From inclusion to illusion: the pitfalls of ethicswashing in Participatory AI practices

Riccardo Corsi[1,2,*,†], Beatrice Melis[1,2,†], Gianluca De Ninno[1,2,†] and Tiziana Nupieri[3,†]

[1]Computer Science Area, Gran Sasso Science Institute: L'Aquila, Italy
[2]Department of Computer Science, University of Pisa: Pisa, Italy
[3]Department of Political Sciences, Sapienza University of Rome: Rome, Italy

Abstract

This article addresses the increasing use of participatory approaches in the development of artificial intelligence (AI), examining their democratic potential and associated risks. While participatory AI promises to democratize technological design and enhance inclusive governance, it often falls short due to power imbalances, vague definitions, and superficial implementation — leading to what as been defined as participatory washing [1, 2] . Drawing from studies on participation in political science, public policy, and technology, the paper proposes a conceptual framework to critically assess participatory practices along four dimensions: power, goals, actors, and arenas. The goal is to support practitioners in avoiding instrumental uses of participation and to foster genuinely empowering and accountable AI development processes.

Keywords

Participatory AI, AI democratization, participatory washing, ethics washing, public policy

1. Introduction

The rapid advancement of artificial intelligence (AI) has presented new societal opportunities and challenges. Participatory approaches have emerged to align AI with ethical values and societal needs. The need for increased inclusiveness in relation in AI development process stems from the significant impact that AI-based technologies have in numerous fields, even critical ones, such as healthcare, safety, education and employment in the labor market. Alongside, as highlighted by the European Commission itself, there is a growing demand of participation from non-governmental and non-technical societal actors, with the goal of creating effective and socially responsible systems [3].

One of the major strengths of the call for participatory AI relies on its double promise to democratize the design and development of AI-driven technologies and services, while being integrated as well in democratic processes which involve participatory practices. However, despite these claimed advantages, participatory AI and connected approaches are not free of risks, and the definitional vagueness makes it prone to misunderstanding and misuse, the more significant one being defined as the phenomenon of *participatory-washing*, where the effort to include and collaborate with externals results more in a virtuous simulation than an actual power-sharing and genuine empowerment.

In particular, this work aims to: (1) contribute to a theoretical understanding of participatory AI implications, especially concerning power asymmetries and related issues; and (2) support practitioners and developers seeking to adopt participatory approaches by helping them avoid instrumental or superficial uses of participation. In doing so, we refer to the tradition of the analysis of public policy in political science and political sociology, with a specific focus on participation e.g., [4, 5] and related critical studies [6, 7, 8, 9], as well as existing literature on participatory AI more technology centered, e.g., [10, 11, 12]. By integrating these contributions, this article offers an initial development of a

HHAI-WS 2025: Workshops at the Fourth International Conference on Hybrid Human-Artificial Intelligence (HHAI), June 9-13, 2025, Pisa, Italy

*Corresponding author.
†These authors contributed equally.
riccardo.corsi@gssi.it (Riccardo Corsi); beatrice.melis@gssi.it (Beatrice Melis); gianluca.deninno@gssi.it (Gianluca De Ninno); tiziana.nupieri@uniroma1.it (Tiziana Nupieri)

theoretical framework for understanding AI-related participatory practices. Rather than focusing on empirical case studies, it adopts a more conceptual perspective aimed at clarifying the phenomenon itself and offering insights into how to avoid participatory washing [1, 2].

2. The call for participation: from public policy to artificial intelligence

2.1. Participation in public policy: does participating make me more powerful?

Participation is a key concept in social sciences and socio-political research addressing democracy and the relationship between citizens and institutions [13, 14]. Academic discourse has long examined the reasons behind the diffusion of participatory forms [15, 16], their diversity [4], as well as the potential and risks associated with their use and role [17]. It is far from a univocal notion, as it encompasses multiple dimensions. As noted, participation must be problematized and historicized to be fully understood [18]. On one hand, it refers to the capacity of individuals to shape political decision-making actively and influence the quality and outcomes of public policies. In this sense, Segatori [19] defines political participation as a set of attitudes and behaviors through which individuals seek to steer political decisions. On the other hand, participation also serves as a normative foundation for democratic systems, strengthening the legitimacy of both inputs and internal processes of policy-making [20, 21, 22]. Since the 1990s, Western democracies experienced a "deliberative turn" [20], fostering participatory mechanisms supplementing traditional representative channels and promoting direct interactions in policy-making between institutions and civil society beyond conventional interest mediation.

The inclusion of civil society is seen as enhancing democracy through principles like inclusiveness, transparency, and accountability [23, 24]. In this context, participatory governance is viewed as a form of deliberative democracy, aimed at expanding the public sphere and making decision-making processes more open and accountable, as the publication of the European Governance White Paper in the EU formally recognized [25]. At the same time, "new" participatory forms have emerged, following an isomorphic trend and spreading across various sectors — from social policies and welfare to the environment, urban planning, such as Participatory Budgeting, Citizen Juries, and Public Debates[18]. These top-down participatory processes, led by institutional actors in local contexts, differ from the participation processes of the 1960s and 1970s in that they are barely characterized by ideological aims and are more focused on problem-solving for specific policy issues [8, 9, 18].

This shift has been driven by assumptions that participation effectively addresses the challenges of consensus-building, social cohesion, and collaboration [26]. These practices rely on direct engagement between institutions and citizens, beyond traditional channels of interest representation, ideally expanding the access to decision-making [27]. Such informed and reasoned dialogue is expected to enhance the efficiency, legitimacy, and cooperation in public decision-making processes, particularly at the local level [8]. This idea has also been fueled by a normative convergence in both academic literature and among governance actors, where participation is often viewed as inherently positive and capable of producing beneficial effects on decision-making processes and the quality of democracy [8].

