Human-Centered AI for Democracy

Ilaria Mariani, Sabrina Sacco, Francesca Rizzo, Danai Kyrkou, Erilda Goga, Georgia Panagiotidou, Theodore Chadjipadelis, Christos Gogos, Alexandros Tzallas, Xanthi Papageorgiou, Jennifer Edmond, Elizabeth Calderón Lüning, Otilia Kocsis, Maria Zangl, Eva de Lera, and Michael A. Bedek

Abstract This chapter explores the development of human-centered AI (HCAI) systems as a means to support and enhance democratic processes. Building on earlier discussions of data governance, participation, and ethics, it examines how AI technologies can be shaped by democratic values, citizen needs, and participatory design

I. Mariani (✉) · S. Sacco · F. Rizzo
Department of Design, Politecnico di Milano, Milan, Italy
e-mail: ilaria1.mariani@polimi.it

S. Sacco
e-mail: sabrina.sacco@polimi.it

F. Rizzo
e-mail: francesca.rizzo@polimi.it

D. Kyrkou · E. Goga
ViLabs LTD, Limassol, Cyprus
e-mail: danaikyrkou@vilabs.eu

E. Goga
e-mail: eri.goga@vilabs.eu

G. Panagiotidou · T. Chadjipadelis
Department of Political Sciences, Aristotle University of Thessaloniki, Thessaloniki, Greece
e-mail: gvpanag@polsci.auth.gr

T. Chadjipadelis
e-mail: chadji@polsci.auth.gr

C. Gogos · A. Tzallas
Department of Informatics and Telecommunications, University of Ioannina, Ioannina, Greece
e-mail: cgogos@uoi.gr

A. Tzallas
e-mail: tzallas@uoi.gr

X. Papageorgiou
UBITECH Ltd., Limassol, Cyprus
e-mail: xpapageorgiou@ubitech.eu

J. Edmond
Trinity College Dublin, Dublin, Ireland
e-mail: edmondj@tcd.ie

I. Mariani et al. (eds.), *AI for Democracy*, Springer Series in Design and Innovation 66,
https://doi.org/10.1007/978-3-032-23948-8_5

methods. The chapter first outlines the theoretical underpinnings of HCAI and its relevance for democratic governance, drawing on and extending the conceptual foundations proposed by Régis et al. (Human-centered AI: a multidisciplinary perspective for policy-makers, auditors, and users, Chapman and Hall/CRC, Boca Raton, FL, 2024) to build a framework for systematically analyzing AI-enhanced democracy. It emphasizes how co-creation and participatory design approaches enable diverse stakeholders to actively shape the goals, values, and operational logic of AI systems, ensuring that resulting technologies reflect pluralistic and inclusive perspectives. Central to this vision is the concept of 'democracy-in-the-loop,' which calls for embedding democratic principles into the design, governance, and deployment of AI. Moving beyond the state of the art, the chapter then examines how this vision is interpreted and operationalized across four Horizon Europe projects—ORBIS, KT4D, AI4Gov, and ITHACA—each addressing specific challenges and research gaps in the application of HCAI within democratic innovation.

Keywords Human-centered AI · Participatory design · Co-creation · Democracy-in-the-loop

1 A Human-Centered Approach to AI for Democracy

The integration of Artificial Intelligence (AI) into democratic processes marks a significant shift in how citizen engagement is conceptualized and operationalized, with the potential to increase efficiency, inclusiveness, and responsiveness in democratic systems (Sieber et al. 2024). AI increasingly intervenes in how public input is structured, interpreted, and translated into decision-making processes. It functions as a driver for participatory governance by enabling novel modes of interaction between citizens and institutions and by expanding civic input through advanced

E. C. Lüning
Democratic Society, Brussels, Belgium
e-mail: elizabeth.cl@demsoc.eu

O. Kocsis · E. de Lera
Raising the Floor, Brussels, Belgium
e-mail: otilia@raisingthefloor.org

E. de Lera
e-mail: eva@raisingthefloor.org

M. Zangl · M. A. Bedek
University of Graz, Graz, Austria
e-mail: maria.zangl@uni-graz.at

M. A. Bedek
e-mail: michael.bedek@uni-graz.at

computational methods ranging from natural language processing (NLP) to sensor-based data collection (Ghahramani et al. 2021). These innovations include techniques such as real-time translation, automatic key points extraction, sentiment analysis, thematic clustering, and mechanisms for interpreting public feedback—which collectively contribute to enhancing the interface between citizens and public institutions (Fernández-Martínez et al. 2018; Galais et al. 2021). In this setting, there is increasing recognition that deeper civic engagement can not only benefit the functioning of democratic systems but also lead to improved design and user experience in AI applications themselves (Leslie et al. 2021; Straub et al. 2023). Nowadays, a well-rooted stream of research and innovation concerns AI's promise for facilitating more informed, diverse, and inclusive decision-making processes. This direction has been deeply investigated adopting participatory AI design, actively centering community members and impacted stakeholders throughout the process. Evidence from digital government studies highlights the increasing adoption of AI to support more direct citizen involvement, particularly in co-producing and disseminating knowledge (Reis et al. 2019; Simonofski et al. 2017; Wilson 2022). Within the Civic Tech literature, particular attention is given to how AI is used to process large volumes of public input, thus helping to amplify citizen voices (Skaržauskienė and Mačiulienė 2020). This evolution is indicative of a broader shift: on one side, toward using AI to enhance collaboration in public service delivery and policy development (Duberry 2022; Zuiderwijk et al. 2021), such as harnessing social media data to support collective intelligence initiatives (Ghahramani et al. 2021); and on the other, toward more inclusive design practices and stakeholder participation in AI governance itself (Cath et al. 2018; Sousa et al. 2019). Simonofski et al. (2017) articulate multiple civic roles redefined in this new context—from engaged citizens to co-designers and service users—underscoring how digital technologies are reshaping traditional patterns of participation. These approaches are seen as a corrective to the limitations of past governance models, promoting more collaborative and inclusive processes. Still, despite widespread acknowledgment of the importance of engaging citizens and diverse stakeholders in shaping AI systems for public deliberation, the literature lacks depth on how such involvement is practically implemented (Zuiderwijk et al. 2021). Zhang et al. (2023) reflect this orientation in their description of the 'Deliberating with AI' tool—an online, machine learning-supported platform that promotes participatory design and stakeholder deliberation. Rather than replacing human input, this tool is meant to provoke critical reflection, helping participants identify values, tensions, and trade-offs in algorithmic governance. It invites users to collectively define evaluation criteria and shape the logic of AI models, positioning AI not as a decision-maker, but as a facilitator in democratic contexts.

Yet, the integration of AI into democratic, deliberative, civic context is not without its challenges and concerns—not only technical, but normative, political, and epistemic.

The deployment of AI in public contexts raises pressing concerns regarding transparency, inclusivity, accountability, and potential algorithmic discrimination (Manias et al. 2023). These issues are especially critical in the public sector, where AI systems may unintentionally reinforce social inequalities, limit representational diversity, or obscure decision-making logic in ways that hinder public trust. The opacity inherent in many AI systems may exacerbate divides between expert communities and the general public, undermining inclusive and equitable governance (Sachdeva et al. 2015; Sieber et al. 2024) and may also hinder citizens' ability to fully participate in or critically assess these technologies. Scholars such as Buhmann and Fieseler (2023) call attention to the need for a deeper investigation into the distribution of power across stakeholder groups, emphasizing the pronounced knowledge asymmetries and epistemic divides that separate AI experts from the general public. They also question the often opaque influence of private AI companies in setting the terms of access and engagement, arguing for a comprehensive, multi-dimensional governance approach. Advances in AI have also often resulted in a surge in human rights violations, which disproportionately impact marginalised communities (Raso et al. 2018). Accordingly, algorithmic bias is understood as being socially produced and merely amplified by AI, rather than being solely technologically produced. In this setting, current normative solutions, such as ethical frameworks or principles, are seen as tackling technological challenges rather than societal ones. These approaches often protect entitlements defined through top-down processes that do not fully respond to citizens' lived experiences and needs, anticipating the need for more participatory and process-oriented models of governance. In this regard, Gentelet and Mizrahi (2024) argue that current approaches to AI governance fail to equip affected individuals with the capabilities to meaningfully mobilize their rights. Rather than focusing solely on technical risks or outcomes, they advocate for a process-based approach that centers participation, intersectionality, and the socio-political dimensions of human identity.

In response to these limitations, a human-centered approach has been proposed to better support democratic practices.

To ensure that AI enhances human experience and aligns with societal goals, the concept of Human-Centered AI (HCAI) emerges as a compelling approach in circumstances where AI is applied merely because data is available, without considering the benefits to the data subjects—a critique grounded in early reflections on the ethics of information and the risks of data-driven decision-making (Floridi 2005; Nissenbaum 2004). A central tenet is the idea that AI systems should be designed, built, and deployed in a way that adheres to human values and standards, with goals centred around human flourishing (Vallor 2016). The perspective centers on placing AI under human control, while respecting human autonomy and accountability (Régis et al. 2024; Shneiderman 2022). It revolves around the primacy of human dignity, rights, and agency in the design, governance, and deployment of AI technologies. In democratic contexts, this entails treating AI not only as a technical system to be optimized, but as a socio-technical mediator of voice, agency, and legitimacy. The concept thus refers to the design and deployment of technologies that not only enhance human capabilities—such as improving efficiency, empowerment, and decision-making—but also uphold core human values (Shneiderman 2022, p. 4). This perspective is echoed in the European Commission's Ethics Guidelines for Trustworthy AI, developed by the AI High-Level Expert Group (AI HLEG), which emphasizes that a human-centric approach must be grounded in the principle of human dignity (European Commission et al. 2019). Unlike traditional technology development, which might primarily serve institutions, economies, or markets, HCAI is fundamentally about ensuring alignment between those who develop the technology and the humans who are at the core of its impact—an idea strongly supported by theories of participatory governance and inclusive design (Dignum 2019). This means alignment of technology with human interests and needs, as articulated in frameworks such as value-sensitive design (Friedman et al. 2008) and responsible innovation (Stilgoe et al. 2013).

Understanding the need for HCAI approach is also the reflection of Buhmann and Fieseler (2023) on what they term 'responsible AI innovation', a model that advocates for a more profound democratic conversation about AI's societal role and consequences. This entails engagement from all corners of the quadruple helix (Carayannis and Campbell 2010; Schütz et al. 2019)—government, academia, industry, and civil society—allowing for a critical appraisal of AI's purposes and effects. Responsible AI innovation involves not only managing technical and operational risks but also ensuring ethical integrity in the design and deployment of AI systems. It foregrounds the abovementioned values of transparency, fairness, and accountability, requiring mechanisms to trace, audit, and regulate AI use while still encouraging innovation and high performance (Goodman and Flaxman 2017; Herrmann 2023; Voegtlin and Scherer 2017). Within this framework, collaborative participation is widely acknowledged as a cornerstone for building public confidence and fostering ethical innovation. Nevertheless, the complex nature of AI systems often obstructs meaningful public scrutiny or participation, reinforcing existing power asymmetries and epistemic gaps between technical experts and lay citizens. This concern is compounded by the potential dominance of private sector actors in shaping access, framing narratives, and controlling the infrastructure behind AI-driven participatory processes (Buhmann and Fieseler 2023). As such, a human-centered approach to AI is needed to address not only data quality and bias but also questions of legitimacy, access, and knowledge production. In this context, sustained and inclusive engagement is increasingly acknowledged as essential for promoting responsible AI innovation and building public trust. Sieber et al. (2024) advocate for an interdisciplinary strategy in embedding AI into democracy, one that not only includes citizens as participants in discussion forums (Ekman and Amnå 2012) but also empowers them to co-create the digital tools and platforms that support such participation. A meaningful integration of AI into democratic processes demands a more proactive role for civil society in both the design and implementation of such systems. This chapter therefore adopts HCAI as a lens for examining how AI can be integrated into democratic processes in ways that foreground participation, legitimacy, and deliberative quality—setting the foundation for the comparative and governance-oriented analyses developed in the following sections.

2 Current Challenges and Research Gaps

The current discourse on the application of HCAI in democratic contexts primarily revolves around the articulation of principles, the development of conceptual frameworks, and experimental efforts aimed at embedding human values, rights, and participatory mechanisms into AI governance and public policy. Despite promising developments, fully implemented and widely adopted applications of HCAI in democratic settings remain scarce—or, at the very least, poorly documented. Much of the literature emphasizes potential and design rather than deployment and long-term impact, creating a disconnect between theoretization and practical realization. As this section argues, these limitations are not merely technical but directly affect core democratic functions, affecting the capacity to scale participation without eroding deliberative quality. Against this background, the section identifies a set of interrelated challenges and research gaps that must be addressed to meaningfully operationalize HCAI within AI-enhanced democratic systems.

Gap 1. Epistemic asymmetries and transparency deficits. A fundamental challenge lies in the epistemic asymmetry between AI developers and the public. The inherent opacity of many AI systems—often described as their black-box nature (Rudin 2019; von Eschenbach 2021)—makes it difficult for non-experts to understand how these technologies function or to scrutinize their logic. While many initiatives aim to align AI with democratic values, existing deployments have often reinforced biased decision-making, especially in high-stakes public sector contexts. This undermines democratic oversight and exacerbates existing inequalities in digital literacy and access. Citizens are frequently relegated to the role of end-users, with little ability to influence or challenge algorithmic decisions that affect their lives (von Eschenbach 2021). This knowledge imbalance is further compounded when municipalities outsource AI development to private companies, often without establishing clear participatory mechanisms or accountability structures. As Wan and Sieber (2023) highlight, disciplinary confinement and sector-specific knowledge leads to underexplored questions about the role of software engineers and computer scientists in shaping AI governance ecosystems—particularly in cases where AI development is outsourced by municipalities, which can sideline the diverse perspectives necessary for comprehensive governance strategies.

Gap 2. Ethical and representational gaps in system design. While normative calls for responsible AI have proliferated, there remains a significant gap between ethical aspirations and implementation practices. Issues such as algorithmic bias, inadequate representation, and lack of traceability continue to characterize AI applications in public services. These problems are especially acute in high-stakes contexts—such as welfare eligibility or predictive policing—where biased outcomes can have severe consequences for marginalized groups (Buhmann and Fieseler 2023). Although ethical frameworks advocate for fairness, transparency, and accountability, there is limited evidence of their consistent integration into the design and deployment of AI systems. Explainability remains a particularly difficult hurdle in AI-enhanced deliberations, with recent studies underscoring the technical and communicative limitations of current methods (Coussement et al. 2024; Siachos and Karacapilidis 2024).

Gap 3. Fragmentation of knowledge and disciplinary silos. A third challenge is the fragmented nature of research on AI and democracy, with limited focus on how human-centered approaches are applied (Duerinckx et al. 2024). Despite increasing interest in this domain, much of the academic discourse remains confined within disciplinary silos. For instance, urban planning scholars may explore civic engagement through smart cities, while software engineering focuses on technical robustness—yet cross-disciplinary integration remains rare. As a result, comprehensive perspectives that integrate technical, social, ethical, and institutional dimensions are lacking (Duberry 2022; Wilson 2022). This fragmentation also leads to a lack of attention to critical actors—such as civic technologists, civil society organizations, or frontline public servants—who could help bridge the gap between AI innovation and democratic legitimacy, from theory to practice. Moreover, the fragmentation across research disciplines, implementation strategies, and governance practices contributes to a limited body of documented human-centered approaches to AI-enabled or AI-enhanced applications in democratic contexts.

Gap 4. Limited operationalization of participatory design. While public participation and co-creation are widely promoted in policy documents and scholarly literature, there is limited empirical research on how these processes actually shape AI systems. Reviews of AI in civic participation (Reis et al. 2019; Savaget et al. 2019; Zuiderwijk et al. 2021; Sieber et al. 2024) consistently highlight a gap in involving citizens and stakeholders in the design, testing, and evaluation of AI tools for public deliberation. The literature on co-creating AI for democratic deliberation is still in its formative stages (Mariani et al. 2026), with few robust frameworks for integrating diverse user perspectives into the technical development process. Duerinckx et al. (2024) reinforce this view, noting the scarcity of focused studies on how co-design and participatory methodologies can substantively improve deliberative democratic outcomes.

Gap 5. Technical integration and scalability limitations. Another major structural barrier is the poor integration of AI tools within existing civic technology ecosystems, coupled with challenges in scaling these tools beyond the specific contexts for which they were initially designed. Many governments—especially at the local level—grapple with limited technical capacity, rigid procurement frameworks, and fragmented IT infrastructures, all of which hinder the seamless deployment of HCAI systems. Even when AI-enhanced tools prove effective in pilot settings or under ideal development conditions, translating them into broader, real-world applications often encounters resistance due to differences in institutional maturity, policy contexts, and resource availability. The modular or bespoke nature of many civic tech solutions means they are tightly coupled to the conditions of their original implementation, making reuse, adaptation, and upscaling across different domains or jurisdictions especially difficult. Without deliberate efforts to design for interoperability, modularity, and adaptive governance, the scalability of HCAI tools remains limited—jeopardizing their potential to contribute to systemic democratic innovation across diverse governance environments.

3 Linking the HCAI Discourse to Democracy

The academic discourse on HCAI has rapidly expanded in recent years, yet its application to democratic contexts—particularly deliberative democracy—remains relatively underdeveloped. While various studies have explored AI in public services, civic tech, or policy design, there is still a limited body of literature that explicitly connects the principles of HCAI to democratic governance and citizen participation. To advance this conversation, this chapter begins by exploring the broader conceptual foundations of HCAI, and then progressively applies these insights to the democratic domain. To carry out this conceptual transfer from general HCAI principles to the specific context of democracy, this chapter relies on a qualitative, interpretive methodology. First, key perspectives and frameworks within the HCAI discourse are identified through the identification of foundational interdisciplinary literature that addresses ethical, societal, and governance-related dimensions of AI. These contributions are then analysed to extract core principles that define the HCAI paradigm. Finally, these principles are mapped onto the needs and challenges of democratic systems, particularly in relation to deliberative processes and civic participation.

3.1 HCAI as a Socio-technical Construct

Among the pioneering contributions to the HCAI discourse, Régis et al. (2024) stands out as a timely and foundational work which consolidates a diverse set of interdisciplinary perspectives on the ethical, societal, and institutional dimensions of AI. Rather than limiting itself to calls for more responsible or explainable AI, the chapter proposes a deeper reframing of the field: it positions HCAI as both a normative ideal and an operational paradigm that places human dignity, agency, and flourishing at the center of AI systems and their applications. Given its conceptual breadth—spanning theoretical foundations to practical implementations in sectors such as healthcare, education, and governance—the work of Régis and colleagues is here adopted as a key reference for identifying the conceptual principles, trends, and design approaches that structure the evolving landscape of HCAI. These principles are especially relevant to democratic innovation and deliberation, where the question of human-centeredness is deeply tied to collective agency, institutional accountability, and participatory governance.

The theoretical framework that emerges from the curated volume of Régis et al. (2024) conceptualizes HCAI as a socio-technical construct—one that necessitates critical engagement with power asymmetries, value pluralism, and the lived experiences of diverse user groups. In this light, HCAI emerges not only as a design ethos but also as a political project that challenges the dominant imperatives of efficiency and optimization in AI development. It instead emphasizes values such as equity, inclusiveness, transparency, and democratic legitimacy. Furthermore, their work situates HCAI within broader debates on the governance of emerging technologies, highlighting the need for institutional transformation to ensure that AI aligns with public interest and societal priorities. This includes advancing epistemic plurality—the idea that legitimate knowledge about what AI should do must be co-produced with communities and publics—and strengthening deliberative capacity, where citizens can engage meaningfully with the normative choices, trade-offs, and risks embedded in AI systems. Crucially, Régis et al. (2024) also argue that HCAI cannot be reduced to technical add-ons or post-hoc ethical audits. Rather, it calls for participatory governance structures in which stakeholders—including marginalized and affected communities—are directly involved in shaping AI development, deployment, and oversight.

