Systematic Review of the Impact of Artificial Intelligence on Journalistic Practices, News Narratives, and Ethical Challenges. *Journalism and Media* 5, 4 (Dec. 2024), 1554–1570. doi:10.3390/journalmedia5040097

[94] Stefano Sorrentino, Matilde Barbini, and Daniel Gatica-Perez. 2026. FAccT-Checked: A Narrative Review of Authority Reconfigurations and Retention in AI-Mediated Journalism. In *Proceedings of the 2026 ACM Conference on Fairness, Accountability, and Transparency*. Association for Computing Machinery, New York, NY, USA, 2781–2847. https://doi.org/10.1145/3805689.3812300

[95] Katta Spiel. 2021. "Why are they all obsessed with Gender?" — (Non)binary Navigations through Technological Infrastructures. In *Proceedings of the 2021 ACM Designing Interactive Systems Conference* (Virtual Event, USA) *(DIS '21)*. Association for Computing Machinery, New York, NY, USA, 478–494. doi:10.1145/3461778.3462033

[96] Robert Stake. 1995. Case study research. *Cham: Springer* (1995).

[97] Lucille Alice Suchman. 1987. *Plans and situated actions: The problem of human-machine communication.* Cambridge university press.

[98] Yngve Sundblad. 2011. UTOPIA: Participatory Design from Scandinavia to the World. In *History of Nordic Computing 3*, John Impagliazzo, Per Lundin, and Benkt Wangler (Eds.). Springer, Berlin, Heidelberg, 176–186. doi:10.1007/978-3-642-23315-9_20

[99] Harini Suresh, Emily Tseng, Meg Young, Mary Gray, Emma Pierson, and Karen Levy. 2024. Participation in the age of foundation models. In *Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency* (Rio de Janeiro, Brazil) *(FAccT '24)*. Association for Computing Machinery, New York, NY, USA, 1609–1621. doi:10.1145/3630106.3658992

[100] Michael Terry and Elizabeth D. Mynatt. 2002. Side views: persistent, on-demand previews for open-ended tasks. In *Proceedings of the 15th Annual ACM Symposium on User Interface Software and Technology* (Paris, France) *(UIST '02)*. Association for Computing Machinery, New York, NY, USA, 71–80. doi:10.1145/571985.571996

[101] Sina Thäsler-Kordonouri and Michael Koliska. 2025. Journalistic Agency and Power in the Era of Artificial Intelligence. *Journalism Practice* (2025), 1–20.

[102] Emily Tseng, Meg Young, Marianne Aubin Le Quéré, Aimee Rinehart, and Harini Suresh. 2025. "Ownership, Not Just Happy Talk": Co-Designing a Participatory Large Language Model for Journalism. In *Proceedings of the 2025 ACM Conference on Fairness, Accountability, and Transparency (FAccT '25)*. Association for Computing Machinery, New York, NY, USA, 3119–3130. doi:10.1145/3715275.3732198

[103] Nikki Usher. 2021. *News for the rich, white, and blue: How place and power distort American journalism.* Columbia University Press.

[104] Max Van Drunen. 2021. Editorial independence in an automated media system. *Internet Policy Review* 10, 3 (Sept. 2021). doi:10.14763/2021.3.1569

[105] Jo Vermeulen, Kris Luyten, Elise Van Den Hoven, and Karin Coninx. 2013. Crossing the bridge over Norman's Gulf of Execution: revealing feedforward's true identity. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 1931–1940.

[106] Fernanda B. Viégas, Martin Wattenberg, and Kushal Dave. 2004. Studying cooperation and conflict between authors with history flow visualizations. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Vienna, Austria) *(CHI '04)*. Association for Computing Machinery, New York, NY, USA, 575–582. doi:10.1145/985692.985765

[107] Qing Xiao, Xianzhe Fan, Felix Marvin Simon, Bingbing Zhang, and Motahhare Eslami. 2025. "It Might be Technically Impressive, But It's Practically Useless to us": Motivations, Practices, Challenges, and Opportunities for Cross-Functional Collaboration around AI within the News Industry. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*. Association for Computing Machinery, New York, NY, USA, 1–19. doi:10.1145/3706598.3714090

[108] Qing Xiao, Qing Hu, Jingjia Xiao, Hancheng Cao, and Hong Shen. 2026. Can GenAI Move from Individual Use to Collaborative Work? Experiences, Challenges, and Opportunities of Coordinating GenAI into Collaborative Newswork. In *Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems (CHI '26)*. Association for Computing Machinery, New York, NY, USA, Article 1515, 22 pages. doi:10.1145/3772318.3790984

[109] Nur Yildirim, Changhoon Oh, Deniz Sayar, Kayla Brand, Supritha Challa, Violet Turri, Nina Crosby Walton, Anna Elise Wong, Jodi Forlizzi, James McCann, et al. 2023. Creating design resources to scaffold the ideation of AI concepts. In *Proceedings of the 2023 ACM Designing Interactive Systems Conference*. 2326–2346.