But is it always the case? The simple fact of participating in a process will lead to effective empowerment for all of us? Several criticisms have been pointed out concerning participatory practices and they are worth noticing before focusing on the technological domain. First of all, in most cases, these mechanisms are top-down [28], which means that stakeholders are involved mainly for providing feedback and have zero or little influence on the actual content of decisions [8, 6]. Then, it has been pointed out that there is considerable uncertainty about the representative capacity of participatory practices, particularly regarding how well they reflect diverse interests within society, with legitimacy issues and a lack of accountability of the actors involved, if compared with the elected [29, 21]. Another relevant issue is related to the lack of social learning within participatory practices, which strive to affect policy paradigms or guiding ideas in the policy-making [30, 31]. Furthermore, power asymmetry among stakeholders, stemming from unequal resources, often leads to uneven influence in participatory

processes [32, 33, 34]. Moreover, these processes frequently prioritize corporate objectives, aligning interests toward business rather than societal needs [35].

As recognized by Edmundus and Wollenberg [36] these unequal capacities to influence decisions perpetuate inequalities in the policymaking process. Ultimately, participatory practices risk to be aimed at avoiding conflicts and fostering an appeasement that advantages those in power [7, 37], producing a depoliticization of collective interests' issues [9], thus resulting in a democratic illusion [38].

2.2. From e-democracy to democratizing AI: the double promise of Participatory AI

Contemporary governance innovations, such as deliberative assemblies and citizen panels, have sought to shift towards more horizontal structures, where citizens contribute substantive input to policy formation to advocate for openness in decision-making processes [22, 39]. It has been argued that the principles underlying participatory AI closely mirror those pertaining to democratic governance [1], particularly in participatory decision-making processes within political systems [40]. Just as co-design and participatory governance mechanisms in politics aim to involve citizens in public decision-making, participatory AI seeks to *democratize* technological development by incorporating, e.g., user inputs in technological development [41]. However, participatory AI encompasses much more than this.

A key dimension of the relationship between participation and AI lies in the emergence of new forms of engagement mediated by digital platforms and networking tools, sparking discussions regarding the functioning and application of e-participation or digital participation [42]. Digital plataforms, virtual networks and social media are increasingly recognized as facilitating civic participation by fostering direct communication between citizens and administrations, thus enhancing democratic interactions [43]. AI systems have the transformative potential to stimulate democratic participation by significantly reducing informational and cognitive burdens on citizens, thereby fostering more informed and reflective political engagement [44, 45]. AI systems can influence democratic processes through the aid in electoral processes, public consultation, mass online deliberation, or participatory decision-making process [46], facilitating also agenda-setting, opinion synthesis, and consensus-building, thereby enhancing transparency and responsiveness of democratic institutions [47]. Through adversarial machine learning, AI can protect citizens' privacy from invasive political profiling, enabling individuals to deliberate freely and independently, preserving the authenticity of their political choices [48]. Nevertheless, the deployment of AI in democratic processes poses serious risks, including the potential of reinforcing biases and power asymmetries inherent in algorithmic governance [49], threats to freedom of speech and media pluralism, access to public information, truth, and essential services [50, 51, 52]. Recent AI developments reflect a "participatory turn" [10], shifting from top-down technical designs toward inclusive practices actively involving diverse stakeholders — developers, end-users, domain experts, and marginalized communities [53].

Such participatory methods aim to achieve continuous alignment between technological solutions and the diverse needs, values, and preferences of those whom these technologies are designed to support [10]. This recalls how participatory design was initially applied in technological development in the 90s, with the recognition that every societal actor can be a legitimate source of knowledge production [54]. The need of including different stakeholders is recognized as a way to mitigate risks and harms, as argued by the Global Partnership on AI: "products that are not created with an inclusive approach do not serve all users of technology equally and in some cases they can actively harm communities, especially those who have been historically excluded or marginalized" [55]. This aspect has gained attention and is starting to be considered as a driver of technological innovation [56]. Engaging stakeholders throughout the entire development lifecycle can ensure that systems reflect the lived experiences and expectations of those who interact with them. The engagement of citizens and participatory impact assessments are recognized as viable way to prevent risks in our datafied societies and enhance the self-determination in the digital realm [57]. This is because the call for these participatory practices is rooted in the idea of defending the values of democracy, human rights, and autonomy from the risks of being eroded by the concentration of power and unaccountable AI development [58]. The founding goal of this approach is embodied in the pursuit of *democratizing* AI development by ensuring that

those affected by their adoption have a stake in their shaping, even if they are non-experts [59].

However, participation is not a one-size-fits-all solution and it could assume several different meanings and describe totally different practices. As the contribution by the Ada Lovelace Institute [60] highlighted, participatory mechanisms can be integrated according to five different levels, from inform, consult, involve, collaborate, to empower - shaping a spectrum of involvement practices ranging from engagement to deep, systemic co-creation. Other studies, e.g. the one in [61] conceptualize four levels of participatory AI: consultation, contribution, collaboration, and co-creation. Sloane et al.[2] elucidate an epistemological distinction of the concept, distinguishing between participation as a work, as a consultation, and as social justice. The work by Birhane et al. [1] focuses on the aims, identifying three objectives of participatory AI: for algorithmic performance improvements, process improvements, and collective exploration, which should be more centered on stakeholders' needs. The existing body of work provides valuable frameworks for understanding the diverse range of participatory practices in AI development, but there is still difficulty in operationalizing these frameworks effectively in practice [10, 62, 63, 64].

3. Challenges and pitfalls of participatory practices and the spook of ethicswashing

3.1. Challenges and pitfalls of participation practices AI related

Much like in public policy, the state of participatory approaches to AI development is largely consultative in nature, which means that while stakeholders provide input on AI system modules, they are often not integrated as active decision-makers in the broader lifecycle of AI development [10]. Using Arnstein's"Ladder of Citizen Participation" [65], Corbett et al. [41] note participatory AI mostly remains at consultative levels, often merely legitimizing decisions already made.