To examine how Human-Centered AI is concretely applied within democratic settings, this chapter builds on the theoretical contributions of Régis et al. (2024) to construct an analytical framework for systematically interrogating how HCAI is enacted within AI-enhanced democratic systems (Table 1). Building on this foundation, the chapter unpacks the diverse contributions in Régis et al. (2024) into a coherent set of analytical dimensions that characterize HCAI across design, governance, and socio-technical practice. These dimensions are identified through a process of conceptual clustering, resulting in an integrated framework that captures the core orientations of human-centeredness without reducing them to a prescriptive checklist. Adopted as an analytical lens, this framework enables a focused examination of AI applications in democratic contexts, making it possible to assess how participatory, rights-based, and deliberative principles are translated into concrete system features, governance arrangements, and practices of use. Importantly, this approach also reveals how much of the existing literature and practice in AI for democracy already embodies HCAI commitments, even when these are not explicitly articulated in human-centered terms.

Table 1 Mapping foundational HCAI theoretical dimensions from Régis et al. (2024) in the domain of AI-enhanced democracy

Dimensions	Description	Main references in Régis et al. (2024)	Relevance for AI-enhanced democracy	Mapped to deliberative and participatory democracy needs
D1. Participatory design and co-creation	Centers stakeholders throughout AI system design and lifecycle, enabling diverse, context-sensitive input and co-creation. Emphasizes iterative inclusion of marginalized voices	Costanza-Chock (2020), Shneiderman (2022)	• Empowers stakeholders to co-create, ranging from the scale of solutions to that of governance frameworks • Enhances legitimacy through procedural inclusion and responsiveness	Supports inclusive agenda-setting and shared ownership of deliberative tools, ensuring processes are designed with plural stakeholder input
D2. Civic participation and democratic engagement	Promotes citizens' voluntary and informed participation in AI design and governance. Advocates for inclusive, accessible participation channels rooted in rights	Gentelet and Mizrahi (2024), Teller (2022), Zimmermann et al. (2020)	• Ensures democratic engagement with AI systems • Grounds participation in lived experience and civic rights, reducing exclusion	Fosters civic empowerment and direct involvement in shaping AI's role in public discourse, enhancing both voice and agency
D3. Human rights-based approach	Uses human rights as a legitimizing framework across the AI lifecycle, promoting dignity, entitlements, and capability-based approaches to justice	Council of Europe (n.d.), European Commission et al. (2019), UNICEF (2021)	• Anchors AI in legal and ethical principles essential to democratic legitimacy • Helps protect vulnerable populations and enable civic agency	Establishes rights-based boundaries in public-facing AI, protecting civic freedoms and ensuring accountability in public deliberation
D4. Decolonial ethics	Positions HCAI within decolonial and relational ethics, emphasizing interdependence, lived experiences, and community-centered governance of AI, opposing epistemic domination	Birhane (2021), Bohler-Muller Narnia (2005), Liebenberg (2012),Mbiti (1990)	• Challenges epistemic injustice in AI • Brings historically excluded voices and collective knowledge traditions into AI governance	Counters power imbalances in civic tech by including epistemologies often excluded from mainstream policy debates

(continued)

Table 1 (continued)

Dimensions	Description	Main references in Régis et al. (2024)	Relevance for AI-enhanced democracy	Mapped to deliberative and participatory democracy needs
D5. Adaptive governance	Supports flexible, iterative rule-making aligned with societal values. Involves multi-stakeholder dialogue and continuous risk responsiveness instead of static regulation	CAO—Japan's Cabinet Offce (2019), Manifesto and Beck (2001, p. 200)	• Aligns governance with democratic responsiveness and public values • Facilitates inclusive rule-making and agile consensus-building	Allows public deliberation frameworks to evolve responsively with societal input, maintaining alignment with changing civic needs
D6. Policy evaluation and impact assessment	Applies systematic evaluation to assess AI policy alignment with HCAI principles like fairness, trust, and sustainability. Guides iterative improvement in governance	Kappenberger et al. (2022), Vedung (2020), Wollmann (2017)	• Improves transparency, fairness, and equity by evaluating AI outcomes • Enables data-informed governance and democratic oversight	Provides mechanisms to measure whether deliberative processes lead to fair and just outcomes, validating democratic performance
D7. Socio-technical systems and human factors	Analyzes AI within real-world work environments. Considers the physical, cognitive, and organizational conditions of users, highlighting worker agency and system integration	Cabour et al. (2023), Rasmussen (1985), Vicente (1999)	• Strengthens democratic input in workplace technologies • Prioritizes worker well-being and participatory governance in sociotechnical design	Ensures that AI supports—not replaces—deliberative engagement by adapting to user capacities and organizational cultures
D8. Transparency and algorithmic explainability	Focuses on making AI systems understandable to various publics. Addresses the trade-off between transparency and communicative clarity to support oversight	Larsson and Heintz (2020), Lepri et al. (2021), Miller (2019)	• Supports public scrutiny and debate by making AI operations intelligible • Contributes to informed civic deliberation	Makes complex deliberation outcomes intelligible to all participants, thereby increasing transparency and trust in democratic tools

(continued)

Table 1 (continued)

Dimensions	Description	Main references in Régis et al. (2024)	Relevance for AI-enhanced democracy	Mapped to deliberative and participatory democracy needs
D9. Accountability and oversight mechanisms	Establishes institutional and social mechanisms for AI responsibility. Includes internal ethics practices and tools to facilitate public scrutiny and adaptive correction	Gebru et al. (2021), Mitchell et al. (2019), Raji et al. (2020)	• Enables public and institutional checks on AI power • Facilitates democratic accountability through adaptive governance mechanisms	Supports institutional legitimacy by embedding checks and balances in the deliberative process, reinforcing public control over AI
D10. Algorithmic legitimacy and democratic governance	Emphasizes procedural legitimacy and public reasoning in algorithm governance. Encourages citizen influence over digital norm-setting and algorithmic design	Bersini and Badinet (2023), Castelluccia and Le Métayer (2019), Danaher (2016)	• Advances participatory democracy by involving citizens in algorithmic norm development • Reinforces legitimacy and inclusive policymaking	Legitimizes digital deliberation through procedural fairness and civic co-authorship of algorithmic rules and evaluation criteria

Synthesis table mapping ten foundational theoretical dimensions of Human-Centered AI, drawn from Régis et al. (2024), to the domain of AI-enhanced democracy. For each dimension—ranging from participatory design and civic engagement to human rights, decolonial ethics, adaptive governance, explainability, accountability, and algorithmic legitimacy—the table summarises its conceptual grounding, key references, relevance for democratic governance, and implications for deliberative and participatory democracy.

4 HCAI for Democratic Innovation: Four Horizon Europe Projects

The following sections examine how the four Horizon Europe projects contributing to this volume—ORBIS, KT4D, AI4GOV, and ITHACA—respond to the current challenges and research gaps identified in the application of HCAI for democratic innovation (Table 2). Each project presents a distinct approach to integrating AI into democratic processes, often addressing multiple challenges simultaneously. Through co-creation methodologies, technical experimentation, and participatory governance frameworks, these initiatives contribute practical insights into overcoming barriers such as epistemic asymmetries, limited stakeholder engagement, ethical operationalization, and technical scalability. The table below presents an initial mapping of how each project aligns with the dimensions, offering a comparative overview before the in-depth discussion that follows in the subsequent sections.

Table 2 Mapping how each of the four HEU projects addresses HCAI dimensions by Régis et al. (2024), and adapted to the domain of AI-enhanced democracy

HEU project	HCAI dimension tailored on democracy addressed	How is it addressed
ORBIS	D1. Participatory design and co-creation	Addressed through its structured co-creation methodology, involving pilots and communities from early needs mapping to iterative tool refinement
	D2. Civic participation and democratic engagement	Addressed since ORBIS tools are designed not only with but for civic actors, enhancing their capacity to engage in deliberation and influence decision-making
	D5. Adaptive governance	Operationalized by tailoring AI functionalities to diverse pilot contexts, allowing flexible integration across local, institutional, and thematic settings
	D7. Socio-technical systems and human factors	Achieved by grounding tool development in real-world deliberative environments, ensuring compatibility with the workflows, capacities, and cognitive needs of both moderators and participants
	D8. Transparency and algorithmic explainability	Advanced through the explanation generator, which provides user-friendly justifications for AI outputs, promoting understanding and trust
	D9. Accountability and oversight mechanisms	Embodied in the human-in-the-loop model, where all AI-generated content undergoes moderator validation, embedding oversight into the deployment process and reinforcing institutional accountability
KT4D	D1. Participatory design and co-creation	The KT4D Use Cases brought together the research and development team with representatives of three major stakeholder groups (including two culturally different groups of citizens) at three key points in the project for participatory design and valourisation sessions
	D2. Civic participation and democratic engagement	The KT4D Digital Democracy Lab acts as a structured exploration of the place of AI in civic participation in AI-enabled contexts. It also acts at a meta-level to encourage consideration of how a technology platform intervenes in such applications
	D3. Human rights-based approach	Significant reports were produced in the project on the role of cultural diversity, critical digital literacy, participatory machine learning and design justice in understanding the interface of AI and democracy to inform its innovation activities
	D5. Adaptive governance	The project looks at issues of governance and control both from the top down (informing the work of policymakers) and from the bottom up (bringing new perspectives on ethics to software designers)
	D6. Policy evaluation and impact assessment	The KT4D Social Risk Toolkit provides a targeted, but high-level, guide to some of the key social issues underpinning the relationship between AI, big data, and democracy, such as trust, autonomy, and privacy

(continued)

Table 2 (continued)

HEU project	HCAI dimension tailored on democracy addressed	How is it addressed
	D7. Socio-technical systems and human factors	Places a significant emphasis on issues of HCI (as in the DDL developed, described later in this chapter), critical digital literacy through education for citizens and the role of values and cultural practices in both democratic practice and AI adoption
	D8. Transparency and algorithmic explainability	As a part of its commitment to fostering critical digital literacy, KT4D developed a suite of educational games and experiential explainers targeted at capacity building for citizenship in the digital age
	D10. Algorithmic legitimacy and democratic governance	Through stakeholder engagement and a Delphi survey, KT4D facilitated policy recommendations aimed at addressing gaps in the current regulatory landscape and the needs of democratic practice
AI4GOV	D1. Participatory design and co-creation	Addressed through the implementation of theoretical and methodological steps that were followed to conduct a field study in the form of focus groups on the targeted underrepresented populations to complementary support the HRF of AI4Gov
	D2. Civic participation and democratic engagement	Different groups of stakeholders (i.e., policymakers, public servants, citizens) have access to the tools offered by the AI4Gov project and can edit the ownership of policies, data, as well as influence the policy making procedures
	D3. Human rights-based approach	The project explores the challenges and discriminatory practices faced by vulnerable or underrepresented citizen groups when accessing government services and implements a concrete methodology towards identifying how traditional biases (non-AI) currently impact the rights and values of underrepresented groups
	D5. Adaptive governance	Operationalized by enabling citizens and policymakers to join the platform and leverage the blockchain capabilities to participate in open democracy schemes
	D7. Socio-technical systems and human factors	Achieved by validating the tool development in real-world use cases, ensuring compatibility with the current workflows and needs of public institutions and all relevant stakeholders
	D8. Transparency and algorithmic explainability	Advanced through the creation of a novel scale for pragmatic assessment of the quality of explanations as perceived by users of business processes. A key innovation is the use of causal execution knowledge to generate more interpretable explanations

(continued)

Table 2 (continued)

HEU project	HCAI dimension tailored on democracy addressed	How is it addressed
	D9. Accountability and oversight mechanisms	Accountability is ensured through defined roles, robust governance, and the inclusion of traceability features and mechanisms for addressing failures
ITHACA	D1. Participatory design and co-creation	Addressed by involving a wide range of stakeholders (including citizens with different needs, ages and socio-economic backgrounds, legal practitioners, policy makers, tech specialists, data scientists, etc.) in the analysis and design of AI-powered solutions for civic participation, through co-creation workshops and focus groups
	D2. Civic participation and democratic engagement	ITHACA developed an intelligent discussion platform for civic engagement in local governance in two pilot cities from Slovakia and Romania
	D3. Human rights-based approach	Addressed by defining a value-based data governance framework in the AI context, embedding relevant ethical principles and practices in the design, development and deployment processes for AI-powered civic engagement platforms
	D6. Policy evaluation and impact assessment	Facilitated by providing policy recommendations on how to ensure that philosophical, legal and ethical values are embedded in the development of AI-powered technologies that support civic participation with a view to amplifying positive effects and preventing or minimizing possible adverse effects
	D8. Transparency and algorithmic explainability	Facilitated by the development of a suite of guidelines, techniques and tools that facilitate the design, development and evaluation of the compliance of AI-powered technologies for civic engagement under the principles for trustworthy AI
	D9. Accountability and oversight mechanisms	Ensured through robust governance, traceability, logging and accessible mechanisms that facilitate reporting of potential vulnerabilities, risks or biases in the ITHACA platform

Comparative table showing how four Horizon Europe projects—ORBIS, KT4D, AI4GOV, and ITHACA—address key Human-Centered AI dimensions adapted to AIenhanced democracy. For each project, the table maps specific HCAI dimensions to concrete design choices, tools, methodologies, and governance practices, outlining how participatory design, civic engagement, human rights, adaptive governance, explainability, accountability, and democratic legitimacy are operationalised in different project contexts.

5 Co-creating Innovation in ORBIS

In the HCAI setting, ORBIS stands out as a compelling case study for advancing AI-enhanced democratic deliberation through co-creation. Its contribution extends beyond technological advancements to include the exploration of the three scaling mechanisms explored in Chap. "Why AI and Democracy?". Co-creation stands at the core of ORBIS's approach to HCAI. Beyond being a supportive element, it serves as a strategic method for shaping both the technological and governance dimensions of innovation. From the outset, co-creation is employed to guide two intertwined innovation domains: the design and functionality of AI components on the one hand, and the envisioning of how these tools and deliberative practices can be sustainably scaled across democratic governance settings on the other. To support this dual agenda, ORBIS has developed a multi-phased, multi-stakeholder co-creation methodology specifically aimed at rooting AI innovation in democratic values and deliberative practices. In ORBIS, co-creation is understood as a structured, iterative, and dialogic process. It integrates diverse epistemologies, priorities, and practices from both technology developers and deliberative actors into a shared innovation pathway. In doing so, ORBIS offers a privileged setting to demonstrate how inclusive and participatory methodologies—anchored in co-creation—can significantly enrich AI development for democratic deliberation and foster more informed, legitimate, and responsive governance.

Through the combination of exploratory workshops, iterative design feedback, and real-world validation, the ORBIS methodology aims to ensure that the development of AI-enhanced deliberative tools was grounded in the actual needs, values, and dynamics of participating communities, placing users—namely civil society, grassroots organizations, and civic tech practitioners—at the core of its development strategy. The six ORBIS pilots bring a unique deliberative scope, engagement model, and target audience to the project—ranging from youth involvement and local community planning to European-level policymaking. This heterogeneity enables the project to explore democratic participation across diverse institutional levels and thematic domains.

The ORBIS methodology is structured into three sequential phases, complemented by a fourth, parallel synthesis phase. The first phase focuses on mapping pilots' needs and practices in order to ground innovation in real-world deliberative dynamics. This is followed by two iterative co-design phases, during which pilot-specific challenges and opportunities are translated into concrete design directions for AI tools. Running alongside these phases—mapping, engagement and scenario development—, the synthesis layer consolidates insights across iterations, progressively informing technical specifications and guiding the final design and development of the AI-enhanced tool (Fig. 1).

5.1 Mapping Democratic Attitudes and Participation Needs

The process begins with a comprehensive mapping phase aimed at capturing the diversity of democratic orientations and digital engagement practices across the six

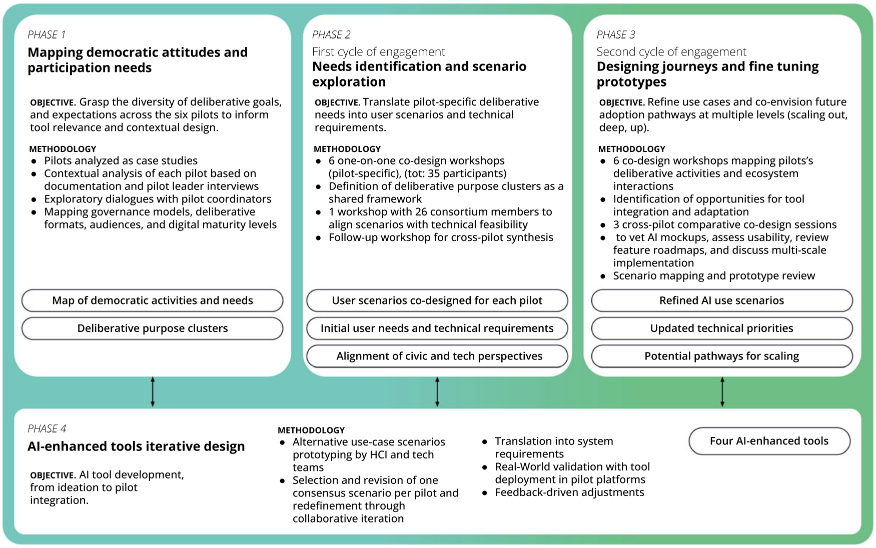

Fig. 1 The ORBIS co-creation methodology: from mapping democratic needs to the iterative development and contextual integration of AI-enhanced tools for deliberation

ORBIS pilots.[1] Each pilot is analyzed as a case study across five key dimensions: the issues being deliberated (Q1), the origins and phases of the process (Q2), the ecosystem of involved institutions and stakeholders (Q3), the formats and technologies used for participation (Q4), and the key democratic outcomes achieved to date (Q5). These findings were then complemented by a contextual analysis conducted through internal document reviews, interviews with pilot leaders, and exploratory dialogues with coordinators. This provides a foundational understanding of each pilot's specific context and operational practices, resulting in a clear mapping of their governance models, deliberative formats, level of institutional and digital maturity.

This mapping reveals a heterogeneous spectrum of democratic attitudes and aspirations: while some pilots aimed at fostering systemic democratic reforms on broader national or supranational levels, others focused on strengthening local participation

[1] Co-creation within ORBIS was implemented through a cross-domain team composed of Grazia Concilio, Anna Moro, and Michelangelo Secchi (Politecnico di Milano, Department of Architecture and Urban Studies), Ilaria Mariani (Politecnico di Milano, Department of Design), and Anna De Liddo and Lucas Anastasiou (The Open University, Knowledge Media Institute).

This core team actively engaged representatives from all pilots—Elina Makri (Democratic Cultural Foundation, Pilots 1 and 2); Timothy Yu-Cheong Yeung, Marta Dell'Aquila, and Berta Mizsei (CEPS, Pilot 3); Caterina Berardi, Lydia Aguirre, and Erika Stael von Holstein (RIE, Pilot 4); Francesco Purpura, Martina Ferruzzi, and Anna Moro (RDM, Pilot 5); and Spyros Papaspyrou and Elena Bekri (Region of Western Greece, Pilot 6)—as well as members of their respective communities.

See Deliverable D2.1 *"Research and innovation landscape and pilots definition"*: https://ec.europa.eu/research/participants/documents/downloadPublic?documentIds=080166e506723d7a&appId=PPGMS.

mechanisms and addressing community-specific challenges. This diversity reflects the pilots' varying scopes, governance cultures, and organizational maturities. It provides a rich foundation for understanding how deliberative goals, practices, and expectations vary across contexts.

These initial findings are analyzed against the Spectrum of Collective Intelligence Activities (Buckingham Shum et al. 2014), a well-established reference model for understanding how groups engage in technology-mediated problem-setting, problem-solving and sensemaking. Through a qualitative analysis, the ORBIS team recognized limitations in the application of the original Collective Intelligence framework when applied to deliberation. As a result, building upon the original five components of the spectrum (collective sensing, sensemaking, ideation, decision-making, action), the project adapts and extends the framework into seven Deliberative Democracy Building Blocks: collective sensing, collective awareness, open discussion, structured discussion, co-creation, decision-making, and power influencing—which are specifically tailored to reflect the democratic goals, engagement formats, and civic ambitions emerging from the pilots. These Building Blocks then serve as an overarching reference model guiding both the design of AI functionalities and the interpretation of their role in scaling democratic innovation, and informed the following two cycles of engagement with their co-design sessions.