[110] Jieyu Zhao, Tianlu Wang, Mark Yatskar, Vicente Ordonez, and Kai-Wei Chang. 2018. Gender Bias in Coreference Resolution: Evaluation and Debiasing Methods. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers)*, Marilyn Walker, Heng Ji, and Amanda Stent (Eds.). Association for Computational Linguistics, New Orleans, Louisiana, 15–20. doi:10.18653/v1/N18-2003

A Supplementary Material: Mapping of Empirical Findings to the Five GVP Principles

In this appendix, Table 3 maps the five GVP principles to their empirical grounding in the interview findings. The table identifies the *dominant* empirical grounding of each principle across perception-gap dimensions and participation patterns. For each principle, it indicates the most relevant perception-gap dimensions (4.2), participation patterns (4.4), and participant statements, and then makes explicit the analytical move through which these situated findings are elaborated into design-level commitments. The table should therefore be read not as a mechanical derivation from data to framework, but as a structured account of how recurring empirical concerns informed the conceptual development of GVP.

Table 3. Mapping of empirical findings to the five GVP principles. The table identifies the dominant empirical grounding of each principle, while also indicating overlap across perception-gap dimensions and the analytical move from situated findings to design-level commitments.

Principle	**Dominant empirical grounding**	**Empirical grounding in the interviews**	**Framework elaboration**
P1: Recognition Without Coercion	Primarily grounded in the *legitimacy of participation demands* dimension of the perception gap (4.2), especially Pattern C, with overlap with Pattern A.	Participants consistently framed willingness to participate as conditional on workload protections, recognition as work, and respect for role boundaries. P6 stated that vendor-facing participation "is extra work and should be compensated/considered working time. I would not want it to be imposed on me as another job". P10 articulated the role-boundary issue even more sharply: "I am a journalist, everything else I consider an appendage to my job, so I would not want to have an important and overly active role in this aspect". Pattern A reinforces this same concern in the governance arena: P10 expressed interest in participating in discussions while insisting on protecting time "for the work I was actually hired to do". P9 similarly described pervasive time scarcity as a barrier across participation venues. The external feedback participation dimension (4.3) provided the most detailed evidence for the specific recognition forms participants proposed, including formal working time allocation, compensation for feedback contributions, and acknowledgment in role descriptions or performance evaluations.	The framework generalizes these conditional-willingness patterns into a design commitment centered on *resourced voluntariness*: participation must remain optional, visibly recognized, and supported by protected time or equivalent organizational acknowledgment, rather than being treated as an informal extension of journalistic duties. The analytical move is from individual resistance to imposed burden and unpaid extra labor (an empirical finding about when participation is perceived as illegitimate) to an organizational design requirement that structures participation so that non-participation carries no professional penalty and contributions are formally recognized.

Table 3. Mapping of empirical findings to the five GVP principles. The table identifies the dominant empirical grounding of each principle, while also indicating overlap across perception-gap dimensions and the analytical move from situated findings to design-level commitments.

Principle	Dominant empirical grounding	Empirical grounding in the interviews	Framework elaboration
P2: Continuous Value Alignment	Primarily grounded in the *procedural agency* dimension of the perception gap (4.2) and in Pattern A, with reinforcement from Pattern B and some overlap with accountability-related concerns later developed in P5.	Participants repeatedly framed participation as normative work concerning standards, rules, and accountability, explicitly distinguishing it from just improving a product or service [P4, P7, P2, P3]. P7 stated: "participation can't just be about the product or making it work better from a technical point of view. It has to be normative, regulatory. Journalists should be involved in setting the rules and must figure this out all together". P4 similarly emphasized the need for "clear guidelines about what's acceptable and how we keep our standards when using these tools". Section 4.3 further showed that internal governance participation concerned defining permissible use cases, disclosure requirements, verification standards, and acceptable boundaries of use. Pattern B reinforces the processual character of this principle: participants valued training and capacity-building only when it was ongoing and institutionally embedded, integrated into working hours, connected to concrete newsroom policies, and sustained through on-the-job experimentation rather than treated as a one-off event [P7, P9]; a preference that directly parallels the logic of continuous rather than episodic alignment.	The framework extends these findings from a demand for normative voice into a design commitment to ongoing alignment: values, standards, and newsroom rules should not be treated as fixed once established, but as matters for periodic re-evaluation and collective revision. This step goes beyond the interview material in a design-oriented way. The analytical move is from journalists' insistence that participation should concern the *substance* of norms (an empirical finding about what participation should be about) to a processual framework requirement in which value negotiation becomes an ongoing organizational practice rather than a one-off consultation event.