Additionally, there is a real risk of reifying or amplifying existing power dynamics through participatory AI processes. As the work by Delgado et al. [10] highlights, while participatory mechanisms in AI aim to include marginalized voices, there is always the danger that these efforts reinforce rather than challenge the *status quo*, especially when the underlying power structures are not addressed early in the design process.

Birhane et al. [1] address another key challenge in the inclusion debate. They emphasize that while being included in participatory processes might have practical consequences on engagement, inclusion alone is not equivalent to participation. As they note, an individual can be part of a group but still not actively participate, for example by not voting, writing, or acting [1]. This distinction between inclusion and true participation is crucial when discussing participatory AI, as it underlines the difference between being invited to the table and having actual influence over decisions.

Main challenges in achieving meaningful participation in AI arise from the inherent complexity of data processing and the technical issues of AI systems. The massive amounts of data and computational demands involved in developing AI systems, particularly Large Language Models (LLMs), raise concerns regarding the feasibility of meaningful human participation at the scales at which these systems operate. As Ananny and Crawford[66] discuss, the ideal of transparency applied to algorithmic accountability may not be realistic due to the inherent complexities and black-box nature of AI systems. These challenges complicate the notion of inclusion, as they directly impact the interpretability and comprehensibility of AI development process and implementation.

Moreover, scaling of AI systems poses a significant issue for defining useful forms of participation. As noted, meaningful human participation may not even be feasible on the large scales at which many AI systems operate, whether in terms of data collection across vast geographies or system deployment in diverse contexts [66]. The operation of AI at these scales often results in "fossilized preference models", which may create a substantive gap between the preferences of human agents and algorithmic decision-making, thus functioning as a technology of depoliticization [10]. Proxies, such as aggregated data or simulations, are increasingly used in participatory AI contexts but carry significant risks. They

can inadequately represent marginalized voices, exacerbate power imbalances, and reduce the agency of affected communities, thereby distorting democratic processes and sidelining authentic stakeholder engagement [1, 10].

Moreover, focusing on a social side effect of machine learning approaches which are defined as participatory, Sloane et al. [2] coined the term "participatory washing" by referring to an extractive and exploitative approaches of community involvement - that is specifically called ethicswashing when it addresses ethical concerns.

3.2. Participation as a branding device for digital products?

In fact, amid growing recognition of the widespread influence of socio-technical systems and the mounting risks they entail, companies and developers have increasingly turned to ethics-based initiatives as a means of signaling trustworthiness and strengthening brand legitimacy [67]. The notion of *ethicswashing* — adapted from the environmental domain where it denoted superficial displays of sustainability [68] — has become a critical lens through which these practices are analyzed. It describes the strategic invocation of ethical discourse by private or public actors in ways that are largely performative or semantic. A salient example is the intervention of Thomas Metzinger, a philosopher and former member of the EU's AI Ethics Guidelines drafting panel, who condemned the process of as "ethicswashing" and referred to the industry's approach as an "ethicswashing machine" [69].

This form of merely rhetorical ethics is best understood as a communicative strategy — intended not to reform practices, but to shape perception. As Freiman argues, ethicswashing functions as a deliberate branding technique aimed at cultivating consumer confidence through the illusion of ethical integrity [70]. Floridi terms this phenomenon *ethics bluewashing*, echoing its environmental counterpart - the *green* - and defines it as the "malpractice of making unsubstantiated or misleading claims about, or implementing superficial measures in favor of, the ethical values and benefits of digital processes, products, services, or other solutions in order to appear more digitally ethical than one is" [71].

Further challenges in operationalizing the ethics of socio-technical and AI systems contribute to what is now recognized as *digital ethicswashing* [72]. These include the unchecked proliferation of ethical codes and guidelines, their opportunistic use to delay or deter legislation, and the outsourcing of ethically dubious research to jurisdictions with weaker oversight [71]. Another form of digital ethicswashing is known as *ethics-bashing*: this refers to the dismissal or trivialization of ethical discourse, reducing it to modular tools — ethics boards, self-governance frameworks, or stakeholder assemblies —stripped of their normative force and treated instead as bureaucratic formalities [73]. At the user level, ethicswashing erodes trust when a dissonance emerges between corporate rhetoric and practice — the proverbial gap between the "talk" and the "walk" [74]. Importantly, ethicswashing is not confined to the private sector. It pervades other institutional domains, including academia, the public sector, and policymaking [72]. Viewing the growing interest in inclusive and participatory methods of designing AI systems [56], it could be used to gain a more "ethical" appeal, if we consider also the limits of current participatory practices, it is well sounding to claim that participation in the digital realm could be a form of digital ethicswashing.

4. Assessing participation: Power, Goals, Actors, Arenas and Formats as analytical dimensions

In this section, we turn to propose an initial theoretical framework which aim to help the conceptual understanding of participatory AI implications, and support practitioners to avoid instrumental uses of it. We do so by drawing on theories from public policy [4, 17, 65, 75, 76, 77, 78, 79], and AI-related domain [10, 41, 62].

For this contribution, the idea is to select a series of elements that have raised reflections on our part about perspectives and attentions to be added to the existing analyzed landscape. According to the aim of the present work and based on the works discussed along the paper, we propose 4 critical analytical

dimensions of participation which must be encompassed in order to avoid participatory washing. These are the following: *power*, *goals*, *actors*, *arenas*. Two fundamental considerations have guided our decision to focus on these four specific dimensions of participation as central to our framework. First, AI development is inherently political [80, 81], making it crucial to critically examine power relations embedded in participatory processes, including who decides, how influence is distributed, and whose interests are ultimately served. Second, while many of the frameworks discussed in the literature provide valuable tools for understanding participation, they often adress certain aspects — especially those related to power, actors' agency, and the deeper goals of participation — in a limited or overly formalized way [1]. These elements are either insufficiently explored or treated in ways that do not fully acknowledge their broader societal implications. For instance, there is a tendency to treat participation as a procedural matter, rather than recognizing the socio-political dynamics at play, such as the risk of reinforcing existing inequalities or enabling washing practices by dominant actors, including tech companies - instead of working for the social good [81, 82].