5.2 Needs Identification and Scenario Exploration

The second phase of the ORBIS co-creation methodology focuses on identifying concrete deliberative needs and exploring usage scenarios for AI-enhanced tools.[2] This phase, shaped as a first series of workshops, operationalizes the Deliberative Democracy Building Blocks by mapping them onto the actual practices, constraints, and aspirations of each pilot (Fig. 2). Meanwhile platforms—BCause, PolisOrbis, Democratic Reflection—and tools are introduced.

Between July and September 2023, six pilot-specific online co-design workshops were conducted (Fig. 5.2) involving pilot organizers, members of their deliberative ecosystems, and technical partners from the ORBIS consortium (over 35 participants in total). These one-to-one sessions created a structured dialogue between civic actors and technologists, enabling reciprocal learning: pilots articulated challenges

[2] The activities related to Needs Identification and Scenario Exploration were primarily conducted by Anna De Liddo and Lucas Anastasiou (The Open University, Knowledge Media Institute), in close collaboration with Grazia Concilio, Anna Moro, and Michelangelo Secchi (Politecnico di Milano, Department of Architecture and Urban Studies), and Ilaria Mariani (Politecnico di Milano, Department of Design).

See Deliverable D2.2 *"Technical and operational specifications and reference architecture"*: https://ec.europa.eu/research/participants/documents/downloadPublic?documentIds=080166e50532bb37&appId=PPGMS.

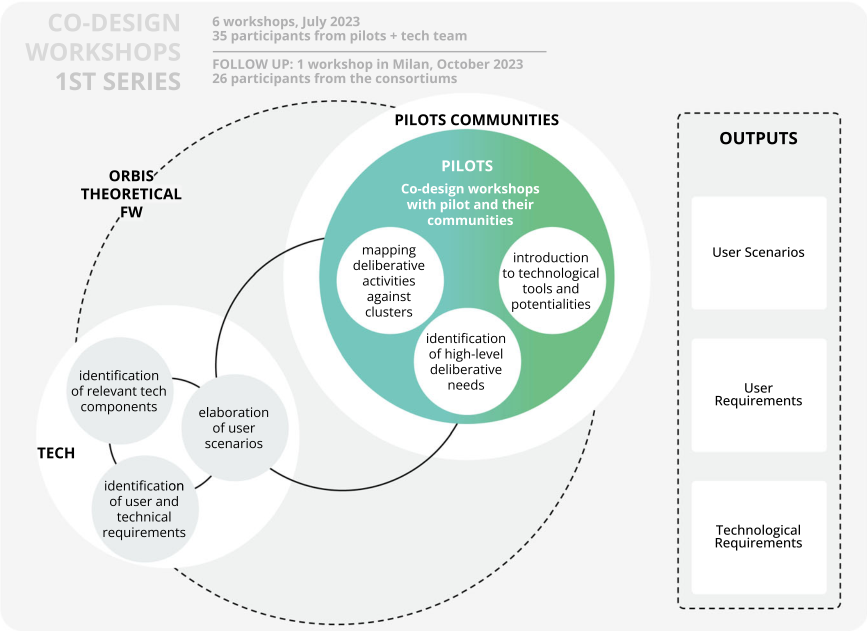

Fig. 2 The structure of the first series of co-design workshops, highlighting iterations and interplays between the tech team, the project six pilots, and their communities, and the tech-related outputs they fed

and expectations emerging from their deliberative practices, while technical partners translated these into assessable technical requirements and preliminary design directions.

Pilots and their communities examined their existing deliberation formats through the lens of the Building Blocks, unpacking each block into associated democratic purposes and user needs. This process supported the identification of critical gaps in current practices—such as limitations in structuring discussion, synthesizing contributions, ensuring inclusiveness, or linking deliberation to policy outcomes—and enabled the joint exploration of how AI functionalities could purposefully intervene in these areas. The outputs of this phase consisted of a set of pilot-specific usage scenarios describing how AI tools could support deliberative activities across different moments of the participation process. These scenarios were consolidated and discussed during an in-person consortium-wide workshop in October 2023 (26 participants), which served as a collective validation and alignment point. Cross-pilot discussions enabled the identification of shared patterns and context-specific divergences, informing the refinement of scenarios into a coherent set of user requirements.

This alignment process informs the technological requirements of the tools and platforms, laying the groundwork for component prototyping and early integration.

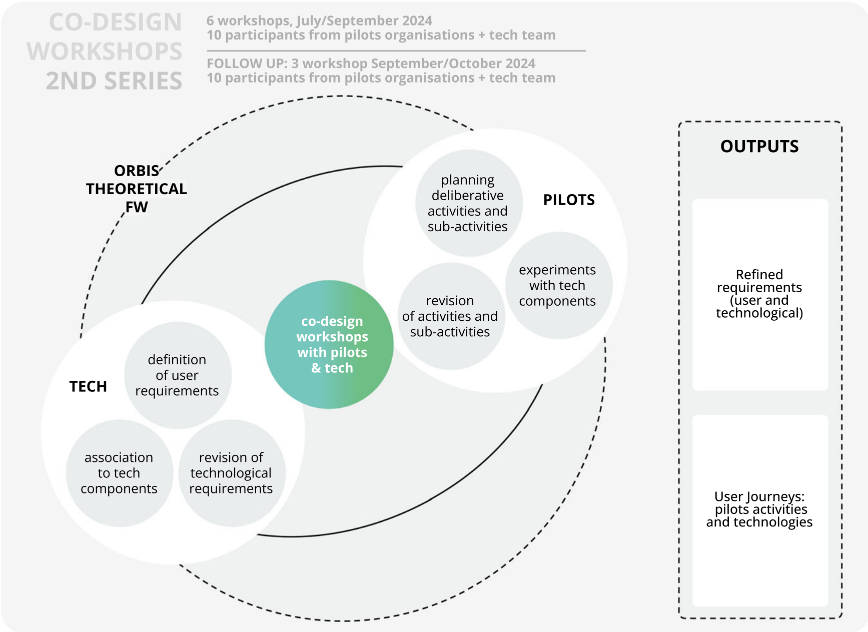

Fig. 3 The structure of the second series of co-design workshops, highlighting iterations and interplays between the tech team, the project six pilots, and their communities, and the tech-related outputs they fed

5.3 Designing Journeys and Fine Tuning Prototypes

A second series of workshops is aimed at fine tuning the results[3] of the first workshop. It engaged pilots in designing their incoming activities, use cases, and desired outcomes, consisting of two stages (Fig. 3). The first involved six workshops—one per pilot—each attended by 2–3 pilot organizers (total = 15). These workshops aim to document the evolution of each pilot's deliberative activities, as well as their interactions with stakeholder ecosystems and ongoing digital engagement practices. This series results in a collaborative planning of the pilots' journeys as progressive steps with specific deliberative activities, methodologies and possible tools.

A follow-up involves all pilot organizers and technical partners in three cross-pilot workshops. These sessions were dedicated to vetting and co-evaluating the proposed technological developments. Participants are introduced to mock-ups, prototype

[3] The final phase of the project involved an expanded group of contributors from the technical side. For the Argument Mining and Explanation Generator, contributions were provided by Cristian Cardellino and Sofiane Elguendouze (Université Côte d'Azur). For the Feedback Aggregator and Policy Recommendation components, the team included Nikos Karacapilidis, George Domalis, and Ioannis Efstathiou (Novelcore). Platform development involved Anna De Liddo, Lucas Anastasiou, and Alberto Ardito (The Open University's Knowledge Media Institute) for BCause and Democratic Reflection, and Davide Formica and Luca Monti (Copernicani) for PolisOrbis.

demonstrations, and feature roadmaps for the AI tools. These sessions produce feedback on interface design, usability, ethical implications, and expected outcomes of the tools, enabling a dialogue on how to align implementation with normative expectations and everyday practices. These workshops also examine the tools' integration within different deliberative formats (e.g. structured platforms like BCause, open-discussion settings, or hybrid models) and consider how outputs could be routed to policymaking or scaled across institutions and contexts.

5.4 *Shaping AI-Enhanced Tools*

In ORBIS, the development of AI-enhanced tools is conceived from the outset as a deeply intertwined, iterative co-design process, where the epistemic, functional, and ethical dimensions of AI were shaped collectively by designers, developers, and civic actors. The starting point for this approach was the recognition that deliberative processes differ greatly across scales, communities, and institutional settings. Hence, ORBIS did not treat the AI toolset as a one-size-fits-all solution but as a modular architecture to be refined through structured collaboration with six pilot organizations. In this process, the pilots acted as co-design partners, offering situated knowledge about deliberative practice, engagement bottlenecks, and social expectations around digital tools. This iterative engagement is structured as a five-step development cycle (Fig. 4). (1) The initial one-on-one workshops identified pilot-specific challenges and goals, followed by the elaboration of needs into relevant insights and elaborate user requirements—informed by phase 2 of the methodology; (2) preliminary design scenarios were drafted by HCI and technical teams to translate user requirements into functional objectives; (3) scenario refinement workshops co-designed a consensus scenario for each pilot—informed by phase 3 of the methodology; (4) these scenarios informed technical specifications; and (5) final testing and validation occurred within real-world pilot deliberations.

This process resulted in an AI toolkit consisting of four components to be implemented in the three platforms adopted by ORBIS—BCause, PolisOrbis, Democratic Reflection.[4] The Argument Mining (AM) module identifies and structures argumentative content from textual input; the Feedback Aggregator (FA) synthesizes and

[4] The ORBIS toolkit components were designed and implemented by multiple partners. The Argument Mining and Explanation Generator were developed by Elena Cabrio, Cristian Cardellino, and Sofiane Elguendouze (Université Côte d'Azur), together with Serena Villata (National Institute for Research in Digital Science and Technology). The Feedback Aggregator and Policy Recommendation components were developed by Novelcore (Dimitris Tsakalidis, George Domalis, Ioannis Efstathiou, and Nikos Karacapilidis).

The BCause and Democratic Reflection platforms were developed by The Open University's Knowledge Media Institute (Anna De Liddo, Lucas Anastasiou, and Alberto Ardito), while the PolisOrbis platform was developed by Copernicani (Davide Formica, Enrico Fagnoni, and Luca Monti).

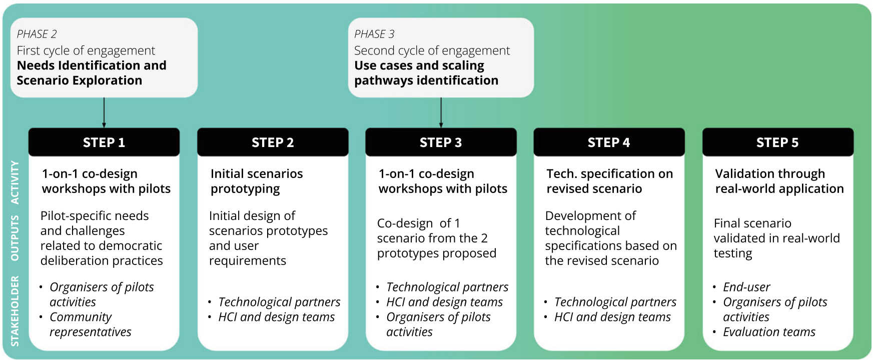

Fig. 4 The five-step development cycle, detailing activities, the tangible outputs generated, and the stakeholders involved

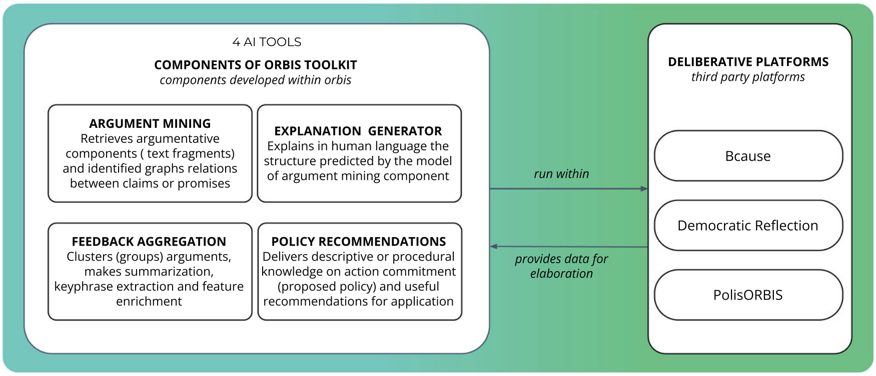

Fig. 5 AI components in the ORBIS toolkit and interaction with deliberative platforms

organizes contributions, revealing key themes and patterns; the Explanation Generator (EG) translates complex algorithmic operations into human-readable justifications, enhancing transparency; and the Policy Recommendation (PR) tool converts deliberative outputs into targeted, actionable policy suggestions. Collectively, they make up the ORBIS AI toolkit, which has been integrated into three deliberative platforms implemented within the project (Fig. 5).

5.5 *A Focus on the Four Components*

The aforementioned four AI components were shaped through continuous feedback from civic actors. Each of them benefited from targeted co-design interventions that reshaped their architecture, usability, and intended role in democratic engagement.

Drawing on advanced techniques from data analytics, machine learning, and natural language processing (NLP), these components are designed to process large volumes of input—from structured argumentation and position statements to unstructured, real-time discussion streams. A short summary of the components and their respective contributions to the system architecture is provided below.

Argument Mining (AM). AM is developed to extract and classify argumentative statements in deliberative texts (Cabrio and Villata 2018; Goffredo et al. 2023). Pilot feedback revealed two divergent data realities: (1) structured discussions when inputting comments in e-participation platforms—such as BCause, where users enter statements tagged as positions or arguments; and (2) free-form textual data from less structured platforms. This prompted the creation of a dual-path AM system: one tailored to structured, statement-level inputs, and another optimized for parsing continuous discourse. In workshops, pilots emphasized the need not only to identify argumentative components (positions, supports, attacks) but also to expose logical relations within and across discussions. In response, the AM system was extended to include inter-discussion linkage, allowing the identification of thematic and argumentative connections between separate threads—enabling users to see how their ideas align or conflict with contributions from parallel discussions. Iterative feedback loops with pilot teams were critical in refining relation classification models, balancing machine learning accuracy with interpretability.

Feedback Aggregator (FA). FA is designed to summarize, cluster, and enrich public input, transforming large volumes of user-generated content into digestible and actionable insights. The co-design journey begins with pilot organizers expressing a need for more than basic summarization: capturing the semantic depth, diversity of viewpoints, and topic-sensitive granularity. Based on these inputs, FA is developed to perform multiple tasks. Workshops revealed that clustering needed to reflect both thematic and conceptual proximities, not just linguistic similarity. As a result, metadata from trusted knowledge repositories (e.g., Wiki) was integrated into the clustering logic to provide contextual enrichment. Abstractive summarization models are trained to generate cluster-level titles and summaries, providing high-level orientation while preserving argumentative diversity. A keyword extraction feature is added to identify salient terms, enabling users to trace emergent discourse themes without reading entire threads. Pilots also expressed concern about automated outputs being potentially misleading or oversimplifying complex debates. This led to the introduction of a human-in-the-loop moderation layer, allowing facilitators to review and validate AI-generated summaries before publication.

Policy Recommendation (PR). PR addresses a common bottleneck in deliberative systems, recurrently voiced by ORBIS pilots: the gap between discourse and policy action. The co-design process revealed strong demand for features that translate the outputs of discussions into policy-relevant knowledge. Informed by this, PR is designed to extract structured recommendations from user discourse, supported by argument strength and sentiment trends. Co-design workshops emphasized that recommendations should reflect both volume and diversity of support (related posts), leading to the inclusion of metrics that show how many posts substantiate each policy suggestion. This provided traceability and gave moderators a traceable chain of support behind each recommendation and breadth of civic consensus. Another key output of the co-design process was the integration of policy matching capabilities, using large language models (LLMs) to compare recommendations against existing regulations—highlighting gaps, alignments, or novel contributions.

Explanation Generator (EG). EG covers a central challenge in HCAI: making algorithmic processes transparent and accountable. It addresses Explainable AI (XAI) (Coussement et al. 2024; Karacapilidis et al. 2023; Siachos and Karacapilidis 2024) by translating complex algorithmic processes into clear, human-readable insights that enhance transparency, trust, and user comprehension. From the beginning, pilots expressed the need for AI systems not only to function well but to justify their outputs in ways accessible to non-experts, allowing participants to understand algorithmic logic without needing technical literacy. Co-designed with pilot organizers and tested in live deliberative settings, EG was developed in response to a critical need for participants and facilitators to understand how AI-generated outputs—such as argument classifications or discussion clusters—are produced. One of its foundational features is attention mapping, which highlights the specific parts of user input that most influenced the model's decisions, enabling intuitive traceability of outputs. Building on pilot insights, EG is extended to express its confidence levels and argumentation logic, such as why a statement was categorized as a premise or claim, and how it relates to others via support or attack links, shaped as user-friendly rationales for argument classifications and relationship detection. A second wave of co-design focuses on linking EG to FA outputs, offering meta-level insights into discourse structure and emerging themes. Pilots requested the ability to see how summaries are derived, how clusters form, and what keywords shape thematic emphasis. This results in narrative-style explanations that map data provenance and processing steps. Discussions also explore the use of explanation scores to guide participants toward weaker or stronger arguments—turning EG into a deliberation aid, not just a transparency feature. By enabling users to interrogate, contest, or reflect on algorithmic reasoning, EG transforms explainability from a static reporting function into an interactive, participatory feature, supporting informed engagement and reinforcing democratic accountability in AI-supported deliberation.

6 Democracy in the Loop. The Case of KT4D

Within the comparative framework of this chapter, KT4D addresses AI-enhanced democracy not by embedding AI directly into deliberative platforms, but by strengthening the deliberative capacities of citizens and institutions through critical literacy, participatory design, and collective sensemaking.

The literature on HCAI places a strong emphasis on the importance of human agency in any democratic process, regardless of the extent to which said process might be supported, facilitated or enhanced by the application of AI in its delivery.

In spite of the ethical centre this implies, however, the wider environment of AI ethics has not necessarily aligned as strongly with these goals as human-centric democratic innovation would necessarily require. Bolte and van Wynsberghe (2025) usefully characterise the development of AI ethics over time as having followed a series of historical waves, the first of which tended to focus broadly on technological hypotheticals, the second of which became almost too narrowly focussed on specific components of the wider challenge of ethical AI, such as opacity or data bias. These precursors have enabled a more holistic approach to emerge under the rubric of 'Sustainable AI,' which is able to bring together the long term considerations of the first wave with the more tactical approaches of the second.

Sustainability, as per Bolte and van Wynsberghe, is 'a property of large-scale systems, not a property of the technology or object under consideration' (2025, p. 1737). In this way, this emerging paradigm embraces many of the principles of HCAI described above by resisting the tendency to view technology platforms and products as somehow independent of the ecosystems in which they are deployed, or of the people whose voices they are intended to channel, amplify or moderate. Seen from this perspective, the challenge of developing AI systems that respect the tenets of HCAI and which can foster rich democratic exchanges emerges as one most often framed in terms of technology development, but truly situated in the associated field of Human–Computer Interaction (HCI). HCI has historically been viewed as limited in its suitedness to the needs of human collectives, rather than individuals, however. The manner in which it has been construed at the intersection of fields such as psychology, sociology and design has left the mainstream approaches that inform the field less than fully suited to ensure the cultural layers that underpin our democratic and civic participation—our stories, values, discourses, etc.—are taken fully into consideration when considering how we might or should interact with technology. Though HCI of course continues to evolve, its roots in the paradigm of usability maintain a bias toward professional contexts, weaknesses vis-à-vis the social dimension of collaboration and contextual interactions (Bardzell and Bardzell 2015); overemphasis on a simplifying vision of binary relationship between humans and devices, rather than software systems (Mortier et al. 2015) and a tendency to see individuals only as consumers of knowledge technologies (Sunstein 2007) when they are as much formed by these technologies as they are in control of them.