Table 3. Mapping of empirical findings to the five GVP principles. The table identifies the dominant empirical grounding of each principle, while also indicating overlap across perception-gap dimensions and the analytical move from situated findings to design-level commitments.

Principle	**Dominant empirical grounding**	**Empirical grounding in the interviews**	**Framework elaboration**
P3: Contextual Knowledge Scaffolding	Primarily grounded in the *epistemic controllability* dimension of the perception gap (4.2) and in Pattern B, with some overlap with legitimacy concerns where access to training was constrained by workload.	Participants emphasized the need for profession-specific knowledge about what AI can and cannot be relied upon for, including normative and practical boundaries [P1, P2]. P1 stated: "What I need are clear guidelines about what I must not do with AI specifically as a journalist. And management needs to commit to training employees". P2 wanted "specific courses on AI in journalism, not generic AI courses …the regulatory and legal side, applied to our work and profession, specifically". P7 preferred practical, on-the-job experimentation ("trying things out with AI, understanding where it fails") over theoretical instruction, while P9 stressed that time scarcity made formal courses difficult to attend. Section 4.3 further documented that participants distinguished journalism-specific training from generic AI literacy and identified existing knowledge silos as barriers to collective participation.	The framework translates these findings into a design commitment to modular, situated, profession-specific scaffolding distributed across multiple learning modalities, including peer learning, tailored guidance, and on-the-job experimentation. The analytical move is from participants' expressed needs regarding training format, content, and accessibility (an empirical finding about the conditions under which epistemic access becomes possible) to a framework requirement that knowledge support be embedded within ongoing work practices and made available through multiple entry points, so that competence does not function as a gatekeeping prerequisite for participation.

Table 3. Mapping of empirical findings to the five GVP principles. The table identifies the dominant empirical grounding of each principle, while also indicating overlap across perception-gap dimensions and the analytical move from situated findings to design-level commitments.

Principle	Dominant empirical grounding	Empirical grounding in the interviews	Framework elaboration
P4: Sovereignty Preservation	Primarily grounded in *procedural agency*, with important overlap with *epistemic controllability* where understanding AI behavior serves as a precondition for exercising editorial authority over its use.	Participants did not evaluate AI merely in terms of output quality, but through whether editorial sovereignty would be preserved and whether they could exercise informed professional judgment over AI-mediated work. Pattern A most directly grounds this principle: participants wanted influence over "what is allowed", "who decides", and "what happens when something goes wrong", framing these not as interface-level preferences but as questions of editorial authority and decisional primacy. Section 4.3 further showed that internal governance participation included demands for clarity about accountability escalation when AI outputs raise editorial concerns. The epistemic controllability dimension reinforces these concerns from a complementary angle: participants asked for practical and normative boundaries, including recognizing failure modes and what the epistemic controllability theme describes as boundaries of acceptable reliance [P1, P2], suggesting that the ability to understand system behavior is a necessary condition for exercising meaningful editorial override.	The framework sharpens these concerns into a requirement that human editorial judgment must retain decisional primacy over AI outputs, with explicit override mechanisms and protected spaces for professional disagreement with the system. The analytical move is from journalists' situated insistence on retaining authority over AI-mediated work (an empirical finding about the conditions under which AI remains professionally acceptable) to a design requirement that systems and workflows encode human override, document divergences from AI suggestions, and preserve editorial primacy throughout implementation.

Table 3. Mapping of empirical findings to the five GVP principles. The table identifies the dominant empirical grounding of each principle, while also indicating overlap across perception-gap dimensions and the analytical move from situated findings to design-level commitments.