For this reason, we propose these four dimensions which are still under-theorized in relation to their political substance and practical consequences, especially within the AI domain, as represented in Figure 1. Additionally, Table1 provides a concise synthesis of these dimensions, offering an overview of critical reflections and key analytical questions practitioners should consider to ensure genuine participatory practices.

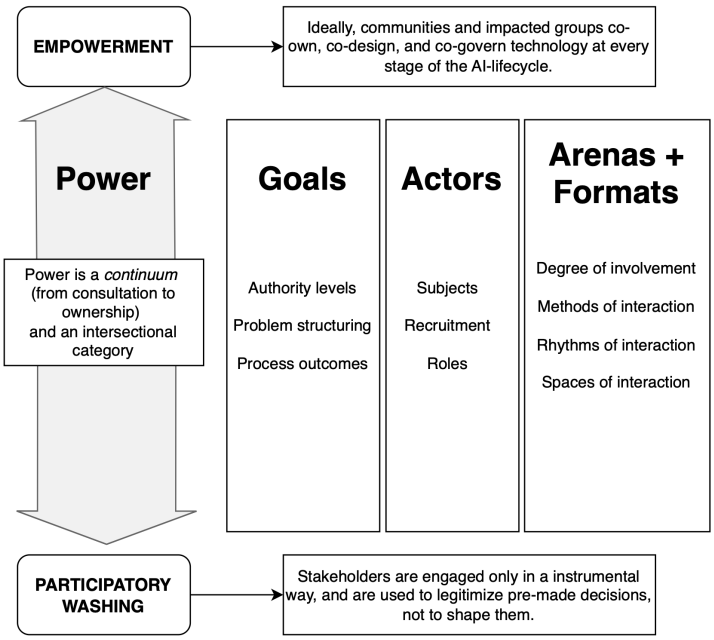

Figure 1. Visual representation of the framework proposed to assess participation in AI-related practices.

Power. A crucial aspect concerns power and has extensively being studied in relation to participation, e.g., [65, 4]. Delgado et al. [10] discuss power as a scale that ranges from *consult* (minimun level) to *own* (full participation form), distinguishing between four varying degrees of decision-making authority investing stakeholders. Participation ranges from consultative feedback without decision authority, to ownership where stakeholders actively shape the entire AI lifecycle, a level never observed in current practices [10]. A crucial example of limited power redistribution can be found in a project involving predictive AI for child welfare risk assessment [83]. In this case, community members such as caregivers and child protection workers were invited to share their perspectives, but were ultimately excluded from any decision-making about the design or deployment of the AI model. The assumption that AI would be implemented was never questioned, and participation remained restricted to consultation on pre-defined aspects. This underscores how stakeholders, although present, may remain marginal to critical decisions when the power structure is not fundamentally rebalanced [10]. Besides how many levels of wield power and authority one should claim, in our perspective, this power dimension must be

intended as a *continuum*, and as a relational issue which will be reflected also in other dimensions, from the selection of stakeholders [84], to the roles of interaction, recruitment techniques, possibilities of social learning - for example redefining problems besides mere solutions [4, 75].

Goals. Participation has always a goal, there is always a *why* to be answered. It is possible to conceptualize this by distinguishing the *level of authority* exercised by the participants, ranging from the improvement of user experience, to alignment with values or preferences of people engaged, to deliberation on specific features, and finally to engagement across the entire life cycle[10]. The study in [75] allows the conceptualization of two further elements to define the aim of a participatory process. Drawing on the classical study in [76], they argue that determining the appropriate level of participation depends also on *how structured the problem at hand is* — whether it is fully structured, moderately structured, or unstructured. Clearly, a more open problem to be solved could lead to social learning and much more power redistribution among stakeholders, if compared to e.g., a simple request of expressing a preference among two options. Furthermore, the outcomes of the participatory process contribute to define its purpose [75]. Drawing on the study by [78], it becomes clear that analyzing both the immediate results and the longer-term substantive outputs and *outcomes* of the participatory process is essential to understanding how participation is intended to be enacted. A relevant example is the WeBuildAI project [85], in which researchers engaged Pittsburgh residents in the co-construction of preferences to be used in a public resource allocation algorithm. While the project was supposed to go beyond an instrumental goal of performance optimization by encouraging reflection on civic values and collective priorities, the scope of citizen influence remained limited. Participants were constrained to selecting among predefined options, without the opportunity to shape the design of the algorithm itself. As Delgado et al. observe, even advanced participatory frameworks like this often remain anchored in preference elicitation rather than effective co-decision-making - even if it can be recognized as a valuable starting point in the involvement of impacted stakeholders to express preferences towards the goal [10]. Participation, as pointed out and recognized by many, is also an ”end in itself” [77, 10, 65], and technology could even allow new forms of participation [46].