The practical implementation of measures to engage citizens in the design and deployment of the AI systems that will influence the contexts and conditions of their democratic participation requires a reconsideration of the manner in which we construct HCI interventions in this space. For this reason, rather than build a system designed to ultimately become a product-like part of the democratic AI landscape, the KT4D project decided to focus on contributing more indirectly to meeting the goal of ensuring HCAI in deliberative contexts. Through its demonstrator system and extensive handbook-style documentation (which will be described in more detail below) the KT4D Digital Democracy Lab (DDL) aims to promote meaningful participatory design processes by enhancing the critical digital literacy of the potential users or commissioners of, or contributors to, future AI platforms. Existing as it does in the dual embodiments of a set of integrated technical components (the Demonstrator) and a set of prompts for human communication about and consideration of these components (the Handbook and Facilitation Guides) the DDL seeks to empower citizens and their representatives to use their voice in the development of systems that could ultimately be used to frame the future of their societies and cultures.

In the course of developing this hybrid object as a contribution to the HCI of HCAI, a number of conceptual innovations were able to be solidified into the framework that would ultimately shape the DDL. The first of these was the importance of meaningful friction. Friction is generally viewed negatively within technical development teams and contexts. It is perceived as contributing to inefficiency, creating extra work (Bradley et al. 2022) and resulting in both 'technical debt' (Ozkaya et al. 2019) and the 'cognitive friction' of a system which visually seems to promise one thing, but delivers another (IxDF - Interaction Design Foundation 2016). More recently, however, the potentially positive role of friction in design has been emerging as an area of interest for fields in which design processes need to ensure meaningful user engagement. Therefore, from a deliberative perspective, meaningful friction functions as a safeguard against premature consensus, uncritical delegation to algorithmic authority, and the erosion of citizens' responsibility for judgment. In their literature review, Benedetti and Mauri (2023) look at the ways in which frictions in design might achieve 'the goal of promoting reflection' (2023, p. 138). While this work finds a hugely varied terminological landscape (encompassing descriptors from 'microboundaries' to 'slow technology' to 'value-sensitive design'), it ultimately concludes that interest in mechanisms that slow down technology interactions are both growing and necessary. To embrace meaningful friction is to recognise that no system, no matter how intuitive and well-designed, can make its full apparatus of inputs and processing limitations, biases and optimizations, transparent at the single glance we give to a user interface. While the maximal reduction of friction may be viewed as a necessity for user retention in circumstances when both their goals and the impact of any decision that can be made are straightforwardly transparent, there are many cases where this is not possible or where a 'good enough' decision-making strategy impinges on core values, such as citizen voice in democratic debate. In these cases, the introduction of friction can serve a positive function to ensure that users do not make incorrect assumptions about system neutrality or authority, or delegate their own civic responsibilities to agents without a stake in the operation of their societies or communities.

The second key enabling concept utilised by KT4D was that of the AI system metabolism, and how it might be meaningfully revealed to users. The value among technology developers of reducing cognitive friction is abetted by the inherent complexity and opacity of AI systems so as to make it very difficult to ensure that users are able to access even the most rudimentary information about how that system functions, and on what criteria its outputs are predicated. Even the most basic questions that might be asked about the function of such a system—what inputs does it base its conclusions on? How has its predictive model been informed? How reliable are its conclusions?—tend to be obscured behind a mixture of corporate policy, system complexity and a software developer's biases and assumptions about how much a user needs and wants to know. The epistemic disparities and deficits in transparency represented by this situation create one of the primary research gaps described in the introduction to this section. This can be seen nowhere better than in this ethically challenging transposition of design principles, formulated according to a paradigm of a consumer purchase, to the context of democracy. In fact, recent work has shown that the less a user or citizen knows about AI, the more receptive they are to its use (Tully et al. 2025). As a result, individuals may approach the challenge of understanding the kinds of abstractions inherent in AI systems through a comparative lens drawing on previous life experience—a fully understandable strategy, but one that lead to miscorrelations of causes and effects that can have a deleterious effect on the veracity of sensemaking processes (Greenwald et al. 2021). These findings may be good news for marketers based in tech companies, but they are bad news for AI and democracy, as the latter requires citizens to be active, engaged and informed about the impact of their choices. By creating a platform that allows the use of an AI tool to be experienced alongside a systematic exposure to the data, processing and presentation decisions being made in the background, encouraging each of these layers to be queried and potentially tweaked, the KT4D DDL shifted the paradigm of use from one of maximising convenience (low friction) to enhancing active engagement (high critical literacy).

The overarching guiding principle that emerged in the development of the KT4D DDL became known as Democracy-in-the-Loop. This term is obviously drawn from the more commonly recognised technology design paradigm known as 'human-in-the-loop' (HITL). HITL can imply a number of different things, however, and be used in a number of different contexts. In some cases, it refers to a development approach that requires a combined intelligence and distributed responsibility between a human actor and an expert system; in other cases, the term can refer more to an ethical imperative to ensure human oversight over certain kinds of critical decision-making. Democracy-in-the-loop embraces both of these senses, in recognition of the nature of democracy as complex, critical and predicated upon human decision-making and action. The term also, however, cuts to the core of the challenge of HCI in democratic contexts, as described above, placing an emphasis on the manner in which any interaction with AI that acts as a contributor to democracy must be understood as a collective, rather than isolated individual, act. From a deliberative perspective, Democracy-in-the-Loop reframes human-in-the-loop paradigms by foregrounding collective judgment, reciprocal justification, and shared responsibility over individual oversight, positioning AI as a participant in processes of public reasoning and collective sensemaking, rather than a substitute for democratic deliberation. The speaker cannot be considered without the audience; the direct effect cannot be disentangled from its reverberations; the single intervention must always be understood in the context of its many possible intended and unintended consequences. This conceptual evolution is described in more detail in (Ohren et al. 2025), but for the purposes of understanding the KT4D DDL as a contribution to the HCI of HCAI, this framing of the interaction between AI and democracy as a challenge of collective sensemaking (Edmond forthcoming) designates an important set of conditions for the interactions facilitated by the project.

The convergence of this conceptual framework with the mandate to realise an applied experiment in AI innovation within the KT4D project shaped both the design and ultimately the participants' experience of the DDL. This intersection was not necessarily a simple or indeed comfortable venue for operations, however. It is easy to develop ethical frameworks and software systems in isolation from each other, but attempts at bringing them together tend to reveal all too clearly the gaps and weaknesses in each. As such the resulting form of the DDL is that of an experimental model of a mini-public which integrates AI in novel, exploratory ways. As a prototype, the DDL is neither a perfect representation of a mini public nor a seamless deployment of AI. However, it is an ideal opportunity to investigate the added-value and risks such technology can afford, within the settings of democratic exchange. In essence, the DDL was designed not as a platform to be implemented, but as a one-day workshop aimed at exploring and demonstrating the potential and limitations of AI and big data for enhancing democratic exchange processes. The Lab incorporates these technologies within established best practices for civic participation, creating a unique environment for understanding and evaluating AI's role in democratic processes.

The Lab employs a systematic approach, beginning with addressing a wickedly complex and multifaceted challenge. For the first iteration of the DDL, the policy focus was on the EU's role in labour market transitions during the AI era, but any sufficiently well-bounded problem of a similar scope could have been chosen. The software platform (available on GitHub) through which the participants engaged with this issue included a number distinctive modules, with break points built in between them. The issues that would be explored through this framework included the nature of training data, and how it is chosen and curated (encouraging consideration of data cooperatives); the process of tagging training and testing data (consideration of participatory ML design); and the viewing, application and proposing of profiling cases (to understand the trade-offs between profiling and ethics). Finally, through collaborative exercises, participants generated prompts for a specialised GPT model using a custom knowledge source of 30 documents—comprising 10 from newspapers, 10 from policy papers, and 10 from academic journals—leveraging the Retrieval-Augmented Generation (RAG) technique. Participants then critically reviewed the AI-generated policy positions and engaged in iterative improvement cycles to address any misalignment between their own expectations and values and the assumptions encoded in the specialised GPT regarding those expectations and values.

It is perhaps important to note that, in spite of our own initial expectations, in the persona description process, we did not ultimately use participants' own pre-existing online footprint, and/or individually provided data. Although this would in some ways have been interesting for the participants to see how their own actual profiles resonated with those of the perspectives used to train the system, the ethical risks of collecting, using and deleting their data seemed too high, and the risk of this having a chilling effect on user participation seemed too high to take. Our original participatory design sessions did not indicate that our overarching goal of increasing the critical digital literacy of the participants required that their own personal data be used. Instead, we did ask the participants to fill out a profiling quiz with questions concerning their stance on social, political and economic questions (which, it goes without saying, they could answer as truthfully or creatively as they wished). This was then the basis of the profiles used by the RAG. This data was then deleted at the end of the session, meaning the model did not train further on the data.

While maintaining this common structure, each guide was carefully tailored to address the specific needs and objectives of its target audience. The Brussels Use Case emphasised policy implications and governance frameworks, aligning with the expertise and interests of policymakers and AI policy experts. The implementations of the Use Cases in Madrid and Krakow focused primarily on citizen engagement and democratic participation, creating accessible entry points for the general public to engage with these complex topics. The Dublin use case took a more technical approach, facilitating deep-dive explorations and practical implementation considerations for software engineers and technology professionals, especially for those who are involved in the creation of technologies of AI and Big Data.

In all of this work, we foregrounded participatory and human-centric AI approaches that leveraged human science and qualitative big data to understand the deeper needs, aspirations and drivers that underlie human behaviours in our target societies. Accordingly, we focussed on implementing the basic benefits of human-centric AI, such as informed-decision-making, reliability and scalability, and more fair and trustworthy software and product-building. In this way, we were able to develop a comprehensive and legally sound approach to the design, development, and deployment of information systems that learn from and collaborate with humans in a deep, meaningful way, not just exploring and informing, but implementing in software the project's expertise and approach to digital ethics.

In the comparative perspective developed across this chapter, KT4D complements the other projects by addressing a foundational but often overlooked dimension of AI-enhanced deliberation: the conditions under which citizens can meaningfully understand, question, and shape AI-mediated democratic processes. Rather than optimizing participation at scale or institutional deployment, KT4D foregrounds deliberation as a practice of collective sensemaking, positioning critical literacy, reflective friction, and participatory design as democratic safeguards in their own right.

7 Engaging the Marginalised. The Case of ITHACA

Positioned within the comparative HCAI lens adopted in this chapter, ITHACA addresses a core democratic challenge that consistently emerges across AI-enhanced participation: how to ensure that deliberative innovation does not reproduce—or deepen—existing forms of exclusion. In this sense, the project operationalizes human-centeredness primarily through an inclusion-first strategy, combining participatory governance methods (Citizens' Juries) with AI-enabled accessibility features (e.g., language simplification, translation, speech interfaces) to broaden who can enter, remain in, and benefit from deliberative processes.

Digital Inclusion has been an important research topic during the last two decades, as it has evolved along with the rapid adoption of Information and Communication

Technology (ICT) in all aspects of everyday life, thus in the modern society becoming mandatory to ensure that all citizens have equal opportunities to participate when it comes to the emerging digital engagement platforms and eGovernance tools. In the European Union, as shown by the *Digital Decade reports for 2025*,[5] significant progress has been achieved by many of the countries in regard to digitalization through connectivity, adoption of digital technologies by businesses, provision of digital public services, etc. However, many countries are still beyond, in particular when it comes to the human capital (e.g. development of digital skills), thus digital inequalities persist on a large scale.

Online civic engagement platforms could help reduce the digital inclusion gap, by empowering citizens who experience marginalization for various reasons (e.g. age, gender, ethnicity, education) or become vulnerable due to life course (e.g. unemployment, immigration, health issues, etc.) and enabling participation regardless of location or time, particularly benefiting the elderly, citizens with mobility constraints, or those in rural areas (Deng and Fei 2023; Pietilä et al. 2021). However, the very individuals that are already underrepresented in politics and are often unheard in their opinions by political decision-makers (Skorge 2023), include those lacking digital skills or having diverse abilities (e.g. functional or cognitive disabilities), thus such platforms further enhancing the risk of exclusion. Thus, technology should be designed and leveraged in an inclusive manner, to assist marginalized and vulnerable groups instead of keeping them away from technological innovations. For instance, AI-based systems, such as LLMs, as well as tools for Text-to-Speech and Speech-To-Text have the potential to enhance inclusivity by facilitating digital accessibility of interfaces, including language simplification and translation (Garcia Valencia et al. 2024; Yenduri et al. 2023). Understanding what benefits citizens with different needs and diverse backgrounds is the main goal of ITHACA project, in order to ensure that their requirements for active civic engagement can be met. In the process of generating and validating user requirements, a multi-step process has been adopted, including: (i) a focused state-of-the art analysis and desk research to establish the context of use and special user groups data (e.g. census data from the two pilot cities, Brasov in Romania and Martin in Slovakia), (ii) focus groups with various stakeholders (e.g. experts working on a daily basis with marginalized and vulnerable groups, ethics and legal experts, AI experts, municipality workers, etc.); (iii) participatory activities (e.g. workshops) with the citizens from the pilot cities; (iv) prototyping and mock-ups testing; and (v) validation and evaluation of the developed prototypes. In this section, the focus is on the participatory activities with the end-users, the so-called 'Citizens' Juries' of the two ITHACA pilot sites. The selection of the pilot sites is justified by their particular contexts.

Romania was ranked on the last place among the 27 European Union Member States in regard to the *Digital Economy and Society Index (DESI) in 2022*,[6] as it was beyond all countries in regard to most of the indicators and it appears to be constantly beyond, as demonstrated by the 2025 Digital Decade report. While the

[5] https://digital-strategy.ec.europa.eu/en/library/state-digital-decade-2025-report.

[6] https://digital-strategy.ec.europa.eu/en/library/digital-economy-and-society-index-desi-2022

connectivity infrastructure is well developed and it is above the average, Romania is ranked on the last place for key indicators such as digital public services for citizens and for e-government users and pre-filled forms. Furthermore, when it comes to human capital, only 1/3 of people aged between 16 and 74 have at least basic digital skills, and only 10% of the individuals have above-basic digital skills. In regard to the local context, Brașov municipality, located in the central part of the country, has a total population of over 400,000 inhabitants and consists of 19 communities (7 cities and towns and 12 rural localities). The area is characterized by a high degree of urbanization (>80% of the population lives in urban areas) and multiculturalism, with 77.82% being Romanians and many other ethnicities being present (e.g. Germans, Hungarians, Romani people). Older citizens (65+) represent 20.35% and immigrants represent 2.21% of the total population.

Slovakia was ranked on the 23rd place among the 27 European Union Member States in 2022 in regard to DESI, and little advance has been noticed since then, remaining well under the EU average in particular in regard to digital skills (e.g. 55% of Slovaks have basic digital skills, 21% have advanced digital skills). Furthermore, when it comes to digital public services, Slovakia is below the EU average for all monitored indicators. There is also a gap for digital public services for businesses, with the future plans being focused on minimizing the required administrative steps for citizens and business, and improving the user-friendliness of digital public services. The City of Martin, located in northern Slovakia, has a population of over 50,000 inhabitants, of which the vast majority (92.4%) is Slovak, but other nationalities and ethnic minorities are also represented (i.e. Czech, German, Romani people, Icelanders, Norwegians, etc.) [10]. Elderly (60+) represent over 27% of the total population and younger/youth (18–30 years old) is almost 11%. Refugees and migrants (e.g. Ukrainians, Syrians, Afghans—0.64% in total), homeless people (0.38%) and people of color (1.75%) are also present. Currently, the local and regional administration is not using any online platform to engage citizens in civic activities and discussions.

7.1 Selection Criteria for the Citizens' Juries in the ITHACA Pilot Sites

To ensure a representative and inclusive selection of participants for the AI Citizens' Juries in the two ITHACA pilot sites—Martin (Slovakia) and Brașov (Romania)—a multi-faceted approach was adopted to identify vulnerable groups and the variables constituting vulnerability. This process combined extensive reviews of relevant academic literature and governmental reports, consultation with project partners working in the field of inclusive digitalization, and the analysis of discussions and statistical data provided by partners based in the two pilot cities (Forbes-Mewett and Nguyen-Trung 2019; Pérez-Escolar and Canet 2023; Raisio et al. 2014; Rosenberg 2024; Kocsis et al. 2024; Liu et al. 2025). The outcomes of these activities were consolidated into two analytical outputs: first, the identification of a set of vulnerable groups:

- Elderly/pensioners (i.e., 60+ years of age)
- Younger/youth (i.e., between 18 and 30 years of age)
- Refugees and Migrants (e.g. Ukrainians, Syrians, Afghans, etc.)
- Roma people (including Sinti and Travelers)
- People with physical disabilities (e.g., mobility-impaired or blind people, etc.)
- People with mental disabilities and/or problems (e.g., depression, addiction, etc.)
- People living in rural areas
- Homeless people
- People of colour
- Women
- Pregnant women and women with young children
- Families with many children (more than 3 children)
- Single parents, and
- LGBTQIA + community.

And second, the definition of vulnerability-related variables used to inform participant recruitment:

- Low income
- Poor living conditions
- Precarious employment/(repeated) unemployed
- Lack of insurance
- Low educational background
- Limited educational opportunities
- Low digital literacy and/or particular need for support w.r.t. to digital platform
- Limited access to infrastructure and mobility
- Limited access to cultural programs, resources and information
- Social isolation/loneliness
- Structural/systematic discrimination concerning (democratic and social) participation
- Facing physical and/or verbal violence.

Initial binary assignments of vulnerable groups and vulnerable variables were conducted, and these were further refined by the project partners of the pilot cities, as depicted in Tables 3 and 4. The assignments highlighted in green have been added, and those in red have been removed, as compared to the initial ones.

For the recruitment of Citizens' Juries participants, a comprehensive participation form was created, that beyond the contact data and motivation, captured also the belonging to the identified vulnerable groups and the vulnerable variables applicable to those individuals that were motivated to participate. When asking about those vulnerabilities, we pursued the principles to (i) ask as less questions as possible and as many as necessary, (ii) formulate them as easy as possible, preferring a multiple choice format, (iii) providing background information and GDPR information beforehand, and (iv) covering both, objective socio-demographic variables

Table 3 Refinements of the binary assignments between vulnerable groups/communities and vulnerability criteria by Brasov Municipality (Romania)

		Vulnerable (and marginalized) social groups and communities													
		Elderly / pensioners (i.e., 60+)	Younger / youth (i.e., 18-30)	Refugees & Migrants	Romani people	People with physical disabilities	People with mental disabilities and/or problems (e.g., depression, addiction, etc.)	People in rural areas	Homeless people	People of colour	Women	Pregnant women & women with young children	Families with many children (i.e. >3)	Single parents	LGBTQIA+
Criteria that constitute vulnerability	low income	x		x	x	x			x	*			x	x	
	poor living conditions	x	x	x	x		x		x				x		
	precarious employment (repeated) unemployed	x	x	x	x	x	x	x	x			*			
	lack of insurance			x	x				x						
	low educational background				x								x		
	limited educational opportunities	*		x	x		x		x			x	x		
	low digital literacy and/or particular need for (IT) support	x			x		x		x						
	limited access to infrastructure / mobility				x			x	x						
	limited access to cultural resources			x	x	x	x	x	x				x		

(continued)

Table 3 (continued)

	social isolation / loneliness	x	x				x								
	structural discrimination concerning participation			x	*		x		x	*					*
	facing physical and/or verbal violence			*	*				x	*	x				x

Table 4 Refinements of the binary assignments between vulnerable groups/communities and vulnerability criteria by the city of Martin (Slovakia)

		Vulnerable (and marginalized) social groups and communities													
		Elderly / pensioners (i.e., 60+)	Younger / youth (i.e., 18-30)	Refugees & Migrants	Romani people	People with physical disabilities	People with mental disabilities and/or problems (e.g., depression, addiction, etc.)	People in rural areas	Homeless people	People of colour	Women	Pregnant women & women with young children	Families with many children (i.e. >3)	Single parents	LGBTQIA+
riteria that constitute vulnerability	low income	x		x	x	x			x	x		x	x	x	
	poor living conditions	x		x	x		x		x						
	precarious employment (repeated) unemployed	x	x	x	x	x	x		x			x			
	lack of insurance			x	x				x						
	low educational background				x										
	limited educational opportunities	x		x	x		x		x			x	x		
	low digital literacy and/or particular need for (IT) support	x			x		x		x						
	limited access to infrastructure / mobility				x	x		x	x						
	limited access to cultural resources			x	x	*	x	*	x				x		

(continued)

Table 4 (continued)

	social isolation / loneliness	x					x		x						
	structural discrimination concerning participation			x	x		x		x	x		x		x	*
	facing physical and/or verbal violence			x	x		x		x	x	x				x

and subjective evaluations on social and economic aspects. The last principle had the specific reason for significant differences between the self-perception and the perception of others commonly observed in research with vulnerable groups (Aldridge 2014). While most of the questions explicitly addressed the vulnerability criteria or group, such as 'I have physical limitations and/or illnesses' with the answer options: 'yes'/'no'/'I don't want to answer', addressing physical limitations, some vulnerable criteria could have been inferred implicitly. One example for that is the variable 'Square metres per person in the household' which was not addressed explicitly (since it might be not that easy to do the calculation) but it was inferred based on the responses to the items 'Number of people in your household' and 'Size of your household'. Despite the information on the absolute confidentiality of the personal data provided and the information on the rights of individuals with regard to the protection of personal data, almost all questions in the participation form had the option of not answering in order not to exclude those who feared negative consequences from the disclosure of personal data. In addition, no specific group affiliations were requested for particularly sensitive personal data that could be associated with fear of social exclusion or discrimination, such as belonging to ethnic or sexual minorities.