Principle	Dominant empirical grounding	Empirical grounding in the interviews	Framework elaboration
P5: Transparent Accountability	Primarily grounded in *procedural agency*, with overlap with *epistemic controllability* where provenance and traceability concerns support the demand for audit mechanisms, and with sovereignty concerns where accountability and editorial control intersect.	Participants consistently linked meaningful participation to knowing who is responsible for decisions, how responsibilities are distributed, and what happens when AI-related failures occur. In Section 4.2, procedural agency was framed as influence over standards and accountability; in Pattern A, participants raised questions about "who decides" and "what happens when something goes wrong". Section 4.3 further specified that internal governance participation included accountability escalation when AI outputs raise editorial concerns, attribution of responsibility, and disclosure requirements when AI is used in published work. The epistemic controllability dimension reinforces this principle from a complementary direction: participants' concerns about provenance, meaning where AI gets its information, what it can and cannot be relied upon for, and how to recognize failure modes [P1, P2], directly support the need for traceable audit mechanisms that make system behavior legible. Additionally, the external feedback participation arena (4.3) extended accountability demands beyond the newsroom: participants who envisioned reporting errors and limitations to vendors were articulating a form of cross-organizational traceability in which accountability structures span institutional boundaries.	The framework formalizes these concerns into transparent accountability mechanisms: traceable decision processes, clear attribution of responsibility, and disclosure structures that make governance visible to differently positioned stakeholders. The analytical move is from journalists' demand for traceability and responsibility in cases of error or contestation (an empirical finding about the conditions of legitimate governance) to a design requirement that decision processes themselves be rendered visible, auditable, and legible across the newsroom and, where relevant, beyond it.

B Supplementary Material: Operationalizing GVP Depth and Scope

In this appendix, Table 4 provides an operationalization of the two dimensions of the GVP framework, *depth* and *scope*, through working definitions, observable indicators, typical decision rights, and example governance artifacts or practices. The table is intended to clarify how positions in the GVP matrix may be interpreted in practice across newsroom AI governance settings. In particular, it specifies what can count as participation at different levels of depth, from more preliminary forms of involvement to more durable infrastructuring arrangements, and across different levels of scope, from local newsroom practices to broader sectoral and public governance arenas.

The table should be read as a lightweight interpretive aid rather than as a rigid classificatory instrument. Its purpose is not to reduce participation to a fixed measurement scheme, but to make explicit the kinds of activities, forms of influence, and governance objects associated with different positions in the matrix. In this sense, it complements the checkerboard representation introduced in Section 7 by clarifying how participation can be located, compared, and discussed analytically. The operationalization draws on prior work on layered participation in foundation-model ecosystems and on multi-level tensions in journalism, which inform respectively the *depth* and *scope* dimensions adapted in GVP.

Table 4. Operationalization of GVP depth and scope levels through working definitions, observable indicators, decision rights, and example governance artifacts. The table adapts the *scope* dimension from prior work on journalism's macro/meso/micro tensions and organizational levels [102], and the *depth* dimension from prior work on layered participation in foundation-model ecosystems, including journalism as a relevant case study [99]. GVP extends these foundations by specifying how participation can be located and compared in newsroom AI governance through concrete activities, artifacts, and forms of influence.

Dimension	Level	Working definition	Observable indicators	Typical decision rights	Example artifacts / practices
Depth	**Policy definition**	Stakeholders articulate concerns, values, red lines, and acceptable-use boundaries for AI in journalism.	Participation in policy discussions; comments on draft rules; identification of prohibited uses; requests for disclosure, attribution, or verification standards.	Can propose, contest, or review norms and acceptable-use rules, but does not directly shape workflow execution or technical implementation.	AI-use policy drafts; acceptable-use checklists; ethics memos; disclosure rules; newsroom charters on AI use.

Table 4. Operationalization of GVP depth and scope levels through working definitions, observable indicators, decision rights, and example governance artifacts. The table adapts the *scope* dimension from prior work on journalism's macro/meso/micro tensions and organizational levels [102], and the *depth* dimension from prior work on layered participation in foundation-model ecosystems, including journalism as a relevant case study [99]. GVP extends these foundations by specifying how participation can be located and compared in newsroom AI governance through concrete activities, artifacts, and forms of influence.