Actors. As the work by Birhane et al. [1] points out, inclusion does not necessarily imply participation, as individuals may be included in a process but lack the agency to make meaningful contributions. In the study in [75] the authors further divide this concept into three aspects, namely subjects, roles and recruitment. Firstly, *subjects* are classified into civil society, government, and business sector, with varying degrees of institutionalization. Civil society actors often represent marginalized or disadvantaged communities, while government and business actors are typically more institutionalized and have greater influence over the process - often representing a reflection of existing power imbalances. This is why the selection phase is a critical component of the participatory process, and it is not just about *who* is included, but how those actors are empowered within the process (*roles*). About the *recruitment* issue, Kallina and Singh [62] discuss how stakeholder recruitment strategies must be inclusive, ensuring that impacted groups especially those who are harder to reach — are not excluded. The role of knowledge is also crucial, as the level of expertise and experience directly influences the extent to which stakeholders can shape the outcome of AI projects. The work in [86] further clarifies that stakeholders selection often occurs based on the knowledge and experience they possess — that is, whether they have more or less specialized cognitive resources in the domain of interest. A case that exemplifies the risks of tokenistic recruitment involves the participatory process in the educational research context of human-robots interaction patterns for children, as highlighted by Delgado et al. [10]. In this project, only educators were consulted, with the assumption that they could effectively represent children’s needs through role-playing. This kind of proxy-based participation raises concerns about extractivism [63], where the knowledge and lived experience of the target group are interpreted and filtered by institutional actors, potentially misrepresenting their interests [10]. A way to prevent misrepresentation is to promote digital literacy [87], with attention also to the *rhythm* of the participatory practices.

Arenas and formats. Delgado et al. [10] distinguishes participatory approaches based on the *degree of involvement*: some processes allow stakeholders to provide input and feedback, others facilitate group discussions, while still others go even further, involving stakeholders in decision-making or

even allowing them to reflexively choose how to participate. Participation formats further shape these dynamics by defining the methods and modes of interaction. As Fung [4] and Gaventa[79] explain, formats include traditional *methods* such as public hearings, deliberative forums, participatory budgeting and online engagement. Accordingly, the work in [75] distinguishes between invited, created, or closed spaces, based on their level of institutionality and the degree of access provided to stakeholders within the arena. These types of arenas shape how collaborative and participative spaces are both socially and physically constructed. The dynamics of participation within these spaces — that is, the interpersonal interactions and, especially, the social performance of participants, are as well elements worth considering. In this regard, the roles of mediators within participatory processes, an element which is often underdeveloped in current frameworks, could be further investigated. As highlighted by Delgado et al.[10], this level often involves professionals with expertise in user experience (UX) and human-computer interaction (HCI), who act as mediators between the stakeholders and designers. Their role is to ensure that input is translated into practical outcomes, bridging the gap between technical knowledge and stakeholder needs — which is more a mediation towards the aim. However, we think that another level of mediation should be addressed and introduced, namely a specific mediation for the process, to ensure that all participants' perspectives are equally valued, especially in complex participatory processes like AI design. In our view, this level of mediation requires professionals with expertise in sociology, psychology, anthropology, conflict resolution, and facilitation, who are trained to manage the complex social dynamics that arise in participatory settings. An example of a more inclusive and effective participatory arena is found in a project co-designing social robots with a small community of adults with depression [88]. The process was structured across multiple phases: initial interviews to gather narratives, collaborative workshops for idea generation, and final sessions for validation. While ultimate control remained with the designers, the iterative nature and openness of the format allowed stakeholders to deeply engage and refine their input over time. This example demonstrates how well-designed formats can promote social learning and build proper legitimacy. Accordingly, another crucial aspects defining arenas of interaction is time. As the study in [75] recognizes, participation practices has *rhythms* - whether one-off events, longer-term engagements, or ongoing processes — which influence the depth and quality of involvement. In AI projects, this issue is particularly critical, as short-term consultations may not provide stakeholders with enough time to understand the complexity of AI systems and influence their design. In our view, the long-term involvement of stakeholders not only deepens their understanding of the technology, but also enables them to observe and react to its integration into real-world contexts.

Table 1 aims to provide practitioners and researchers with concrete analytical criteria to critically assess and structure participatory AI processes, explicitly addressing common vulnerabilities linked to participatory washing. For instance, the Power dimension urges clarification about stakeholders' actual influence on decision-making across the AI lifecycle, emphasizing transparency and genuine power redistribution. The Goals dimension encourages explicitly defining participation objectives, cautioning against vaguely stated purposes and superficial engagement. It underscores how different problem structures, ranging from highly structured to open-ended issues, directly influence the nature of participation required. Concerning Actors, the table highlights the importance of thoughtful recruitment and representation, addressing the pitfalls of tokenism and proxy participation, thus ensuring participants' genuine empowerment and contribution. Finally, the Arenas and Formats dimension stresses the necessity of carefully designed participatory environments, recognizing how spatial, temporal, and methodological configurations profoundly impact the legitimacy, depth, and effectiveness of stakeholder engagement.

Overall, the table operationalizes theoretical insights into practical guidance, promoting deeper, reflective, and genuinely democratic participatory practices in AI development. It emphasizes that participatory quality hinges not only on who is included, but on how power is distributed, what goals are set, how actors are recruited and supported, and in which settings and formats deliberation unfolds. Participation in AI should be an ongoing process of acceptance and feedback, that cannot be captured in short-term consultations, as stakeholders' perspectives on AI may shift as they witness its effects, both positive and negative, over time. As acknowledged in the broader discourse, achieving genuine

Dimension	Critical Reflections	Key Analytical Questions
Power	Clarify how and where participants influence decisions across the AI lifecycle (design, deployment, feedback); Risk of limited stakeholder agency; Need for genuine redistribution of power and transparency about intent	Who holds decision-making authority? How is power distributed and exercised? What power dynamics are challenged through the participatory process? To what extent can stakeholders shape outcomes across the lifecycle of AI systems?
Goals	Clearly articulate the objectives of participation and link them to the type of problem addressed (structured vs. open-ended); Avoid vague or general goals (e.g., "improving inclusivity") without structure; Acknowledge that participation is not inherently positive and must be critically assessed	What are explicit and implicit goals of participation? Are participants co-defining objectives or selecting predefined options? How does the structure of the problem influence the type of participation required?
Actors	Clarify who is involved and the rationale behind recruitment (representation, diversity); Avoid assuming the views of a few reflect all stakeholders; Ensure participants have capacity and support to engage (AI literacy, social performance); Include social mediators (not just technical ones) to foster equity and manage interactional dynamics	Who is included/excluded? What roles do actors assume? How are participants recruited? Are mediators included to balance participation and guide process dynamics?
Arenas and Formats	Describe arenas in practical terms (physical/digital, open/invited); Consider participation rhythm: one-off vs. long-term engagment; Ensure clarity on what participants can expect to do and influence; Recognize that effectiveness depends on continuous engagement	In what spaces does participation occur (invited, created, closed)? What interaction formats are employed? Are arenas anchored in realistic and understandable terms for participants? How do time and rhythms shape participation depth and legitimacy?