For all of the variables and criteria that were part of the participation form, except for gender, weights have been assigned to each of them in order to build the overall ranking score. Additional weights were assigned to each of their criteria answer options, between 0 and 1, such that option weight with value 1 corresponds to the maximum criteria weight. The weights were aligned with the refinements of the binary assignments in each pilot city, and are thus strongly overlapping in both cities. The more vulnerable a variable or its characteristic, the higher the assigned weight is. For example, the current occupation 'unemployed' has a higher weight than 'student', as the situation as an unemployed individual in both pilot cities is in general more vulnerable than the one of a student. The highest prioritisation of the selection process was gender equality, that was determined from the beginning and thus, no weights were assigned to this variable. The assigned weights were used to calculate an individual 'vulnerability score' of each citizen that registered for the AI Citizen's Juries, based on their data, voluntarily shared in the participation form. The vulnerability score facilitated the creation of a ranking list for each pilot city, and the final selection of Citizens' Juries participants.

7.2 Participatory Activities with the Citizens' Juries

Two rounds of on-site participatory workshops were implemented in the cities of Martin (Slovakia) and Brasov (Romania) in order to engage with the AI Citizens' Juries. The activities were carried out in the local languages, and were facilitated by expert moderators, aware of the local context and the specific needs of the selected participants.

In the first round, the implementation methodology was based on a mixture of sessions, including: (i) informative sessions, regarding the project objectives and implementation approach, and the topics of civic engagement in modern society, and AI use for civic engagement and relevant examples; (ii) three sessions dedicated to group discussions and activities (e.g. card sorting, co-creation), and (iii) individual activities, dedicated to individual surveys.

In the second round of workshops, the sessions included: (i) informative sessions, where presentations, explanations and examples were given for a better understanding of the topics of human and algorithmic bias, discrimination, ethics and fundamental rights; (ii) group discussions related to concerns and potential harms when participating in civic activities (both offline and online), along with the perceived importance, urgency and difficulty to identify bias and discrimination of AI-powered tools; and (iii) an individual activity dedicated to filling in a questionnaire targeting to capture user requirements for the ITHACA platform.

A total number of 21, and respectively 20, citizens attended the first and second workshops in Brasov, and a total number of 23, and respectively 19, citizens attended the first and second workshops in Martin. In Brasov, out of the 20 participants in the second round, 15 were female and 5 male, while 5 were in the younger age group (18–30), 10 in the middle age group (31–60), and 5 were older citizens (>61 years old). Among the participants in the Brasov Citizens' Jury, there were unemployed, students, women on maternity leave, part-time workers, retired citizens, immigrants, people living in social or assisted living homes, people who completed only compulsory education, people with physical barriers, and people belonging to an ethnic minority. In Martin, out of the 19 participants, 12 were female and 7 were male, while 4 were in the younger age group (18–30), 9 in the middle age group (31–60) and 6 were older citizens (>61 years old). Similarly to Brasov, the Citizens' Jury participants in Martin spanned a variety of vulnerabilities. At the start of the workshop the participants were separated into 3 groups (no particular grouping criteria was followed), to facilitate the implementation of the group activities. A group leader was chosen by each group among the participants, who supported the moderators to fill-in the group activity form.

7.3 Main Outcomes of the Participatory Activities

From a deliberative democracy perspective, the participatory activities conducted within ITHACA can be interpreted not merely as mechanisms for civic engagement, but as tests of the conditions under which meaningful deliberation becomes possible for marginalized groups. In this sense, the Citizens' Juries reveal four core deliberative functions that AI-enhanced democratic systems must support: (i) voice and recognition, ensuring that participants can express experiences and viewpoints without fear or exclusion; (ii) quality of participation, enabling citizens to understand, contribute to, and contest discussions despite linguistic, cognitive, or digital barriers; (iii) consequentiality, making participation meaningful through traceability of inputs and visibility of outcomes; and (iv) safeguards, protecting participants from harm, discrimination, or misuse of data through moderation, transparency, and human oversight. The following outcomes are therefore read through this deliberative lens, rather than as isolated usability or accessibility findings.

The participants in both pilot cities have been actively involved in the past in general civic activities, with over 70% of each Citizens' Jury members having voted and participated in volunteering and environment protection activities. As expect, time limitations are one of the main barriers for civic engagement, but other important factors have negative impact, including: (i) community factors (e.g. people they know and are close to, such as relatives, friends or neighbors, do not participate in such actions), (ii) education and missing knowledge and skills (e.g. digital skills), (iii) residence-related factors (e.g. limited initiatives at local level), and (iv) financial limitations. It is important to observe that physical/cognitive or health issues are not seen as limiting factors by the participants.

A high percentage (>25%) of the participants in both cities were not aware of the possibility to use digital technologies and tools for civic participation, and only 45% of the participants in Brasov and 30% in Martin have used online platforms in the past. The participants highlighted the fact that the registration process complexity is an important barrier that is holding them back from engaging more in online civic activities and projects. They are missing the necessary digital skills, encounter difficulties to locate the information due to complexity of platforms, and overall find it difficult to follow or subscribe to the topics of discussion within their interest.

The perceived importance of civic participation for the local community and the neighborhoods is primarily linked to (i) improved quality of life, (ii) being able to express opinions and being heard, (iii) active involvement of citizens in the decision making, and (ii) proper function of society.

When it comes to AI, many of the participants (~70%) see it as being both a facilitator and a barrier for civic engagement. The main reasons for which AI is perceived as a facilitator include: increased speed and improved access to data and information, improved (potentially impartial) decision-making processes and quicker problem solving resulting in improved quality of every-day life, and eventually reduced bureaucracy and quicker communication between citizens and administration. The main reasons for which AI is perceived as a barrier include: potential intentional and non-intentional misuse of data, platform and tools complexity which may require advanced skill and knowledge, and potential technical limitations (e.g. backend connectivity, incapacity to adapt to the local context).

The integration of a Personal Information Management System (PIMS) was mainly seen as a positive factor, being justified by the responsible use of the platform, as the citizens undertake responsibility for their actions/postings, the possibility to provide personalized and faster services and content/information, and ensure data privacy and security. Among the reasons for not using a PIMS, the insecurity and fear of potential misuse of personal data along with the re-assurance provided by anonymity (e.g. feeling more confident to make negative critique in anonymous mode, fear of potential social consequences from public administration when speaking out) were the most important.

The vast majority (~90%) of the Citizens' Juries participants, in both pilot sites, have perceived some negative impact (e.g. felt vulnerable, discriminated against or were emotionally negatively affected) as a result of their active involvement in public and/or online discussions or debates in the past. The main occasions and reasons include:

- Feelings of emotional distress and negative impact on moral, when trying to express an opinion in public as a citizen belonging to a minority group,
- Feelings of being ignored, no matter how simple are the topics under discussion (e.g. even regarding weather, health, diet, education and society, etc.),
- Feelings of being judged against some standards (e.g. knowledge, education) when giving some feedback, both negative or positive, or relaying a personal experience,
- Feelings of frustration when initiatives and projects potentially very important for a particular community, fail to be realized (e.g. adoption of a 'mass-centered approach' instead of problem importance).

When it comes to awareness of own rights, only 1/3 of the participants in the Citizens' Jury in Brasov were aware of their own rights when it comes to discriminatory behavior, while in the case of Martin the percentage is much higher (63%). However, in both cases only 50% of them reacted or have taken action to fight back discrimination. The main reasons for not taking any actions or reacting to fight back discrimination included: missing knowledge regarding the procedures and rights, the bureaucratic process and the time and resources required, feeling ashamed or intimidated, fear of potential repercussions, perceived difference in social position/status linked to highly decreased chances to get a positive result of any action.

All participants highlighted as being extremely important to be able to identify potential harms, biases and discrimination related to online civic activities, in order to safeguard yourself online, in particular regarding the misuse of personal data. The main enablers for the citizens to be able to detect/report harms and enforce their rights are:

- Continuous information, the mass-media being mentioned as one important channel to be used in this scope,
- Provision of online platforms with control mechanisms, in particular for protecting personal data,
- Existence of a multi-level (from simple language to complex information) information/explanatory approach for terms and conditions of use,
- Making everything more transparent,
- Using a clear and concise language,
- A system to monitor, report, provide feedback and moderate the usage of personal data,
- The possibility to receive human feedback on request, and
- Alignment between recommendation and implementation when it comes to data privacy.

The information about the online platforms and tools for civic engagement that would help the participants to identify potential harms, and feel safer when engaging in civic actions online, include:

- Detailed information about the platform and the AI/smart tools employed,
- Educational materials,
- Information regarding the legal and regulatory aspects of the platform/tools,
- Sources of information used by the platform and AI tools,
- Information about where and for how long are the personal data stored,
- How and who decides how the personal data is used (e.g. by AI tools),
- Clear description of the benefits of personal data usage, and
- Existence of human assistants (e.g. oversight, guarantee of trust) supporting the AI-powered functionality.

The requirements analysis, regarding potential AI-powered features and tools that would help them engage more in online civic activities, showed as most preferred:

- Security measures (e.g. personal data storage and use)
- Text translation in other language
- Toxicity sensor/Spam-/Phishing detection
- Language simplification
- Recommendations (e.g. personalized)
- Auto Tagging
- Sentiment Analysis
- Information filtering
- Interactive tools (polls, visualizations)
- Comment moderation.

Read through a deliberative lens, the outcomes of the ITHACA Citizens' Juries highlight how inclusion alone is insufficient if not coupled with conditions enabling meaningful deliberation. Participants' experiences of being ignored, judged, or emotionally affected point to failures of voice and recognition, particularly for individuals belonging to marginalized groups. Difficulties related to language complexity, digital literacy, and platform navigation reveal that the quality of participation depends not only on access, but on citizens' capacity to understand, articulate, and contest positions within deliberative settings. Equally significant is the emphasis placed on traceability and feedback—participants consistently stressed the need to see how their contributions move through stages of review and decision-making, underscoring consequentiality as a prerequisite for perceiving participation as legitimate rather than symbolic. Finally, concerns around data misuse, discrimination, and exposure to harm foreground the role of safeguards, including moderation mechanisms, transparent data practices, and the availability of human feedback, as central to sustaining trust and willingness to deliberate—and keep deliberating. Together, these findings position ITHACA as an inclusion-oriented civic tech initiative which concretely articulates the deliberative requirements that AI-enhanced democratic systems need to satisfy when engaging vulnerable and underrepresented populations.

Overall, ITHACA's experimentation underscores how inclusion goes beyond auxiliary design requirement, serving as a democratic precondition for AI-mediated participation and deliberation. Without accessible interfaces, confidence-building safeguards, and sensitivity to lived vulnerabilities, civic engagement risks remaining symbolic, socially skewed, or systematically dominated by already empowered groups. In this sense, ITHACA operationalizes HCAI through a combined strategy of participatory governance—most notably Citizens' Juries and iterative co-creation—and AI-enabled accessibility measures (e.g., language simplification, translation, text-to-speech and speech-to-text). At the same time, it surfaces deliberative requirements that cut across democratic settings, including transparent moderation, multi-level explainability of platform rules, and mechanisms for tracking whether contributions are received, reviewed, and acted upon. Crucially for the cross-project comparison developed in the following section, the ITHACA experience also makes visible a central HCAI tension: the very mechanisms that enable inclusion and personalization—such as identity management, data collection for targeted support, or PIMS-like solutions—may intensify privacy concerns and perceived risks of retaliation, especially for marginalized groups. This tension reinforces the need for human oversight, user control, and accountable governance as indispensable complements to accessibility-driven innovation.

8 Inclusive, Ethical, and Transparent AI. The Case of AI4GOV

The integration of AI systems into democratic governance processes demands a multidimensional approach that is both theoretically grounded and operationally responsible. At its core, this approach must address the foundational and conceptual counterparts between technological innovation and democratic values. AI systems are often optimized for efficiency, prediction, and automation, while democratic societies require systems that prioritize accountability, inclusivity, and the protection of fundamental rights. Hence, the development of AI in the public sector has to go beyond operational efficiency and include normative dimensions that align with the public interest and reflect collective values, ethics, and diverse societal needs.

Central to this transformation is the concept of HCAI, which highlights the active involvement of end users and stakeholders in the design, deployment, and evaluation of AI systems. This requires participatory mechanisms, accessible design interfaces, and the translation of abstract principles (such as bias, discrimination, fairness and transparency) into concrete, actionable design choices. Within the AI4GOV project, human-centeredness is operationalized through co-creation workshops, user-led evaluations, and the development of tools that facilitate meaningful public engagement. AI4GOV's contributions in this include the Holistic Regulatory Framework (HRF), which establishes the normative foundations for trustworthy AI in governance, and the Virtualized Unbiasing Framework (VUF) Catalogue, a tool designed to expose and mitigate algorithmic bias through an accessible, visual platform. Through these two components, the project introduces democratic oversight and social responsibility into the technical core of public sector AI.

8.1 Theoretical Foundations and Design Principles

The HRF developed in AI4GOV provides a conceptual and operational foundation for HCAI. The HRF was developed using a bottom-up and inclusive methodology that combines literature review, theoretical modeling, and extensive fieldwork. The framework pays special attention to underrepresented and vulnerable groups, aiming to minimize systemic bias and discriminatory outcomes in AI-supported governance. It is informed by interdisciplinary perspectives and introduces a set of normative and technical guidelines for building trustworthy, transparent, and inclusive AI systems for the public sector by always keeping the human in the center of the design. It integrates participatory principles into regulatory design, ensuring that citizens are not passive data points but active contributors to the governance and accountability of AI tools.

This approach reflects core HCAI design values:

- Transparency by design, ensuring AI systems can be scrutinized by non-technical stakeholders.
- Knowledge distribution and multidisciplinary perspective, raising awareness with regard to the AI systems and bridging the technical, research, and societal domains.
- Participatory governance, incorporating feedback loops that give stakeholders a say in system behavior.
- Bias mitigation, embedding fairness in data practices and model evaluation.

In more detail, the HRF developed under the AI4GOV project is a comprehensive framework designed to regulate the use of AI technologies in governance. Its primary purpose is to ensure that AI systems used in public services are developed, deployed, and operated in a manner that is ethical, transparent, and compliant with applicable laws and regulations such as the GDPR addressing in parallel the issues of bias and discrimination. The HRF also incorporates ethical recommendations from various AI bodies, including the High-Level Expert Group on Artificial Intelligence (HLEG) and aligns with regulations such as the GDPR and the AI Act. The HRF is structured around fifteen key dimensions that guide the governance of AI systems. These dimensions include regulations compliance, fairness, non-discrimination and inclusivity, human oversight, human-centered design, privacy, explainability and transparency, auditing and continuous monitoring, accountability, safety and security, claims and redress, public engagement, awareness, societal benefit and well-being, and sustainability. Each dimension addresses specific aspects of AI governance to ensure that AI technologies are used responsibly and equitably.

To achieve the key dimensions and finalize the HRF, a multi-level mixed-methodology approach was employed. This comprehensive process included literature reviews of relevant bibliography and legislation, Delphi rounds with experts, SWOT analysis, surveys, focus groups, and interviews, as depicted in the below figure. In that direction, the HRF bridges insights from literature, expert input, and perspectives from citizens with various backgrounds (Fig. 6).

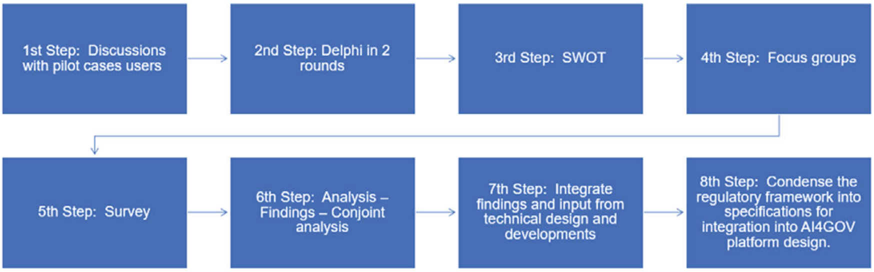

Fig. 6 Holistic regulatory framework (HRF) co-creation approach

In more detail this methodology outlines a structured, multi-step process aimed at developing and integrating a regulatory framework into the AI4GOV platform design ensuring inclusiveness, representation and human-centeredness. It begins with discussions involving users from the pilot cases of the project, ensuring that stakeholder needs, expectations, and context-specific insights are captured at the outset. These qualitative inputs inform the next stage, a two-round Delphi process, which engages expert participants to converge on key priorities, challenges, and opportunities through structured iterative feedback.

Following the Delphi rounds, a SWOT analysis is conducted to assess the internal strengths and weaknesses, as well as the external opportunities and threats relevant to the adoption of AI technologies in public administration. These insights are then deepened through focus groups, which facilitate more targeted and dialogue-based exploration of stakeholder perceptions and domain-specific concerns.

To complement these qualitative approaches, a series of surveys with general public and underrepresented citizens is conducted in Greece, Spain and Slovenia to collect broader quantitative data. This data is analyzed using conjoint analysis to reveal preferences and trade-offs stakeholders make when considering different design or policy options. The analysis phase synthesizes the findings from all previous steps to generate actionable insights.

These insights are then integrated with technical design and development considerations, ensuring alignment between user-centered findings and the architectural and functional decisions of the platform. The final step of the process involves condensing the regulatory framework into concrete technical specifications, allowing it to be directly embedded into the AI4GOV platform's overall design. This methodological progression ensures a balance between participatory input, empirical analysis, and technical feasibility, thereby supporting the creation of an inclusive, trustworthy, and policy-aligned AI governance solution.

The AI4GOV HRF aims to establish a fair, transparent, and inclusive regulatory environment for public organizations ensuring they serve the public good while safeguarding individual rights and societal values. More specifically, HRF's core purpose is to provide a holistic approach to AI governance which ensures the protection of citizens from potential misuse and biases associated with AI technologies in public services, and public decision-making as well as discrimination.

Its key objectives and goals are the following:

- Ensure regulatory compliance: Guarantee that AI systems adhere to all relevant laws, standards, and ethical guidelines.
- Promote fairness and equity: Design AI systems to treat all individuals equitably and reduce potential biases.
- Foster non-discrimination and inclusivity: Prevent discrimination against any individual or group and promote inclusiveness in AI development and deployment.
- Implement human oversight and control: Establish mechanisms for human intervention in AI decision-making processes.
- Enhance transparency and explainability: Ensure AI decision-making processes are transparent and explainable.
- Protect privacy: Implement robust measures to protect individual privacy and handle data securely.
- Ensure safety and security: Design AI systems with measures to prevent harm and protect against misuse.
- Encourage public engagement and awareness: Engage stakeholders and raise awareness about AI capabilities and limitations.
- Establish accountability and redress mechanisms: Define clear accountability mechanisms and provide channels for individuals to seek redress.
- Promote sustainability and societal benefit: Prioritize the development of AI systems that promote societal benefits and are environmentally sustainable.