Dimension	Level	Working definition	Observable indicators	Typical decision rights	Example artifacts / practices
Depth	**Process shaping**	Stakeholders influence how AI is incorporated into newsroom workflows, including review procedures, escalation paths, and accountability routines.	Input on verification steps; specification of override points; definition of escalation procedures; participation in training design; allocation of responsibilities across roles.	Can shape workflow conditions, review requirements, and accountability procedures surrounding AI use.	Workflow diagrams; verification checklists; escalation protocols; approval rules; training modules; override procedures.
Depth	**Co-design**	Stakeholders participate in iterating tool use, interaction patterns, prompts, interface requirements, or evaluation criteria at the level of situated use.	Structured feedback on tool behavior; collaborative testing; pilot participation; prompt or template refinement; development of evaluation rubrics for newsroom tasks.	Can materially influence the design or adaptation of tool use at the application layer, without necessarily affecting upstream model development.	Prompt libraries; pilot feedback forms; interface mock-ups; annotated error cases; evaluation rubrics; newsroom-specific usage templates.

Table 4. Operationalization of GVP depth and scope levels through working definitions, observable indicators, decision rights, and example governance artifacts. The table adapts the *scope* dimension from prior work on journalism's macro/meso/micro tensions and organizational levels [102], and the *depth* dimension from prior work on layered participation in foundation-model ecosystems, including journalism as a relevant case study [99]. GVP extends these foundations by specifying how participation can be located and compared in newsroom AI governance through concrete activities, artifacts, and forms of influence.

Dimension	Level	Working definition	Observable indicators	Typical decision rights	Example artifacts / practices
Depth	**Infrastructuring**	Stakeholders shape durable sociotechnical arrangements that persist beyond a single tool or pilot and organize ongoing participation, accountability, or inter-organizational coordination.	Creation of shared repositories; establishment of audit structures; definition of procurement conditions; coordination of cross-newsroom standards; development of long-term governance routines.	Can shape persistent governance infrastructures, organizational memory, and inter-organizational conditions for AI adoption and oversight.	Audit-log schemas; vendor clauses or SLAs; shared repositories; cross-newsroom standards; governance boards; accreditation or certification mechanisms.
Scope	**Micro**	Participation concerns individual or team-level work practices within a newsroom.	Local workflow changes; desk-level use norms; role-specific training; team discussions about acceptable use.	Decisions affect a single role, desk, or team.	Desk-specific checklists; prompt templates; local guidance notes; team review routines.

Table 4. Operationalization of GVP depth and scope levels through working definitions, observable indicators, decision rights, and example governance artifacts. The table adapts the *scope* dimension from prior work on journalism's macro/meso/micro tensions and organizational levels [102], and the *depth* dimension from prior work on layered participation in foundation-model ecosystems, including journalism as a relevant case study [99]. GVP extends these foundations by specifying how participation can be located and compared in newsroom AI governance through concrete activities, artifacts, and forms of influence.

Dimension	Level	Working definition	Observable indicators	Typical decision rights	Example artifacts / practices
Scope	**Meso**	Participation concerns organization-wide coordination or collaboration across multiple desks or newsrooms.	Shared policy formation; inter-desk standards; cross-functional meetings; newsroom-wide training; coordination across organizations.	Decisions affect the newsroom as an organization, or a collaboration among multiple newsrooms.	Organization-wide policies; shared training repositories; joint codes of conduct; common review procedures.
Scope	**Macro**	Participation concerns sectoral, professional, or industry governance beyond a single organization.	Input into professional guidelines; consultation through associations or unions; contribution to sector-wide standards; negotiation of shared governance arrangements.	Decisions affect professional norms, industry standards, or collective bargaining positions.	Professional charters; journalism-association guidance; industry standards; model contractual clauses.

Table 4. Operationalization of GVP depth and scope levels through working definitions, observable indicators, decision rights, and example governance artifacts. The table adapts the *scope* dimension from prior work on journalism's macro/meso/micro tensions and organizational levels [102], and the *depth* dimension from prior work on layered participation in foundation-model ecosystems, including journalism as a relevant case study [99]. GVP extends these foundations by specifying how participation can be located and compared in newsroom AI governance through concrete activities, artifacts, and forms of influence.

Dimension	Level	Working definition	Observable indicators	Typical decision rights	Example artifacts / practices
Scope	**Public**	Participation includes readers, civic actors, or other public-facing accountability processes concerned with the legitimacy of AI use in journalism.	Reader feedback channels; visible provenance or disclosure; public consultation; contestation mechanisms; community oversight.	Decisions affect public-facing transparency, accountability, and legitimacy conditions.	Disclosure templates; provenance markers; public complaint channels; community review procedures; reader-facing audit summaries.