Table 1
Framework to avoid participatory washing

empowerment in the context of AI development is a significant challenge. One of the primary barriers to this is the knowledge gap. This is why a fundamental requirement for empowerment is AI literacy [87], which facilitates informed decision-making and ensures the integration of diverse perspectives throughout the entire design process.

5. Conclusions and future work

Participatory approaches in AI offer a double promise: democratizing both technology development and decision-making. However, without clear frameworks and genuine power-sharing, such practices risk becoming symbolic — falling into a specific form of digital ethicswashing: participatory washing. To avoid this, drawing in the public policies study on participation and existing literature on participatory processes in AI development, we proposed to conceptually articulate participation in four key dimensions (power, goals, actors, and arenas) highlighting salient aspect in each dimension to allow practitioners and developers to avoid ethicswashing. This structured framework allows a more rigorous and politically-aware assessment of participatory practices, explicitly targeting vulnerabilities associated with participatory washing. It serves both as an analytical tool for researchers and a practical guide for practitioners.

Future works requires a deeper and more exhaustive analysis of the current evolving landscape of participatory AI processes based on case studies. Further research could empirically test the proposed analytical framework across AI projects, explore the role of mediators in participatory processes, and investigate long-term stakeholder engagement combining interviews with diverse stakeholders with a more robust analysis of existing literature in the field. In this regard, future work would also benefit from the development of more actionable and practice-oriented tools for participatory design which could support practitioners in structuring and operationalizing participation in concrete AI development workflows.

References

[1] A. Birhane, W. Isaac, V. Prabhakaran, M. Diaz, M. C. Elish, I. Gabriel, S. Mohamed, Power to the People? Opportunities and Challenges for Participatory AI, in: Proceedings of the 2nd ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization, EAAMO '22, Association for Computing Machinery, New York, NY, USA, 2022. URL: https://doi.org/10.1145/3551624.3555290. doi:10.1145/3551624.3555290.

[2] M. Sloane, E. Moss, O. Awomolo, L. Forlano, Participation is not a Design Fix for Machine Learning, in: Proceedings of the 37th International Conference on Machine Learning (ICML), volume 119 of *Proceedings of Machine Learning Research*, PMLR, Vienna, Austria, 2020. Copyright 2020 by the author(s).

[3] European Commission. Directorate-General for Research and Innovation, Taking European Knowledge Society Seriously, 2007. URL: https://op.europa.eu/en/publication-detail/-/publication/5d0e77c7-2948-4ef5-aec7-bd18efe3c442/language-en, retrieved April 11, 2025.

[4] A. Fung, Varieties of participation in complex governance, Public Administration Review 66 (2006) 66–75.

[5] F. Raniolo, La partecipazione politica. Fare, pensare, essere, Il Mulino, Bologna, 2024.

[6] Y. Sintomer, J. De Maillard, The Limits to Local Participation and Deliberation in the French «politique de la ville», European Journal of Political Research 46 (2007) 503–529.

[7] G. Gourgues, S. Rui, S. Topçu, Gouvernementalité et Participation. Lectures Critiques, Participations 2 (2013) 7–33.

[8] G. Moini, Teoria critica della partecipazione. Un approccio sociologico, Franco Angeli, Milano, 2012.

[9] G. Moini, Participation, Neoliberalism and Depoliticisation of Public Action, SocietàMutamentoPolitica 8 (2017) 129–145. URL: https://doi.org/10.13128/SMP-20853. doi:10.13128/SMP-20853.

[10] F. Delgado, S. Yang, M. Madaio, Q. Yang, The Participatory Turn in AI Design: Theoretical Foundations and the Current State of Practice, in: Proceedings of the 3rd ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization, EAAMO '23, Association for Computing Machinery, New York, NY, USA, 2023. URL: https://doi.org/10.1145/3617694.3623261. doi:10.1145/3617694.3623261.

[11] C. Stimson, R. Raper, Participatory AI: A Method for Integrating Inclusive and Ethical Design Considerations into Autonomous System Development, in: M. Huda, M. Wang, T. Kalganova (Eds.), Towards Autonomous Robotic Systems. TAROS 2024, volume 15051 of *Lecture Notes in Computer Science*, Springer, Cham, 2025. doi:10.1007/978-3-031-72059-8_13.

[12] M. Coeckelbergh, AI Ethics, The MIT Press Essential Knowledge Series, MIT Press, Cambridge, Massachusetts, 2020.

[13] F. De Nardis, Sociologia politica. Per comprendere i fenomeni politici contemporanei, nuova edizione ed., McGraw-Hill Education, Milano, 2023.

[14] F. Raniolo, La partecipazione politica, Il Mulino, Bologna, 2007.

[15] D. Della Porta, La partecipazione nelle istituzioni: concettualizzare gli esperimenti di democrazia deliberativa, Partecipazione e Conflitto (2008) 15–42.

[16] R. A. W. Rhodes, Understanding Governance: Policy Networks, Reflexivity and Accountability, Open University Press, Buckingham, 1997.

[17] A. Fung, E. O. Wright, Deepening Democracy, 2003.

[18] G. Moini, T. Nupieri, Forme emergenti e "nuove" pratiche della partecipazione, in: Sociologia della politica contemporanea, Carocci, 2024, pp. 269–279.

[19] R. Segatori, Sociologia dei fenomeni politici, Laterza, Bari-Roma, 2012.