By integrating these objectives, the HRF aims to create a regulatory framework that supports ethical, inclusive, and transparent AI governance. The vision of the HRF is to create a framework that will be continuously updated and strengthened in response to advancements in AI technology that enhance public services while protecting the rights and interests of all citizens, promoting a more equitable and just society.

The HRF prescribes certain dimensions related to AI governance when AI is involved to a lesser or to a greater extent at services that are provided to the public by governmental entities. These dimensions are the following:

- **Regulations compliance**: Ensuring that the AI system complies with all relevant laws, standards, ethical guidelines and codes of conduct. For example, compliance with the EU Charter of fundamental rights, compliance with EU AI Act, compliance with EU Data Act, etc.
- **Fairness**: The AI system should be designed and deployed in a way that treats all individuals fairly and equitably, even if the bias is not related to protected characteristics (gender, age, religion, disability, etc.). It considers the overall distribution of benefits, burdens, and outcomes across different groups and communities, and aims to prevent or mitigate any unfair disparities or disadvantages.
- **Non-discrimination—inclusivity**: The AI system should be designed and deployed in a way that promotes inclusivity and does not discriminate against or exclude any individual or group based on protected characteristics.

- **Human oversight**: Appropriate human oversight and control mechanisms should be in place to monitor the AI system's decision-making processes, intervene when necessary, and ensure that the system is operating as intended and in accordance with ethical principles.
- **Human-Centered design**: The AI system should be designed with a focus on human needs, values, and well-being, considering human factors such as usability, accessibility, and user experience.
- **Privacy**: The AI system should be designed and deployed in a way that respects and protects the privacy of individuals, including through measures such as data minimization, anonymization, and secure data handling practices.
- **Explainability—transparency**: The AI system's decision-making processes should be transparent and explainable to relevant stakeholders, allowing for scrutiny, accountability, and trust-building.
- **Auditing—continuous monitoring**: There should be mechanisms in place for continuous monitoring and auditing of the AI system's performance, outputs, and impacts, enabling timely detection and correction of issues or unintended consequences.
- **Accountability**: Clear lines of responsibility and mechanisms for holding entities accountable should be established for the development, deployment, and use of the AI system. Safety and security: The AI system should be designed and deployed with robust measures to ensure safety and security, protecting against unintended harm, misuse, or malicious attacks.
- **Claims and redress**: There should be clear and accessible mechanisms for individuals to seek redress and compensation in cases where they have been adversely affected by the AI system's decisions or actions.
- **Public engagement**: Stakeholder engagement and public participation should be encouraged throughout the AI system's development and deployment, fostering transparency, trust, and alignment with societal values and needs.
- **Awareness**: Efforts should be made to raise awareness and promote understanding of the AI system's capabilities, limitations, and potential impacts among relevant stakeholders and the public.
- **Societal benefit and well-being**: The development and deployment of the AI system should be guided by a commitment to promoting societal benefit and well-being, considering both short-term and long-term impacts.
- **Sustainability**: The AI system should be designed and deployed with consideration for environmental sustainability, resource efficiency, and long-term viability, minimizing negative impacts on the environment and promoting sustainable practices.

While the below figure shows at the perimeter the above dimensions and at the center systems that should facilitate them such as the governance and administration system, the AI ethics and compliance system, the auditing and monitoring system, the explainable system and the AI registry system (Fig. 7).

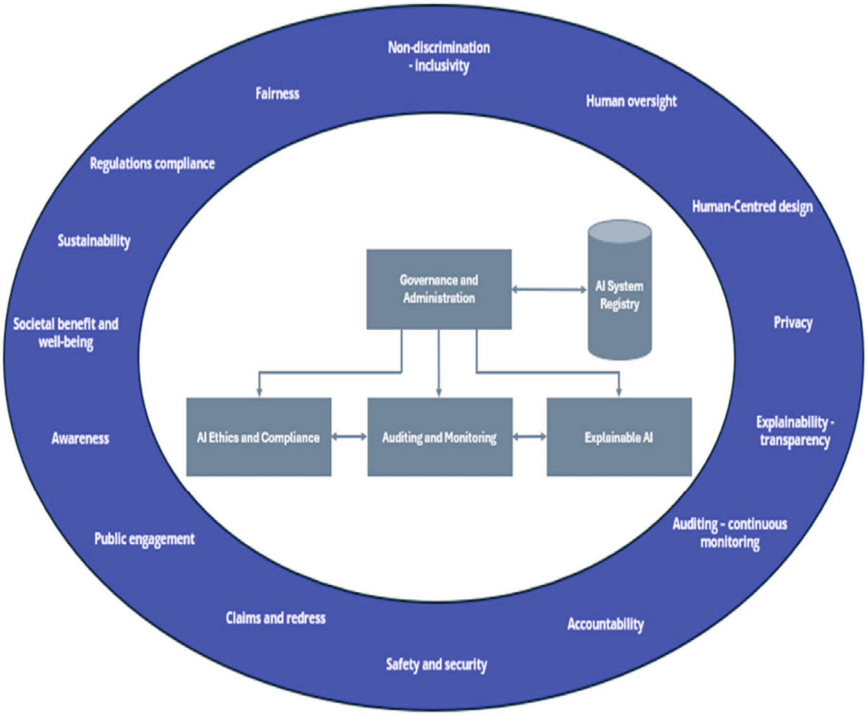

Fig. 7 The 15 regulatory dimensions as defined in the context of AI4GOV's HRF

To bridge the gap between the theoretical foundations of the Holistic Regulatory Framework (HRF) and practical implementation, the AI4GOV project has developed a comprehensive self-assessment tool. This tool enables users to systematically evaluate whether their AI systems adhere to the HRF's provisions across all stages of the AI lifecycle, including design, data processing, model training, deployment, and ongoing monitoring. The assessment is structured around 15 regulatory dimensions, each with specific criteria measured through a combination of a structured Likert scale (1–7) and binary compliance indicators (Yes/No). This dual approach provides both nuanced insights and clear compliance checks, ensuring that human rights considerations are integrated throughout the development and operation of AI systems.

The self-assessment tool employs an advanced scoring system that categorizes results into three compliance levels: non-compliant (0–40%, high risk), partial compliance (40–70%, needs improvement), and compliant (70–100%, low risk). Although still under development, the tool is designed to function as a proactive governance mechanism, offering development teams data-driven insights, automated regulatory tracking, and structured validation of compliance. By embedding these features, the AI4GOV self-assessment tool not only supports organizations in safeguarding human rights but also fosters a culture of ethical, transparent, and trustworthy AI development.

The categorization of the self-assessment questions reflects the breadth and depth of the HRF. Key categories include regulatory compliance (e.g., EU AI Act, GDPR), fairness and non-discrimination, human oversight, data privacy, explainability, safety, accountability, public engagement, and sustainability. Each category delves into specific issues such as bias detection in datasets, user rights awareness, consent and security practices, and the presence of human-in-the-loop mechanisms. The scoring system normalizes all responses on a 0–1 scale, facilitating the generation of percentage-based compliance levels. These are then grouped into risk tiers (non-compliant, partial compliance, compliant) that allow developers and regulators to benchmark system maturity and address deficiencies. This structured evaluation is not static; it incorporates adaptive mechanisms to evolve with changes in law, technology, and ethical standards.

This work has also resulted in a best practice guide, structured around the HRF dimensions. The development process of any AI system that aims to be used by citizens while provided by government entities should be conceptualized, designed, implemented and deployed under continuous scrutiny regarding ethical aspects related with its prospected operation.

Structured in categories aligned with the HRF, a set of best practices applicable to the development of such systems follows (Table 5).

AI systems provided by government entities require proactive governance that evolves with technological capabilities. The integration of these practices early in development cycles will result in developing public confidence and harnessing AI's potential for the social good. Regular updates are essential as both technology and public expectations advance.

8.2 Operationalizing Citizen Participation and AI Transparency Through Blockchain Technology

A key aspect in the overall human-centric approach of the AI4GOV project is the ability of citizens and policymakers to join the platform and leverage the blockchain capabilities to participate in open democracy schemes. The general mechanism that allows such self-governing bodies to form and exist is the so-called Digital Autonomous Organization or DAO. The vision behind DAOs is to provide a means of self-governance to end users. From an eGov perspective, DAOs are the means by which citizens can participate in an inclusive and open scheme of governance. This requires a different vision behind the governance model of the blockchain and a different technical solution to provide the capability for citizens to identify themselves and prove aspects of their identity. The below figure shows a scenario more typical in DAOs. A member proposes a change, such as the modification of tax

Table 5 The categories of the HRF and best practises that can be applied to them

Category	Best practice
1. Regulations compliance	• Ensure adherence to national, European and international AI-related regulations • Keep the corpus of relevant regulations regularly updated • Conduct legal reviews at all development stages
2. Fairness	• Use unbiased datasets and continuously test for bias • Employ fairness-aware algorithms and techniques to mitigate discrimination • Conduct diverse stakeholder reviews to ensure fair outcomes
3. Non-discrimination and inclusivity	• Design AI systems that are inclusive and accessible to all societal groups • Perform impact assessments on marginalized communities • Engage diverse user groups during system testing
4. Human oversight	• Ensure human-in-the-loop mechanisms for critical decision-making • Define clear escalation paths for AI-generated outcomes • Provide training for personnel interacting with AI systems
5. Human-centered design	• Prioritize user experience and ethical considerations in AI system design • Conduct iterative user testing and feedback loops • Align AI goals with human values and societal needs
6. Privacy	• Implement privacy-by-design principles • Use anonymization and differential privacy techniques • Comply with GDPR
7. Explainability and transparency	• Develop inherently interpretable models where possible • Provide documentation on AI decision-making processes • Ensure users can understand AI-generated outcomes
8. Auditing and continuous monitoring	• Establish independent auditing mechanisms • Monitor AI performance and behavior over time • Regularly update AI models based on new data and ethical considerations
9. Accountability	• Clearly define roles and responsibilities within AI development teams • Implement mechanisms for tracking decision-making processes • Assign liability in case of AI-related failures
10. Safety and security	• Follow robust cybersecurity protocols • Use adversarial testing to identify vulnerabilities • Implement fail-safe mechanisms to prevent harm

(continued)

Table 5 (continued)

Category	Best practice
11. Claims and redress	• Provide clear processes for users to challenge AI decisions • Establish mechanisms for correcting erroneous decisions • Offer legal recourse options when necessary
12. Public engagement	• Involve citizens in AI policy discussions • Organize public consultations and transparency initiatives • Communicate AI system capabilities and limitations clearly
13. Awareness	• Implement AI risk-awareness training for developers • Launch awareness campaigns to inform the public • Develop documentation and explainers on AI limitations
14. Societal benefit and well-being	• Design AI to improve social welfare • Prioritize ethical use cases aligned with public good • Conduct ethical impact assessments before deployment
15. Sustainability	• Optimize AI energy consumption • Consider environmental impact when selecting models and infrastructure • Promote AI solutions that contribute to sustainability goals

Synthesis table outlining the fifteen categories of the Holistic Regulatory Framework (HRF) and the associated best practices for each. The table links regulatory and ethical categories—such as regulatory compliance, fairness, inclusivity, human oversight, privacy, explainability, accountability, public engagement, societal well-being, and sustainability—to concrete practices that guide the design, development, deployment, and monitoring of AI systems in line with democratic values and trustworthy AI principles.

rates. A special body of the blockchain approves or disapproves the change; if it is approved, it is sent back to the community for voting (Fig. 8).

DAOs seem perfect tools to decentralize an economy. Expanding this into politics, it may seem that DAOs can also be used to decentralize democracy. This may be a misnomer, since democracy is, by definition, a decentralized system of governance. However, aspects of it, such as participation in opinion forming, transparency of processes, verification of candidate claims etc., may be hindered or obfuscated due to various reasons, such as real-life limitations, limited access to original sources, difficult to identify deep fakes etc. It is to be noted that the term DAO is an umbrella term similar to the term smart contract. It does not refer to any specific functionality but to components implemented via smart contracts and can be self-governed. DAOs can help citizens to somewhat limit the effect of such hindrances by facilitating transparency in decision making and providing more tools for citizen co-creation in policy making. However, it should be noted that the DAOs should ensure that no security breaches and threatening transactions take place and that the governance model

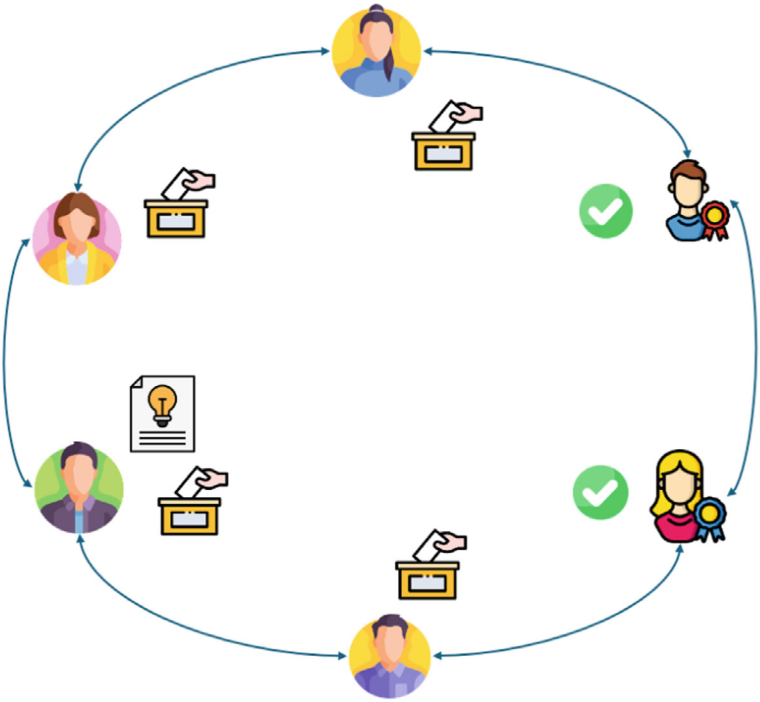

Fig. 8 The utilisation of DAOs and blockchain technology in the eGovernment domain

allows for no such situations to take place. Thus, a strong cybersecurity framework should be also established along with DAOs.

In addition to the above, citizen-centric wallets using the HyperLedger Aries framework have been implemented. The latter is a framework specifically designed for identity management. Aries supports the creation and administration of Decentralized Identifiers (DIDs) that can uniquely be resolved to each subject. Moreover, under Aries, the issuance and verification of credentials and attributes are straightforward. Typical use cases of Aries that are key to implementing mechanisms for citizen trust and participation include:

- Issuing of Self-Sovereign Identity (SSI) credentials which enable citizens to control their own identity.
- Evidence/credential verification mechanisms that allow wallet holders to present and provide proof of credential ownership (e.g., driving license) or attributes (e.g., legal age).
- Mechanisms for providing fine-tuned access to credentials based on the identity of the requesting party.

Finally, the utilisation of the blockchain technology inside the AI4GOV project is also used as a technology enabler to provide transparency in AI output, due to its decentralized and transparent nature. Since current LLMs are trained over a huge corpus of data and use millions or even billions of parameters for producing output, the 'straightforward' way of providing transparent AI by having each node validate the results of training and/or execution is, of course, not applicable. Two main techniques are relevant and have been applied:

- Anchoring the explainability report to any claim or document that is produced by AI. Although a claim may still be fake, producers who are confident in the truth of their statements can anchor explainability reports as further proof of their claims. This will help well-documented claims gain more visibility
- Using DIDs for the AI. By assigning to a generative AI an identity, the AI can behave as a physical actor in the blockchain. AI that has identity can be more easily trusted by citizens; this can create a trend in which AIs with identity are progressively preferred. The benefit of being able to discern the identity of an AI is that a consumer can check its parameters such as training algorithm, past predictions, previous bias reports etc.

9 Conclusions

To conclude this chapter, the insights drawn from the four Horizon Europe projects are recognized to provide a privileged vantage point for understanding how Human-Centered AI is being operationalized within the domain of democratic innovation. Rather than offering abstract principles, the projects are contributing by demonstrating how HCAI takes shape through concrete design choices, participatory arrangements, and governance mechanisms embedded in real-world democratic settings.

By adopting a comparative lens, this section sheds lights on converging trajectories and complementary orientations across the projects. On the one hand, shared commitments emerge around participation, transparency, accountability, and inclusiveness as non-negotiable conditions (and precondition) for legitimate AI-enabled democratic processes. On the other hand, it emerges the presence of different leverage points—ranging from deliberative tooling and co-creation (ORBIS), to critical digital literacy and collective sensemaking (KT4D), inclusion-first design for marginalized groups (ITHACA), and regulatory and infrastructural governance mechanisms (AI4GOV).

What is relevant to stress is how taken together, these projects do not converge on a single model of HCAI for democracy. Instead, they collectively articulate a plural yet coherent and consistent design space, in which democratic values are operationalized through diverse but mutually reinforcing approaches. This comparative perspective provides the basis for identifying key directions in the operationalization of HCAI for AI-enhanced deliberation, while also clarifying where different projects intentionally diverge in response to distinct democratic challenges.

9.1 Comparing HCAI Approaches

This paragraph presents a comparative reading of how the four projects relate to the topic of HCAI for democracy. It adopts a qualitative, theory-driven synthesis grounded in the HCAI dimensions proposed by Régis et al. (2024)—as introduced in par. 5.3.1 and mapped to AI-enhanced democracy in Table 1—and draws on the structured project descriptions presented in paragraphs from 5.5 to 5.7. Rather than running a checklist for compliance against each project, the analysis is intended to empirically examine in a comparative lens how each of them operationalizes, combines, and prioritizes specific dimensions in response to distinct democratic challenges (Table 6).

Viewed comparatively, the four projects articulate distinct yet complementary HCAI configurations. In particular, it is worth noting that no single project answers all the HCAI dimensions mapped in Table 1. Instead, each project emphasizes different combinations, producing complementary approaches to AI-enhanced democracy: ORBIS focuses on deliberative outcomes and traceability; KT4D on epistemic agency and collective sensemaking; AI4GOV on rights-based governance and institutional accountability; and ITHACA on inclusion as a deliberative precondition. This plurality does not signal fragmentation, but rather reflects the multi-layered nature of democratic innovation with AI—where design, literacy, governance, and inclusion must be addressed simultaneously, though not necessarily by the same intervention.

This initial exploration of Table 5 is integrated with a comparative analysis which builds on the mapping of HCAI dimensions presented in Table 1 and the project-specific operationalizations summarized in Table 2. To move beyond descriptive alignment and enable analytical comparison, the projects were examined along four cross-cutting analytical axes: (i) where human-centeredness is primarily enacted (e.g. design processes, governance frameworks, capacity building, or infrastructural safeguards); (ii) which deliberative functions are foregrounded, drawing on deliberative theory (voice and recognition, quality of participation, consequentiality, and safeguards); (iii) the governance implications implied by each project's configuration of HCAI dimensions; and (iv) how the relationship between scale and democracy is conceptualized, particularly in relation to the role of AI in mediating participation beyond small-scale deliberative settings. This lens of observation allows for a systematic comparison that highlights both convergences and complementarities across projects, while making explicit the different normative and operational assumptions through which human-centered AI is mobilized in democratic innovation.