C Supplementary Material: Interview Protocols and Methods

For the empirical component of our study we used two semi-structured protocols: one for journalists and editorial staff (nine participants), and one tailored to the newsroom director (one participant), whose role in setting organizational direction warranted a distinct set of questions on strategy, governance, and policy. Both protocols were administered in person, in the participant's preferred language, and lasted approximately 120 minutes (journalists) and 90 minutes (director).

C.1 Journalist and Editorial Staff Protocol

Background.

(1) Could you tell us about your current role, any prior roles you have held in this newsroom, and how long you have worked here?
(2) What types of stories do you typically cover (e.g., local vs. broader topics; sports, news, culture)?
(3) When you hear about artificial intelligence, particularly applied to journalism, what comes to mind?

Part 1: Current state and awareness.

(1) Do you currently use AI in your work? How - or how do you actively avoid using it?
    - If yes: which systems, and how often (daily, weekly, occasionally)? Internally developed tools, transcription, translation, writing assistants, search/Q&A, image or video tools?
    - If no: what factors keep you from using it, and what conditions or assurances would change that?
(2) Which parts of journalistic work (fact-finding, writing, editing, visualization) do you consider more or less appropriate for AI use?
    - If none: can you explain why?
(3) How does the possibility that sources you use (press releases, social media) may contain AI-generated content affect your work?
(4) Which values do you consider most important in your work as a journalist?
(5) How has AI changed, or how do you think it would change, your daily work? Would the main impact concern efficiency, production volume, fact-checking, or something else?
(6) How have you reacted, or how would you react, to working with AI? What concerns motivate any hesitation, and what would need to change to reduce it?
(7) In your newsroom, how have you prepared, or how would you prepare, before introducing AI (e.g., courses, training, workshops)?

Part 2: Values, identity, and creativity.

(1) How do you think AI might influence your ability to maintain your personal journalistic voice and worldview? What safeguards or practices do you use to preserve your style?
(2) How do you preserve, or how would you preserve, your identity as a journalist when using AI tools - for example, around authorship, reliability, truth-seeking, and personal style?
(3) Do you think AI could stimulate new forms of creativity in your work - in writing, image selection, data visualization, fact-checking, brainstorming, or other practices?
(4) Do you think AI could save journalists time on repetitive tasks and free attention for more creative work? How do you generally evaluate AI-generated text in terms of creativity?
(5) Do you see a risk of flattening or excessive standardization if many journalists begin using AI for their texts, source research, or image generation? How would you address it?
(6) Has AI changed, or would it change, your relationship with readers (trust, certainty of accuracy, etc.)?

(7) As a journalist, how do you perceive the authenticity of a text when content is generated or assisted by AI?
   - Does a fully human article seem more or less authentic than an AI-assisted or AI-generated one?
   - Is AI-produced text perceived as authoritative? Under what conditions (transparency, verification, attribution)?

(8) Do you think you can recognize when a text or image is AI-generated? What cues do you rely on? Do you think the public can recognize it?

(9) Do you think that declaring the use of AI - at the individual or newsroom level -would change readers' trust in your work? What risks or benefits do you see in transparency about AI use?

Part 3: Challenges and responsibilities.

(1) Do you think AI can make mistakes? How do you manage, or would you manage, AI-generated errors in journalistic work? Are there newsroom procedures to identify, correct, or prevent them, or is it left to individual responsibility?

(2) Who do you think should be held responsible for errors in AI-generated content - the newsroom, the individual journalist, or AI developers?

(3) Do you think AI-use disclosure policies should be standardized (nationally or internationally), or left to individual outlets and newsrooms?

(4) Do you think AI produces fair results, or that it can contain biases? What types of bias do you consider most likely in AI-generated journalistic content?

(5) Do you trust AI to be fair? Should journalists act as a critical filter? What evidence informs your view, and what conditions could increase your trust?

(6) Should AI-generated content be subject to different editorial standards than human-created content (e.g., more rigorous checks, restricted topics)?

(7) Do you consider some AI functionalities "safer" than others in journalism, or do you see all uses as comparably risky?

Part 4: Future perspectives and participation.

(1) Which skills do you consider essential today for a journalist who uses, or would like to use, AI in their workflow?

(2) Which human skills do you consider irreplaceable, particularly in journalism?