[20] F. W. Scharpf, Governing in Europe: Effective and Democratic?, Oxford University Press, Oxford, 1999.

[21] V. A. Schmidt, Democracy and legitimacy in the european union revisited: Input, output and 'throughput', Political Studies 61 (2013) 2–22.

[22] V. Schmidt, M. Wood, Conceptualizing Throughput Legitimacy: Procedural Mechanisms of Ac-

countability, Transparency, Inclusiveness and Openness in EU Governance, Public Administration 97 (2019) 727–740. doi:10.1111/padm.12615.
[23] M. Kaldor, The idea of global civil society, International Affairs 79 (2003) 583–593.
[24] J. A. Scholte, Civil society and democracy in global governance, Global Governance 8 (2002) 281–304.
[25] European Commission, European Governance: A White Paper, 2001. URL: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=LEGISSUM%3Al10109, accessed: 2025-04-11.
[26] P. Savoldi, Giochi di partecipazione. Forme territoriali di azione collettiva, in: Giochi di partecipazione, Franco Angeli, 2006, pp. 5–176.
[27] E. D'Albergo, G. Moini, Pratiche partecipative e politiche pubbliche: studi di caso a Roma, Rivista delle politiche sociali (2006) 365–385.
[28] V. Laino, Co-progettazione e processi partecipativi in Regione Toscana: ricognizione e toolkit per l'analisi e la valutazione, Master's thesis, Sapienza Università di Roma, Facoltà di Scienze Politiche, Sociologia, Comunicazione, 2023. URL: https://giovanisi.it/app/uploads/2023/07/LAINO-VITTORIA_TESI-1.pdf.
[29] J. Greenwood, Organized Civil Society and Democratic Legitimacy in the European Union, British Journal of Political Science 37 (2007) 333–357.
[30] H. C. Jenkins-Smith, Analytical Debates and Policy Learning: Analysis and Change in the Federal Bureaucracy, Policy Sciences 21 (1988) 169–211.
[31] P. A. Sabatier, An Advocacy Coalition Framework of Policy Change and the Role of Policy-Oriented Learning Therein, Policy Sciences 21 (1988) 129–168.
[32] E. Wollenberg, J. Anderson, D. Edmunds, Pluralism and the Less Powerful: Accommodating Multiple Interests in Local Forest Management, International Journal of Agricultural Resources, Governance and Ecology 1 (2001) 199–222. URL: https://doi.org/10.1504/IJARGE.2001.000013. doi:10.1504/IJARGE.2001.000013.
[33] N. McKeon, Are Equity and Sustainability a Likely Outcome When Foxes and Chickens Share the Same Coop? Critiquing the Concept of Multistakeholder Governance of Food Security, Globalizations 14 (2017) 379–398.
[34] M. Fougère, N. Solitander, Dissent in Consensusland: An Agonistic Problematization of Multi-Stakeholder Governance, Journal of Business Ethics 164 (2020) 683–699.
[35] J. Roloff, Learning from Multi-Stakeholder Networks: Issue-Focussed Stakeholder Management, Journal of Business Ethics 82 (2008) 233–250.
[36] D. Edmunds, E. Wollenberg, A Strategic Approach to Multistakeholder Negotiations, Development and Change 32 (2001) 231–253.
[37] L. Pellizzoni, Bridging Promises and (Dis)illusions: Deliberative Democracy in an Evolutionary Perspective, in: R. Beunen, K. Van Assche, M. Duineveld (Eds.), Evolutionary Governance Theory, Springer International Publishing, Dordrecht, 2015, pp. 215–232.
[38] M. Johnson, J. M. Bradshaw, P. J. Feltovich, C. M. Jonker, M. B. van Riemsdijk, M. Sierhuis, Coactive design: designing support for interdependence in joint activity, J. Hum.-Robot Interact. 3 (2014) 43–69. URL: https://doi.org/10.5898/JHRI.3.1.Johnson. doi:10.5898/JHRI.3.1.Johnson.
[39] E. Ostrom, Crossing the Great Divide: Coproduction, Synergy and Development, World Development 24 (1996) 1073–1087. URL: https://www.sciencedirect.com/science/article/abs/pii/0305750X9600023X.
[40] L. Bherer, P. Dufour, F. Montambeault, The participatory democracy turn: an introduction, Journal of Civil Society 12 (2016) 225–230. doi:10.1080/17448689.2016.1216383.
[41] E. Corbett, R. Denton, S. Erete, Power and Public Participation in AI, in: Equity and Access in Algorithms, Mechanisms, and Optimization (EAAMO '23), ACM, 2023. doi:10.1145/3617694.3623228.
[42] M. Steinbach, J. Sieweke, S. Süß, The Diffusion of E-Participation in Public Administrations: A Systematic Literature Review, Journal of Organizational Computing and Electronic Commerce 29 (2019) 61–95. URL: https://doi.org/10.1080/10919392.2019.1606849. doi:10.1080/10919392.2019.1606849.

[43] United Nations, United Nations E-Government Survey 2020: Digital Government in the Decade of Action and Sustainable Development, 2020. URL: https://publicadministration.un.org/egovkb/en-us/Reports/UN-E-Government-Survey-2020, accessed: 2025-04-10.

[44] M. Deseriis, Rethinking the digital democratic affordance and its impact on political representation: Toward a new framework, New Media & Society 23 (2020) 2452–2473. doi:10.1177/1461444820929678, original work published 2021.

[45] M. Deseriis, Reducing the Burden of Decision in Digital Democracy Applications: A Comparative Analysis of Six Decision-making Software, Science, Technology, & Human Values 48 (2021) 401–427. doi:10.1177/01622439211054081, original work published 2023.

[46] H. Landemore, Fostering More Inclusive Democracy with AI, *Finance & Development Magazine*, International Monetary Fund, 2023. URL: https://www.imf.org/en/Publications/fandd/issues/2023/12/Fostering-more-inclusive-democracy-with-AI-Landemore, accessed: 2025-04-11.