Axis 1—Where Human-Centeredness is Enacted

A central challenge in HCAI lies in determining where human-centeredness is practically realized within socio-technical systems. HCAI is often articulated normatively—as a commitment to human dignity, agency, and flourishing—but these values can be enacted at different layers of an AI system, ranging from interface design and

Table 6 Comparative synthesis of HCAI approaches across projects, mapped to Régis et al. (2024) dimensions

Comparative analytical dimension	Related HCAI dimensions (Table 1)	Comparative insight across projects
Entry point to HCAI	D1 Participatory design D2 Civic participation	Projects differ in where human-centeredness is 'activated': ORBIS and ITHACA embed participation directly in tool and process design; KT4D operates upstream by shaping citizens' and designers' critical capacities; AI4GOV anchors participation within governance and regulatory infrastructures
Primary deliberative function emphasized	D2 Civic participation D10 Algorithmic legitimacy	ORBIS foregrounds *consequentiality* (connecting deliberation to policy outputs); ITHACA foregrounds *voice and recognition*; KT4D foregrounds *collective judgment and reflexivity*; AI4GOV foregrounds *procedural legitimacy and accountability*
Treatment of epistemic asymmetries	D4 Decolonial ethics D8 Transparency and explainability	All projects address epistemic gaps, but differently: KT4D makes asymmetries visible through friction and literacy; ORBIS and AI4GOV mitigate them via explainability mechanisms; ITHACA reduces them through accessibility, mediation, and human support
Role assigned to AI in democratic processes	D7 Socio-technical systems D10 Algorithmic legitimacy	AI is not treated uniformly: ORBIS frames AI as a deliberative *facilitator*; KT4D treats AI as an *object of collective inquiry*; ITHACA treats AI as an *accessibility enabler*; AI4GOV treats AI as a *regulated institutional actor*
Safeguards against harm and exclusion	D3 Human rights-based approach D9 Accountability and oversight	ITHACA and AI4GOV institutionalize safeguards (rights, redress, governance); ORBIS relies on human-on-the-loop approach to development, and in human-in-the-loop in moderation; KT4D relies on reflective design choices and collective responsibility rather than formal enforcement
Governance model implied	D5 Adaptive governance D9 Accountability	ORBIS and ITHACA emphasize adaptive, context-sensitive governance; AI4GOV formalizes governance through regulatory and blockchain-based mechanisms; KT4D adopts a meta-governance stance, shaping how governance itself is understood and designed

(continued)

Table 6 (continued)

Comparative analytical dimension	Related HCAI dimensions (Table 1)	Comparative insight across projects
Approach to scale and transferability	D5 Adaptive governance D6 Policy evaluation	ORBIS and AI4GOV explicitly address scalability through modularity and infrastructures; ITHACA prioritizes contextual depth over scale; KT4D scales primarily through transferable principles, methods, and educational artefacts
Underlying conception of human-centeredness	All dimensions (D1–D10)	Human-centeredness is enacted as: *procedural co-creation* (ORBIS), *deliberative capability-building* (KT4D), *inclusion as democratic precondition* (ITHACA), and *rights-based institutional alignment* (AI4GOV)

Comparative table analysing how four Horizon Europe projects—ORBIS, KT4D, ITHACA, and AI4GOV—enact Human-Centered AI principles when mapped to the dimensions proposed by Régis et al. (2024). Across a set of analytical dimensions, the table compares how projects differ in their entry points to HCAI, the democratic functions they prioritise, their treatment of epistemic asymmetries, the roles assigned to AI, safeguards against harm, governance models, approaches to scale, and underlying conceptions of human-centeredness.

participatory processes to institutional governance and regulatory frameworks. Identifying the primary locus at which human-centeredness is operationalized is therefore analytically relevant, as it reveals underlying assumptions about how and by whom democratic values are protected and enacted. This axis allows us to distinguish between approaches that embed human-centeredness in *design and use practices*, those that emphasize *epistemic empowerment and literacy*, and those that prioritize *institutional safeguards and governance structures*. From a deliberative perspective, this distinction is crucial, as the site of human-centeredness directly affects who has agency, when intervention is possible, and how responsibility is distributed across actors in AI-enhanced democratic systems.

In ORBIS, human-centeredness is enacted primarily at the design–use interface of e-participation platforms understood as deliberative infrastructures. The project embeds citizens, facilitators, and institutional actors throughout the full lifecycle of AI-supported deliberation—from needs mapping and co-design to testing, deployment, and moderation—ensuring that system functionalities are shaped by situated deliberative practices rather than abstract or generalized user models. Human-centeredness is thus operationalized through iterative co-creation, contextual adaptation across heterogeneous pilots, and the continuous moderation and validation of AI-generated outputs. In this configuration, AI-enhanced tools are designerly shaped to remain intelligible, and aligned with the lived realities of democratic practice.

KT4D enacts human-centeredness primarily at the level of *epistemic capacity and collective sensemaking*. Rather than centering on deployable platforms, the project focuses on empowering citizens, policymakers, and technologists to critically engage with AI systems through the policy framework, capacity-building games, Digital Democracy Lab and accompanying facilitation materials. Human-centeredness is thus located in *how people understand, question, and negotiate* AI's role in democracy.

In AI4GOV, human-centeredness is enacted primarily through *institutional governance frameworks*. The Holistic Regulatory Framework (HRF) embeds HCAI principles into the regulatory, organizational, and procedural layers of public-sector AI systems. Citizens are positioned as rights holders whose interests are safeguarded through structured oversight mechanisms rather than continuous participation in system design.

ITHACA enacts human-centeredness primarily through *inclusion-oriented participatory processes*. Citizens' Juries, co-creation workshops, and iterative validation with marginalized groups ensure that AI-supported civic platforms are shaped by lived experiences of vulnerability, exclusion, and discrimination. Human-centeredness is grounded in social context rather than abstract usability.

Axis 2—Deliberative Functions Foregrounded

Deliberation is not a single activity but a constellation of interrelated functions—such as voice, recognition, understanding, discussion, contestation, consequentiality, and safeguards against harm—that together enable meaningful democratic participation. In the context of HCAI, identifying which deliberative functions are foregrounded by different systems is methodologically important, because AI technologies inevitably privilege certain forms of participation over others through their design choices, interaction models, and governance arrangements. This axis therefore examines how projects implicitly define what 'good' deliberation entails when mediated by AI. By analyzing whether systems prioritize inclusion and recognition, epistemic quality, traceability of outcomes, or protection against abuse and misuse, this dimension makes explicit the normative orientations embedded in technical and organizational decisions. Framing outcomes as deliberative functions rather than usability features also ensures alignment with democratic theory, allowing civic technologies to be assessed not only in terms of access or efficiency, but in terms of their capacity to sustain legitimate public reasoning.

ORBIS foregrounds deliberative functions related to value construction, shaped as quality of participation, explanation, and consequentiality, explicitly aiming to move beyond the production of technological outputs as an end in itself. Explainability mechanisms and structured representations of discourse are designed to help participants and facilitators understand how contributions are processed, aggregated, and translated into policy-relevant insights. A combined human-in-the-loop and human-on-the-loop approach safeguards deliberative quality by ensuring that AI outputs do not displace human judgment, flatten disagreement, or obscure minority positions. While voice is supported through inclusive design choices, its democratic legitimacy is reinforced primarily by making participation visibly consequential—through traceability, synthesis, and feedback loops that render deliberative outcomes intelligible and accessible to different audiences, including participants, moderators, and decision-makers.

KT4D foregrounds quality of participation and voice, understood as informed, reflective, and collectively situated engagement. Through the deliberate introduction of 'meaningful friction,' KT4D resists the tendency to optimize for convenience and instead promotes reflection, contestation, and reciprocal justification. Deliberation is framed as a process of collective reasoning rather than efficient opinion aggregation.

AI4GOV most strongly foregrounds safeguards and consequentiality. Transparency, accountability, redress, and human oversight are treated as foundational requirements for legitimate AI-supported governance. While direct deliberation is not always the focal point, the project reinforces the conditions under which deliberation can occur meaningfully by ensuring that AI-driven decisions remain contestable and accountable.

ITHACA foregrounds voice and recognition and safeguards as preconditions for deliberation. Findings from engagement with the Citizens' Juries show that participation is undermined when citizens feel ignored, judged, or exposed to harm. AI-enabled accessibility features can and should support quality of participation, but only when coupled with transparency, moderation, and human feedback.

Axis 3—Governance Implications of HCAI Configurations

HCAI is inseparable from questions of governance, as decisions about how AI systems are designed, deployed, and overseen directly shape power relations, accountability, and democratic legitimacy. Different configurations of HCAI dimensions imply different models of governance—ranging from embedded oversight within deliberative processes, to formal regulatory compliance, to capacity-building and epistemic governance. This axis is analytically relevant because it surfaces how projects conceptualize responsibility, control, and participation in AI-enhanced democratic systems. By comparing governance implications, the analysis moves beyond surface-level descriptions of participatory practices to interrogate how authority is allocated, how harms are addressed, and how legitimacy is produced. In deliberative contexts, this is particularly salient: governance models determine whether AI reinforces democratic institutions, redistributes agency, or introduces new asymmetries under the guise of neutrality or efficiency.

ORBIS implies a governance model based on embedded accountability, where oversight is integrated directly into deliberative workflows rather than delegated to external audits or post-hoc control mechanisms. Moderators and institutional actors play a central role as stewards of the deliberative process: they remain central in supporting interpretation, validation, and contextualization of AI-assisted outputs, and for explicitly signalling what the AI is doing at different stages of discourse processing. While participants are granted direct access to AI-generated analyses derived from collective deliberation, these outputs are presented as decision-support artefacts, not authoritative conclusions. This configuration reinforces democratic accountability by keeping responsibility anchored in human actors and institutional practices, while preserving institutional legitimacy through transparency, traceability, and the possibility of contestation.

KT4D implies a governance model oriented toward *democratic capacity-building*. Rather than prescribing regulatory solutions, it emphasizes the need for informed publics capable of engaging with and shaping AI governance. Governance, in this configuration, is not only institutional but pedagogical, relying on critical digital literacy as a precondition for legitimate democratic oversight.

AI4GOV implies a governance model based on *procedural legitimacy and regulatory alignment*. Democratic values are operationalized through the adoption of the Quintuple Helix innovation model, as well as compliance with legal frameworks, continuous monitoring, and institutional accountability structures. Governance is thus formalized, scalable, and aligned with public-sector responsibilities.

ITHACA implies a governance model centered on *protective and participatory accountability*. Governance mechanisms must be accessible, responsive, and sensitive to power asymmetries. Importantly, ITHACA highlights tensions between personalization and privacy, underscoring the need for governance arrangements that balance support with protection from retaliation or misuse.

Axis 4—Conceptualizations of the Relationship Between Scale and Democracy

A recurring tension in AI for democracy concerns the relationship between scale and deliberative quality. AI is often justified as a means to scale participation, process large volumes of input, and support decision-making in complex governance environments. However, democratic theory cautions that scaling participation can undermine key deliberative conditions, such as mutual understanding, recognition, and accountability. This axis is therefore essential for examining how HCAI approaches negotiate the trade-offs between inclusiveness, efficiency, and democratic depth. By analyzing how projects conceptualize scale—as an opportunity, a risk, an institutional necessity, or a vulnerability multiplier—this dimension reveals their implicit assumptions about what democracy can and should look like in AI-mediated contexts. Methodologically, this axis enables a critical comparison of whether AI is framed as augmenting deliberation without eroding its normative foundations, or whether democratic values are reinterpreted to fit the constraints of large-scale technological systems.

ORBIS conceptualizes scale primarily as a deliberative and epistemic challenge. AI is mobilized to support collective sensemaking and synthesis across large volumes of contributions, but always recalling the need for explicit human supervision and interpretive control. Scaling deliberation, in this view, requires preserving intelligibility, traceability, and plurality as participation grows, rather than maximizing automation or throughput. This approach treats scale as more than a matter of processing capacity: it becomes a question of deliberative literacy. Participants and facilitators are supported in understanding what happens to contributions at scale, how AI mediates aggregation and synthesis, and what is gained or lost in the process. AI thus becomes a means for rendering large-scale deliberation legible, rather than an invisible mechanism that abstracts democratic processes away from public understanding.

KT4D conceptualizes scale cautiously. Rather than seeking to directly scale deliberation through AI, it interrogates how large-scale AI systems reshape the epistemic conditions of democratic participation. Scale is treated as a source of risk—particularly epistemic asymmetry—requiring deliberate design interventions that slow down, expose, and contextualize AI-mediated processes.

AI4GOV conceptualizes scale as an *institutional necessity*. AI is considered as an integral part of the modern large-scale governance and policymaking ecosystems, and the democratic challenge lies in ensuring that scaling does not erode rights, fairness, or accountability. The HRF positions scale as manageable through robust governance architectures rather than through participatory intensity alone.

ITHACA treats scale as a *fragile condition* for deliberation. While AI can lower barriers to participation at scale, the project demonstrates that scaling without adequate safeguards risks reproducing or amplifying exclusion. Democratic scaling, in this view, requires prioritizing safety, trust, and recognition over sheer participation numbers.

9.2 The Governance Lens

Before turning to a more systematic treatment of democratic AI governance, which is addressed in Chap. "Future Research Directions", this section introduces governance as a focused analytical lens through which HCAI for democracy can be interpreted. Rather than offering a comprehensive governance framework, this chapter vets into on how governance concerns emerge from the interaction between HCAI principles and democratic practice, therefore approaching governance as a conceptual and normative dimension embedded in design choices, participatory arrangements, and socio-technical configurations. Differently and complementarily, Chap. "Future Research Directions" builds on this foundation to address governance more explicitly, examining institutional models, regulatory implications, and pathways for scaling AI-enhanced democratic systems.

In much AI discourse, governance is framed around managing risks and maximizing benefits, frequently under the banner of 'AI for the common good'. Yet this framing raises fundamental political questions about what counts as good, for whom, at what level of collective, and over what time horizon (Coeckelbergh 2025). Re-centering governance within HCAI therefore means re-politicizing the 'common good' claim (Offe 2012; Pettit 2004)—treating it as a site of democratic determination rather than a pre-defined objective to be implemented technocratically.

A key contribution of political-philosophical accounts of the common good is to clarify why current AI governance often suffers from a democratic deficit. Even when broad principles such as human rights, fairness, sustainability, transparency, or accountability are widely endorsed (UNESCO 2022), the crucial matter is how such general goods are/should be translated into operational meaning—and through which procedures. Procedural approaches emphasize that in pluralistic societies there is no final, uncontested definition of the common good; instead, legitimacy depends on inclusive public procedures that keep definitions open to revision and disagreement (Mouffe 1999). The risk here is not only poor or 'wrong' outcomes, rather an erosion of self-rule through a quiet shift toward technocratic closure (Coeckelbergh 2022). In this sense, the governance problem lies in the lack of democratically recognized pathways linking principles to design choices, model behaviors, institutional mandates, and accountability regimes.

A second, often under-theorized issue concerns who gets to define the collective implied by ‘common’, beyond the logics of regulation, sectoral standards, and corporate self-governance. A common-good framing pushes toward a multi-level conception of the collective—from local communities to national polities to humanity—without presuming these levels align neatly or that trade-offs can be avoided (Offe 2012; Pettit 2004). The same applies to time: the common good should be oriented to present populations, near-future generations, and longer horizons, and each choice reconfigures what counts as acceptable risk, acceptable surveillance, or acceptable automation (Coeckelbergh 2022). For AI-enhanced democracy, this matters because ‘scaling participation’ is often presented as inherently good, while the costs of scale—loss of recognition, procedural voice, or contestability—are distributed unevenly across groups and across time. A governance lens therefore requires HCAI commitments (rights, inclusion, explainability, accountability) to be interpreted as collective commitments that remain publicly revisable as contexts change.

From here, HCAI governance can be articulated as a shift from rule-setting about AI to ensuring that those affected can contest, question, and modify the norms embedded in systems that shape public reasoning. Contestability implies concrete design and institutional requirements—such as visibility of where and how AI intervenes, accessible avenues for disagreement, and pathways for revising system behavior and procedural rules—rather than relying on post-hoc transparency statements alone.

Within this frame, a central governance challenge for AI-enhanced democracy is preventing democratic displacement: the gradual substitution of political judgment with managerial optimization. This strengthens the rationale for deliberative innovations that place people at the core, informed and enabled to participate meaningfully. From this perspective, citizens are not only entitled to benefit from the common good, but also bear responsibilities in sustaining it through participation and contribution. Democratic governance therefore cannot rely solely on what institutions and companies do ‘for’ the public; it also depends on civic practices—how communities adopt, adapt, contest, and repurpose AI in everyday public life, and how civil society shapes alternative technological imaginaries. This points to the need for civic education and critical technological literacy that equips people to deliberate about AI and to recognize when ‘common good’ rhetoric is being used to foreclose political disagreement. Within HCAI for democracy, governance is thus best understood as a distributed achievement: a socio-technical arrangement in which institutions guarantee contestability and rights-based safeguards, while publics retain the knowledge, venues, and agency to co-determine how AI reshapes democratic life.

Overall, this governance lens clarifies how HCAI for democracy cannot be reduced to technical safeguards or compliance mechanisms, while leaving the broader governance implications—concerning institutional arrangements, regulatory frameworks, and the scaling of AI-enhanced deliberative systems—to be addressed in depth in

Chap. "Future Research Directions", where these insights are consolidated into a more systematic discussion of democratic AI governance.

Funding The reasoning presented in this work derive from knowledge and insights from four European projects that have received funding from the European Union's Horizon Europe Programme:

1. 'ORBIS. Augmenting participation, co-creation, trust and transparency in Deliberative Democracy at all scales' has received funding under Grant Agreement No. 101094765.
2. 'KT4D. Knowledge Technologies for Democracy' has received funding under Grant Agreement No. 101094302.
3. 'AI4GOV. Trusted AI for Transparent Public Governance fostering Democratic Values' has received funding under Grant Agreement No. 101094905.
4. 'ITHACA. artificial Intelligence To enHAnce Civic pArticipation' has received funding under Grant Agreement No. 101094364.

The opinions expressed herewith are solely of the authors and do not necessarily reflect the point of view of any EU institution.

AI Declaration AI tools were employed in the preparation of this chapter solely for proofreading, language refinement, and accessibility enhancement, not for generating original content or research results. The use of AI-based proofreading support (LLM specifically) was particularly valuable given that not all the authors are non-native English speakers, ensuring improved linguistic clarity, consistency, and readability without altering the scientific meaning or interpretation of the content.