(3) How would you feel about having an active role in training AI models or shaping implementation strategies? Would you be interested? This could mean:
   - Contributing journalistic data or feedback to improve model quality
   - Providing input on ethical or editorial guidelines for AI use
   - Testing or evaluating newsroom AI tools before adoption
   - Helping design internal policies on AI use

(4) Among these roles, where do you see yourself potentially contributing? What interests you most, and what barriers do you anticipate (lack of opportunity, lack of skills, time, other)?

(5) How do you feel about your published work being used by large technology companies to train AI models? Would you be in favor or opposed? Would you request compensation or some form of recognition?

Closing.

(1) How do you feel about readers encountering your work indirectly through AI - for example through search-engine AI overviews - potentially in a version modified by AI?

(2) Do you think readers should be active participants in shaping the discussion on AI in journalism? Are they currently? Do they have the tools to do so?

(3) What advantages or disadvantages do you see in the use of AI compared to traditional journalistic practices?
(4) Do you think there should be professional or industry-specific standards in journalism guiding how AI is and is not used? Are you aware of any standards, guidelines, or discussions on this topic within your organization or the profession?
(5) Does it make a difference, in your view, if AI-generated or AI-assisted content appears in print or online? Should the rules differ?
(6) Which of the topics we discussed might have different implications, effects, or consequences for local newsrooms or local content?
(7) Is there anything about AI in journalism that we have not discussed that you consider important to mention?

C.2 Newsroom Director Protocol

The director's protocol was organized around six thematic blocks rather than a chronological sequence, reflecting the director's responsibility across strategy, organization, editorial standards, and external relationships.

Theme 1: AI adoption, strategy, organization, and internal preparation.

(1) What do you know about generative AI? Do you use it in your work? What is your view on it?
(2) Are you aware of others in the newsroom experimenting with AI? In what ways?
(3) Which factors most influence - or would most influence - your decision to adopt or avoid AI tools (e.g., costs, efficiency, innovation, reputation, ethical considerations)?
(4) Have you implemented a shared AI policy specifying permitted and non-permitted uses and tools within the newsroom? If not, do you plan to? How would you go about it - through a participatory process with staff, or by issuing direct editorial guidance?
(5) Which areas of newsroom activity (news gathering, editing, distribution, audience relations, business models) do you consider more or less appropriate for AI use? Do you think some AI functionalities are "safer" than others in a journalistic context?
(6) How do you balance AI adoption with the need to preserve your publication's identity?
(7) How have the journalists in your newsroom reacted to working with AI (resistance, enthusiasm, adaptation)?
(8) Have you offered, or are you planning to offer, internal training on AI (e.g., courses, workshops, practical guidelines)?

Theme 2: Creativity, content quality, roles, skills, and individual voice.

(1) In your view, can AI stimulate new forms of creativity in your newsroom's output? Or do you see it primarily as a way to support or accelerate journalists' work (freeing time for creative tasks)?
(2) Considering both content style and diversity of perspectives, to what extent do you see a risk of flattening or excessive standardization - of the publication's overall identity and of individual journalists' voices - as AI use grows? If you consider this risk significant, how could the newsroom address it?
(3) Which human skills do you consider irreplaceable in your newsroom and in journalism more broadly? Will journalists need to develop new competencies to work with AI?
(4) Do you see journalists or editors playing a critical-filter role with respect to AI? In what way?

Theme 3: Editorial standards, errors, responsibility, fairness, bias, and risk management.

(1) Should AI-generated or AI-assisted content be subject to different editorial standards than fully human work (e.g., transparency, fact-checking)?
(2) What risks do you see, if any, for the credibility and brand reputation of your newsroom - and of journalism in general - if AI use expands?
(3) Should a newsroom make the limitations and flaws of AI visible to readers, or conceal them?
(4) Who should ultimately be held responsible for errors in AI-generated or AI-assisted content?
(5) Do you consider AI systems fair in editorial contexts?
(6) When you think about AI systems in an editorial context, which types of bias do you consider most problematic for your work (e.g., political, gender, geographic, linguistic)?
(7) Given that press releases, social content, or other sources may now be AI-generated or AI-rewritten, have you changed - or would you change - your source-verification practices? In what way?

Theme 4: Trust, authenticity, transparency, and the relationship with the audience.