References

Aldridge J (2014) Working with vulnerable groups in social research: dilemmas by default and design. Qual Res 14(1):112–130. https://doi.org/10.1177/1468794112455041

Bardzell J, Bardzell S (2015) Humanistic HCI. Springer, Cham, Switzerland. https://doi.org/10.1007/978-3-031-02214-2

Benedetti A, Mauri M (2023) Design for friction. An inquiry to position friction as a method for reflection in design interventions. Converg - J Res Arts Educ 16(31). https://doi.org/10.53681/c1514225187514391s.31.139

Bersini H, Badinet G (2023) Algocratie: Allons-nous donner le pouvoir aux algorithmes? De Boeck Supérieur, Louvain-la-Neuve, Belgium

Birhane A (2021) Algorithmic injustice: a relational ethics approach. Patterns 2(2). https://doi.org/10.1016/j.patter.2021.100205

Bohler-Muller Narnia (2005) The story of an African value: focus: ten years after Makwanyane. SA Publiekreg = SA Public Law 20(2):266–280. https://doi.org/10.10520/EJC97885

Bolte L, van Wynsberghe A (2025) Sustainable AI and the third wave of AI ethics: a structural turn. AI Ethics 5(2):1733–1742. https://doi.org/10.1007/s43681-024-00522-6

Bradley NC, Fritz T, Holmes R (2022) Sources of software development task friction. Empir Softw Eng 27(7):175. https://doi.org/10.1007/s10664-022-10187-6

Buckingham Shum S, De Liddo A, Klein M (2014) DCLA meet CIDA: collective intelligence deliberation analytics. In: 2nd international workshop on discourse-centric learning analytics, LAK14: 4th international conference on learning analytics and knowledge. https://catalyst-fp7.idea.kmi.open.ac.uk/catalyst-data/uploads/2014/05/dcla14_buckinghamshumdeliddoklein.pdf

Buhmann A, Fieseler C (2023) Deep learning meets deep democracy: deliberative governance and responsible innovation in artificial intelligence. Bus Ethics Q 33(1):146–179. https://doi.org/10.1017/beq.2021.42

Cabour G, Morales-Forero A, Ledoux É, Bassetto S (2023) An explanation space to align user studies with the technical development of explainable AI. AI Soc 38(2):869–887. https://doi.org/10.1007/s00146-022-01536-6

Cabrio E, Villata S (2018) Five years of argument mining: a data-driven analysis. In: Proceedings of the twenty-seventh international joint conference on artificial intelligence, IJCAI-18. International Joint Conferences on Artificial Intelligence Organization, pp 5427–5433. https://doi.org/10.24963/ijcai.2018/766

CAO—Japan's Cabinet Office (2019) Social principles of human-centric AI. https://www8.cao.go.jp/cstp/ai/humancentricai.pdf

Carayannis EG, Campbell DFJ (2010) Triple helix, quadruple helix and quintuple helix and how do knowledge, innovation and the environment relate to each other?: A proposed framework for a trans-disciplinary analysis of sustainable development and social ecology. Int J Soc Ecol Sustain Dev (IJSESD) 1(1):41–69http://doi.org/10.4018/978-1-4666-1586-1.ch004

Castelluccia C, Le Métayer D (2019) Understanding algorithmic decision-making: opportunities and challenges. European Parliament. https://www.europarl.europa.eu/RegData/etudes/STUD/2019/624261/EPRS_STU(2019)624261_EN.pdf

Cath C, Wachter S, Mittelstadt B, Taddeo M, Floridi L (2018) Artificial intelligence and the 'Good Society': the US, EU, and UK approach. Sci Eng Ethics 24(2):505–528. https://doi.org/10.1007/s11948-017-9901-7

Coeckelbergh M (2022) The political philosophy of AI: an introduction. John Wiley & Sons, Hoboken

Coeckelbergh M (2025) Artificial intelligence, the common good, and the democratic deficit in AI governance. AI Ethics 5(2):1491–1497. https://doi.org/10.1007/s43681-024-00492-9

Costanza-Chock S (2020) Design justice: community-led practices to build the worlds we need. The MIT Press. https://doi.org/10.7551/mitpress/12255.001.0001

Council of Europe (n.d.) Ethical frameworks—artificial intelligence—www.coe.int. Artificial Intelligence. https://www.coe.int/en/web/artificial-intelligence/ethical-frameworks. Accessed 8 Jan 2026

Coussement K, Abedin MZ, Kraus M, Maldonado S, Topuz K (2024) Explainable AI for enhanced decision-making. Decis Support Syst 184:114276. https://doi.org/10.1016/j.dss.2024.114276

Danaher J (2016) The threat of algocracy: reality, resistance and accommodation. Philos Technol 29(3):245–268. https://doi.org/10.1007/s13347-015-0211-1

de Sousa WG, de Melo ERP, Bermejo PHDS, Farias RAS, Gomes AO (2019) How and where is artificial intelligence in the public sector going? A literature review and research agenda. Gov Inf Q 36(4):101392. https://doi.org/10.1016/j.giq.2019.07.004

Deng G, Fei S (2023) Exploring the factors influencing online civic engagement in a smart city: the mediating roles of ICT self-efficacy and commitment to community. Comput Hum Behav 143:107682. https://doi.org/10.1016/j.chb.2023.107682

Dignum V (ed) (2019) Responsible artificial intelligence: how to develop and use AI in a responsible way. Springer International Publishing, Cham, Switzerland. https://doi.org/10.1007/978-3-030-30371-6

Duberry J (2022) Artificial intelligence and democracy: risks and promises of AI-mediated citizen-government relations. Edward Elgar Publishing, Cheltenham, UK. https://doi.org/10.4337/9781788977319

Duerinckx A, Veeckman C, Verstraelen K, Singh N, Van Laer J, Vaes M et al (2024) Co-creating artificial intelligence: designing and enhancing democratic AI solutions through citizen science. Citiz Sci: Theory Pract. https://doi.org/10.5334/cstp.732

Edmond J (2026) From 'Hello World' to 'How are We Doing?': Fostering collective sensemaking with and around the deafness of code. AI Soc. https://doi.org/10.1007/s00146-026-02991-1

Ekman J, Amnå E (2012) Political participation and civic engagement: towards a new typology. Hum Aff 22(3):283–300. https://doi.org/10.2478/s13374-012-0024-1

European Commission, Directorate-General for Communications Networks C, Technology, Inteligencji, G Ekspertów Wysokiego Szczebla ds Sztucznej (2019) Ethics guidelines for trustworthy AI. Publications Office. https://doi.org/10.2759/346720

Fernández-Martínez JL, Lopez-Sanchez M, Aguilar JAR, Rubio DS, Nemegyei BZ (2018) Co-designing participatory tools for a new age: a proposal for combining collective and artificial intelligences. Int J Public Adm DigitAge (IJPADA) 5(4):1–17. https://doi.org/10.4018/IJPADA.2018100101

Floridi L (2005) The ontological interpretation of informational privacy. Ethics Inf Technol 7(4):185–200. https://doi.org/10.1007/s10676-006-0001-7

Forbes-Mewett H, Nguyen-Trung K (2019) Defining vulnerability. In: Forbes-Mewett H (ed) Vulnerability in a mobile world. Emerald Publishing Limited. https://doi.org/10.1108/978-1-78756-911-920191004

Friedman B, Borning A, Davis JL, Gill BT, Kahn Jr PH, Kriplean T, Lin P (2008) Laying the foundations for public participation and value advocacy: interaction design for a large scale urban simulation. In: Proceedings of the 2008 international conference on digital government research. CiteSeer, pp 305–314

Galais C, Fernández-Martínez JL, Font J, Smith G (2021) Testing the input-process-output model of public participation. Eur J Polit Res 60(4):807–828. https://doi.org/10.1111/1475-6765.12427

Garcia Valencia OA, Thongprayoon C, Miao J, Suppadungsuk S, Krisanapan P, Craici IM et al (2024) Empowering inclusivity: improving readability of living kidney donation information with ChatGPT. Front Digit Health 6. https://doi.org/10.3389/fdgth.2024.1366967

Gebru T, Morgenstern J, Vecchione B, Vaughan JW, Wallach III HD, Crawford K (2021) Datasheets for datasets. https://arxiv.org/abs/1803.09010

Gentelet K, Mizrahi SK (2024) A human-centered approach to AI governance: operationalizing human rights through citizen participation. Human-centered AI. CRC Press, Boca Raton, FL, pp 215–230

Ghahramani M, Galle NJ, Ratti C, Pilla F (2021) Tales of a city: sentiment analysis of urban green space in Dublin. Cities 119:103395. https://doi.org/10.1016/j.cities.2021.103395

Goffredo P, Cabrio E, Villata S, Haddadan S, Torres Sanchez J (2023) DISPUTool 2.0: a modular architecture for multi-layer argumentative analysis of political debates. In Williams B, Chen Y, Neville J (eds) AAAI-23—37th AAAI conference on artificial intelligence, vol 37. AAAI, Washinghton, DC, United States, pp 16431–16433. https://doi.org/10.1609/aaai.v37i13.27069

Goodman B, Flaxman S (2017) European Union regulations on algorithmic decision-making and a "Right to Explanation." AI Mag 38(3):50–57. https://doi.org/10.1609/aimag.v38i3.2741

Greenwald E, Leitner M, Wang N (2021) Learning artificial intelligence: insights into how youth encounter and build understanding of AI concepts. In: Proceedings of the AAAI conference on artificial intelligence, vol 35, no 17, pp 15526–15533. https://doi.org/10.1609/aaai.v35i17.17828

Herrmann H (2023) What's next for responsible artificial intelligence: a way forward through responsible innovation. Heliyon 9(3). https://doi.org/10.1016/j.heliyon.2023.e14379

IxDF - Interaction Design Foundation (2016) What is cognitive friction? IxDF - Interaction Design Foundation. https://www.interaction-design.org/literature/topics/cognitive-friction

Kappenberger J, Theil K, Stuckenschmidt H (2022) Evaluating the impact of AI-based priced parking with social simulation. In: Hopfgartner F, Jaidka K, Mayr P, Jose J, Breitsohl J (eds) Social informatics. Springer International Publishing, Cham, pp 54–75. https://doi.org/10.1007/978-3-031-19097-1_4

Karacapilidis N, Tsakalidis D, Domalis G (2023) An AI-enhanced solution for large-scale deliberation mapping and explainable reasoning. In: Papadaki M, Rupino da Cunha P, Themistocleous M, Christodoulou K (eds) Information systems. Springer Nature Switzerland, Cham, pp 305–316. https://doi.org/10.1007/978-3-031-30694-5_23

Kocsis O, de Lera E, Bedek MA, Zangl M, Colt R, Šaušová V, Birasova L (2024) Facilitators and barriers of marginalized groups for the engagement in online civic activities. In: Kö A, Kotsis G, Tjoa AM, Khalil I (eds) Electronic government and the information systems perspective. Springer Nature Switzerland, Cham, pp 105–121. https://doi.org/10.1007/978-3-031-68211-7_9

Larsson S, Heintz F (2020) Transparency in artificial intelligence. Internet Policy Rev 9(2):1–16. https://doi.org/10.14763/2020.2.1469

Lepri B, Oliver N, Pentland A (2021) Ethical machines: the human-centric use of artificial intelligence. iScience 24(3). https://doi.org/10.1016/j.isci.2021.102249

Leslie D, Burr C, Aitken M, Cowls J, Briggs M, Lord Clement Jones T (2021) Artificial intelligence, human rights, democracy, and the rule of law: a primer. Zenodo. https://doi.org/10.5281/ZENODO.4639743

Liebenberg S (2012) Engaging the paradoxes of the universal and particular in human rights adjudication: the possibilities and pitfalls of 'meaningful engagement.' Afr Hum Rights Law J 12(1):1–29

Liu H, Zhou Q, Liang S (2025) Digital inclusion in public services for vulnerable groups: a systematic review for research themes and goal-action framework from the lens of public service ecosystem theory. Gov Inf Q 42(2):102019. https://doi.org/10.1016/j.giq.2025.102019

Manias G, Apostolopoulos D, Athanassopoulos S, Borotis S, Chatzimallis C, Chatzipantelis T et al (2023) AI4Gov: trusted AI for transparent public governance fostering democratic values. In: 2023 19th international conference on distributed computing in smart systems and the internet of things (DCOSS-IoT), pp 548–555. https://doi.org/10.1109/DCOSS-IoT58021.2023.00090

Mariani I, Anastasiou L, Elguendouze S, Rizzo F, Concilio G, Cabrio E, Villata S, Efstathiou I, Domalis G, De Liddo A (2026) Enhancing democratic deliberations with AI: insights from the ORBIS co-creation journey. In: Artificial Intelligence and Government: Examining the Roles and Uses of AI in Enhancing Government Operations, pp 221-251. Springer Nature Switzerland, Cham. https://doi.org/10.1007/978-3-032-12344-2_10

Manifesto A, Beck K (2001) Manifesto for agile software development. https://agilemanifesto.org/

Mbiti JS (1990) African religions and philosophy. Heinemann, Portsmouth, NH

Miller T (2019) Explanation in artificial intelligence: insights from the social sciences. Artif Intell 267:1–38. https://doi.org/10.1016/j.artint.2018.07.007

Mitchell M, Wu S, Zaldivar A, Barnes P, Vasserman L, Hutchinson B et al (2019) Model cards for model reporting. In: Proceedings of the conference on fairness, accountability, and transparency. Association for Computing Machinery, New York, NY, USA, pp 220–229. https://doi.org/10.1145/3287560.3287596

Mortier R, Haddadi H, Henderson T, McAuley D, Crowcroft J (2015) Human-data interaction: the human face of the data-driven society. https://doi.org/10.48550/arXiv.1412.6159

Mouffe C (1999) Deliberative democracy or agonistic pluralism? Soc Res 66(3):745–758. https://www.jstor.org/stable/40971349

Nissenbaum H (2004) Privacy as contextual integrity. Wash Law Rev 79:119. https://digitalcommons.law.uw.edu/wlr/vol79/iss1/10

Offe C (2012) Whose good is the common good? Philos Soc Crit 38(7):665–684. https://doi.org/10.1177/0191453712447770

Ohren A, Calderón Lüning E, Markov Č, Stearns M (2025) From efficiency to deliberation: rethinking AI's role in institutionalizing democratic innovations. In: Politics and governance; vol 14 (2026): towards an innovative democracy: institutionalizing participation in challenging times. https://doi.org/10.17645/pag.10632

Ozkaya I, Nord R, Kruchten P (2019) Exploring the technical debt landscape. Carnagie-Mellon University, Pittsburg, PA. https://apps.dtic.mil/sti/html/tr/AD1088675/

Pérez-Escolar M, Canet F (2023) Research on vulnerable people and digital inclusion: toward a consolidated taxonomical framework. Univ Access Inf Soc 22(3):1059–1072. https://doi.org/10.1007/s10209-022-00867-x

Pettit P (2004) The common good. In: Justice and democracy, vol 9780521836951. Cambridge University Press, United Kingdom, pp 150–169. https://doi.org/10.1017/CBO9780511490217.009

Pietilä I, Meriläinen N, Varsaluoma J, Väänänen K (2021) Understanding youths' needs for digital societal participation: towards an inclusive virtual council. Behav Inf Technol 40(5):483–496. https://doi.org/10.1080/0144929X.2021.1912182

Raisio H, Valkama K, Peltola E (2014) Disability and deliberative democracy: towards involving the whole human spectrum in public deliberation. Scand J Disabil Res. https://doi.org/10.1080/15017419.2013.781957

Raji ID, Smart A, White RN, Mitchell M, Gebru T, Hutchinson B et al (2020) Closing the AI accountability gap: defining an end-to-end framework for internal algorithmic auditing. In: Proceedings of the 2020 conference on fairness, accountability, and transparency. Association for Computing Machinery, New York, NY, USA, pp 33–44. https://doi.org/10.1145/3351095.3372873

Rasmussen J (1985) The role of hierarchical knowledge representation in decision making and system management. IEEE Trans Syst, Man, Cybern SMC 15(2):234–243. https://doi.org/10.1109/TSMC.1985.6313353

Raso FA, Hilligoss H, Krishnamurthy V, Bavitz C, Kim L (2018) Artificial intelligence and human rights: opportunities and risks. Berkman Klein Center Research Publication (2018–6). https://doi.org/10.2139/ssrn.3259344

Régis C, Denis J-L, Axente ML, Kishimoto A (2024) Human-centered AI: a multidisciplinary perspective for policy-makers, auditors, and users. Chapman and Hall/CRC, Boca Raton, FL. https://doi.org/10.1201/9781003320791

Reis J, Santo PE, Melão N (2019) Artificial intelligence in government services: a systematic literature review. Adv Intell Syst Comput 930:241–252. https://doi.org/10.1007/978-3-030-16181-1_23

Rosenberg D (2024) Older adults' e-government use for bureaucratic and transactional purposes: the role of website-related perceptions and subjective digital skills. Transform Gov: People, Process Policy 18(2):257–269. https://doi.org/10.1108/TG-08-2023-0116

Rudin C (2019) Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. https://arxiv.org/abs/1811.10154

Sachdeva N, Tuikka A-M, Kimppa KK, Suomi R (2015) Digital disability divide in information society. J Inf Commun Ethics Soc 13(3/4):283–298. https://doi.org/10.1108/JICES-10-2014-0050

Savaget P, Chiarini T, Evans S (2019) Empowering political participation through artificial intelligence. Sci Public Policy 46(3):369–380. https://doi.org/10.1093/scipol/scy064

Schütz F, Heidingsfelder ML, Schraudner M (2019) Co-shaping the future in quadruple helix innovation systems: uncovering public preferences toward participatory research and innovation. She Ji: J Des, Econ, Innov 5(2):128–146. https://doi.org/10.1016/j.sheji.2019.04.002

Shneiderman B (2022) Human-centered AI. Oxford University Press

Siachos I, Karacapilidis N (2024) Explainable artificial intelligence methods to enhance transparency and trust in digital deliberation settings. Futur Internet 16(7). https://doi.org/10.3390/fi16070241

Sieber R, Brandusescu A, Sangiambut S, Adu-Daako A (2024) What is civic participation in artificial intelligence? Environ Plan B: Urban Anal City Sci 23998083241296200. https://doi.org/10.1177/23998083241296200

Simonofski A, Asensio ES, De Smedt J, Snoeck M (2017) Citizen participation in smart cities: evaluation framework proposal. In: 2017 IEEE 19th conference on business informatics (CBI), vol 01, pp 227–236. https://doi.org/10.1109/CBI.2017.21

Skaržauskienė A, Mačiulienė M (2020) Mapping international civic technologies platforms. Informatics 7(4). https://doi.org/10.3390/informatics7040046

Skorge ØS (2023) Mobilizing the underrepresented: electoral systems and gender inequality in political participation. Am J Polit Sci 67(3):538–552. https://doi.org/10.1111/ajps.12654

Stilgoe J, Owen R, Macnaghten P (2013) Developing a framework for responsible innovation. Res Policy 42(9):1568–1580. https://doi.org/10.1016/j.respol.2013.05.008

Straub VJ, Morgan D, Bright J, Margetts H (2023) Artificial intelligence in government: concepts, standards, and a unified framework. Gov Inf Q 40(4):101881. https://doi.org/10.1016/j.giq.2023.101881

Sunstein CR (2007) Republic.com 2.0. Princeton University Press. http://www.jstor.org/stable/j.ctt7tbsw. Accessed 5 Jan 2026

Teller M (2022) Vers l'acceptabilité sociale des algorithmes ou comment passer de l' "algocratie" à la démocratie algorithmique. *Revue pratique de la prospective et de l'innovation* (no 1). https://hal.science/hal-03671112

Tully SM, Longoni C, Appel G (2025) Lower artificial intelligence literacy predicts greater AI receptivity. J Mark 89(5):1–20. https://doi.org/10.1177/00222429251314491

UNESCO (2022) Recommendation on the ethics of artificial intelligence. United Nations Educational, Scientific and Cultural Organization. https://unesdoc.unesco.org/ark:/48223/pf0000381137

UNICEF (2021) Policy guidance on AI for children. Office of Global Insight and Policy. https://www.unicef.org/innocenti/media/1341/file/UNICEF-Global-Insight-policy-guidance-AI-children-2.0-2021.pdf

Vallor S (2016) Technology and the virtues: a philosophical guide to a future worth wanting. Oxford University Press

Vedung E (2020) Policy evaluation. In: The SAGE handbook of political science. SAGE, London UK, p 68. https://doi.org/10.4135/9781529714333.n68

Vicente KJ (1999) Cognitive work analysis: toward safe, productive, and healthy computer-based work. CRC Press, Boca Raton, FL. https://doi.org/10.1201/b12457

Voegtlin C, Scherer AG (2017) Responsible innovation and the innovation of responsibility: governing sustainable development in a globalized world. J Bus Ethics 143(2):227–243. https://doi.org/10.1007/s10551-015-2769-z

von Eschenbach WJ (2021) Transparency and the black box problem: why we do not trust AI. Philos Technol 34(4):1607–1622. https://doi.org/10.1007/s13347-021-00477-0

Wan S, Sieber R (2023) Artificial intelligence (AI) adoption in Canadian municipalities: in-house development versus outsourcing. In: The 18th international conference on computational urban planning and urban management. Presented at the 18th international conference on computational urban planning and urban management—CUPUM 2023. https://doi.org/10.17605/OSF.IO/6YR5V

Wilson C (2022) Public engagement and AI: a values analysis of national strategies. Gov Inf Q 39(1):101652. https://doi.org/10.1016/j.giq.2021.101652

Wollmann H (2017) Policy evaluation and evaluation research. Handbook of public policy analysis: theory, politics, and methods. Routledge, New York, NY, pp 393–404. https://doi.org/10.4324/9781315093192

Yenduri G, Kaluri R, Rajput DS, Lakshmanna K, Gadekallu TR, Mahmud M, Brown DJ (2023) From assistive technologies to metaverse—technologies in inclusive higher education for students with specific learning difficulties: a review. IEEE Access 11:64907–64927. https://doi.org/10.1109/ACCESS.2023.3289496

Zhang A, Walker O, Nguyen K, Dai J, Chen A, Lee MK (2023) Deliberating with AI: improving decision-making for the future through participatory AI design and stakeholder deliberation. In: Proceedings of the ACM human-computer interaction, vol 7(CSCW1). https://doi.org/10.1145/3579601

Zimmermann A, Di Rosa E, Kim H (2020) Technology can't fix algorithmic injustice. Boston Rev 9. https://www.bostonreview.net/articles/annette-zimmermann-algorithmic-political/

Zuiderwijk A, Chen Y-C, Salem F (2021) Implications of the use of artificial intelligence in public governance: a systematic literature review and a research agenda. Gov Inf Q 38(3):101577. https://doi.org/10.1016/j.giq.2021.101577