(1) Do you think a newsroom should disclose its use of AI? For all functionalities? How much transparency do you think readers should be guaranteed?
(2) If you already have one, what is your transparency policy regarding AI use?
(3) Do you think disclosing AI-generated content strengthens or weakens trust in your publication?
(4) How do you perceive the authenticity of AI-generated content, and how do you think readers perceive it?
(5) Do you feel able to recognize AI-generated texts or images? Do you think your journalists can? Your audience?
(6) Should readers be active participants in shaping the debate on AI use in journalism?
(7) Does it make a difference, in your view, whether AI-assisted content appears in print or online?

Theme 5: Industry norms, governance, copyright, platforms, business models, and context.

(1) Should standards or regulations be defined at the industry level for AI use in journalism? Who should set them? Should they be uniform across all newsrooms, or left to the judgment of individual outlets?
(2) To what extent do you think AI adoption could reshape the competitive landscape of journalism (e.g., small outlets versus large publishing groups)?
(3) Do you feel threatened by AI models with respect to copyright over your publication's articles? Do you think a deeper agreement with the technology companies developing AI is needed to protect newsrooms from misuse or exploitation of their data? Do you think collaboration among newsrooms is necessary to guard against such risks?
(4) Given your publication's profile, are there aspects of AI use that you consider specific or particularly critical to your context? In what sense?

Theme 6: Participatory design and co-creation with journalists and readers.

(1) Would you be open to adopting a participatory design approach for integrating AI in the newsroom, actively involving journalists in the process?
(2) Do you think this could help mitigate dissatisfaction or resentment among staff?
(3) Do you see other benefits? Would you actively involve readers in defining policies on AI models?

C.3 Analytic Approach

Our analysis followed a hybrid deductive-inductive thematic approach, in the tradition of framework- and template-based qualitative analysis[44, 48, 61, 87]. We adopted this orientation deliberately, in line with the design of the broader study: our interview protocol was developed on the basis of a prior literature review [94] on participation in AI adoption, editorial authority in journalism, and organizational dynamics around newsroom technology, which surfaced a set of conceptual domains we sought to investigate empirically. The protocol operationalized these domains into open-ended prompts and questions (see Appendix C), inviting participants to articulate their understandings, concerns, and conditions within and beyond the questions posed. Coding and theme development therefore combined a deductive layer, in which interview material was organized in relation to the conceptual domains that motivated the study, with an inductive layer, in which finer-grained patterns, tensions, and analytic constructs were developed from the participants' own accounts.

Coding combined semantic and latent registers: semantic codes captured what participants explicitly stated, while latent codes captured underlying assumptions and orientations interpreted from the material.

The analysis was conducted collaboratively by the two PhD researchers who jointly conducted the interviews. Both researchers engaged closely with the full corpus and developed interpretations through repeated discussion. Consistent with the qualitative philosophy of framework- and template-based approaches, the involvement of two analysts served a reflexive function: joint engagement was used to sense-check interpretations, surface alternative readings, and enrich the analysis through dialogue[14].

The analytic process unfolded in three broad movements, which were recursive and not strictly sequential. First, the two researchers familiarized themselves with the transcripts through repeated reading, organizing initial observations in relation to the conceptual domains that had motivated the protocol (e.g., internal governance and policy participation, capacity-building and training, vendor-facing participation, editorial sovereignty and accountability, authenticity and trust). Second, within and across these domains, the researchers identified recurring patterns in how participants articulated the conditions under which participation would or would not be experienced as meaningful, legitimate, or burdensome. These patterns were discussed, refined, and reorganized through analytic conversations between the two researchers. The three dimensions of the *perception gap* (Section 4.2) -epistemic controllability, procedural agency, and legitimacy of participation demands-and the three patterns of conditional willingness (Patterns A-C; Section 4.4) emerged at this stage as analytic constructs developed through interpretive engagement with the corpus. Third, the researchers worked from these constructs back to one of the conceptual question that motivated the study, examining how the identified patterns spoke to existing frameworks for participation in AI adoption[5, 99, 102] and where they suggested the need for an alternative formulation. This third movement informed the development of the GVP framework's two-dimensional structure (Section 5) and its five organizational principles (Section 6), with the analytic mapping from empirical findings to principles documented in Appendix A.

Following the qualitative tradition we draw on, we make no claim that another analytic team would have produced an identical account; the analysis reflects the researchers' situated interpretation of the material at the intersection of the corpus, the conceptual domains that framed the present inquiry, and the analytic sensibilities as researchers.

Received 31 March 2026; revised 17 July 2026; accepted 17 September 2026