[47] B. Schneier, et al., How Artificial Intelligence can Aid Democracy, *Slate*, 2023. URL: https://slate.com/technology/2023/04/ai-public-option.html, april 21, 2023.

[48] S. F. Auliya, O. Kudina, A. Y. Ding, et al., AI versus AI for Democracy: Exploring the Potential of Adversarial Machine Learning to Enhance Privacy and Deliberative Decision-Making in Elections, AI Ethics (2024). doi:10.1007/s43681-024-00588-2, online first.

[49] M. Coeckelbergh, H. S. Saetra, Climate Change and the Political Pathways of AI: The Technocracy-Democracy Dilemma in Light of Artificial Intelligence and Human Agency, Technology in Society 75 (2023) 102406. doi:10.1016/j.techsoc.2023.102406.

[50] A. Jungherr, Artificial Intelligence and Democracy: A Conceptual Framework, Social Media + Society (2023). doi:10.1177/20563051231186353, first published online July 16, 2023.

[51] M. Coeckelbergh, Democracy as Communication: Towards a Normative Framework for Evaluating Digital Technologies, Contemporary Pragmatism 21 (2024) 217–235.

[52] M. Coeckelbergh, LLMs, Truth, and Democracy: An Overview of Risks, Science and Engineering Ethics 31 (2025). doi:10.1007/s11948-025-00529-0.

[53] S. Hossain, S. I. Ahmed, Towards a New Participatory Approach for Designing Artificial Intelligence and Data-Driven Technologies, 2021. URL: https://arxiv.org/abs/2104.04072. arXiv:2104.04072.

[54] D. Schuler, A. Namioka (Eds.), Participatory Design: Principles and Practices, 1st ed., CRC Press, 1993. doi:10.1201/9780203744338.

[55] T. Khan, T. Park, AI Needs Inclusive Stakeholder Engagement Now More Than Ever, https://partnershiponai.org/ai-needs-inclusive-stakeholder-engagement-now-more-than-ever/, 2024. Accessed: 2025-04-11.

[56] M. De Sanctis, P. Inverardi, P. Pelliccione, Do Modern Systems Require New Quality Dimensions?, in: A. Bertolino, J. Pascoal Faria, P. Lago, L. Semini (Eds.), Quality of Information and Communications Technology. QUATIC 2024, volume 2178 of *Communications in Computer and Information Science*, Springer, Cham, 2024. doi:10.1007/978-3-031-70245-7_6.

[57] S. G. Verhulst, Computational Social Science for the Public Good: Towards a taxonomy of governance and policy challenges, ????

[58] T. Davies, A. Colom, L. Velkova, M. Poblet, L. Nobbs, L. Kuneva, Participatory AI: Forging Shared Frameworks for Action, https://www.techpolicy.press/participatory-ai-forging-shared-frameworks-for-action/, 2025. Tech Policy Press.

[59] J. Simonsen, T. Robertson, Routledge International Handbook of Participatory Design, Routledge, 2012.

[60] Ada Lovelace Institute, Participatory Data Stewardship, 2021. URL: https://www.adalovelaceinstitute.org/report/participatory-data-stewardship/.

[61] A. Berditchevskaia, K. Peach, E. Malliaraki, Participatory AI for Humanitarian Innovation: A Briefing Paper, Technical Report, Nesta, London, 2021.

[62] E. Kallina, J. Singh, Stakeholder Involvement for Responsible AI Development: A Process Framework, in: Equity and Access in Algorithms, Mechanisms, and Optimization (EAAMO '24), ACM, 2024. doi:10.1145/3689904.3694698.

[63] M. Sloane, Controversies, contradiction, and "participation" in AI, Big Data & Society 11 (2024)

20539517241235862. doi:10.1177/20539517241235862, original work published 2024.
[64] A. Parthasarathy, A. Phalnikar, A. Jauhar, D. Somayajula, G. S. Krishnan, B. Ravindran, Participatory Approaches in AI Development and Governance: A Principled Approach, 2024. URL: https://arxiv.org/abs/2407.13100. arXiv:2407.13100.
[65] S. R. Arnstein, A Ladder of Citizen Participation, Journal of the American Institute of Planners 35 (1969) 216–224. doi:10.1080/01944366908977225.
[66] M. Ananny, K. Crawford, Seeing Without Knowing: Limitations of the Transparency Ideal and Its Application to Algorithmic Accountability, New Media & Society 20 (2018) 973–989. URL: https://doi.org/10.1177/1461444816676645. doi:10.1177/1461444816676645.
[67] B. Attard-Frost, A. De Los Rìos, D. R. Walters, The ethics of AI Business Practices: a Review of 47 AI Ethics Guidelines, AI and Ethics 3 (2022) 389–406. doi:https://doi.org/10.1007/s43681-022-00156-6.
[68] F. Bowen, After Greenwashing: Symbolic Corporate Environmentalism and Society, Organizations and the Natural Environment, Cambridge University Press, 2014.
[69] T. Metzinger, EU Guidelines: Ethics Washing Made in Europe, https://www.tagesspiegel.de/politik/eu-guidelines-ethics-washing-made-in-europe/24195496.html, 2019. Accessed April 2025.
[70] O. Freiman, Making sense of the conceptual nonsense 'trustwhorty AI', AI and Ethics 3 (2023) 1351–1360. doi:https://doi.org/10.1007/s43681-022-00241-w.
[71] L. Floridi, Translating Principles into Practices of Digital Ethics: Five Risks of Being Unethical, Philosophy and Technology 32 (2019) 185–194. doi:https://doi.org/10.1007/s13347-019-00354-x.
[72] M. Schultz, L. G. Conti, P. Seele, Digital Ethicswashing: a Systemic Review and a Process-Perception-Framework, AI and Ethics (2024) 169–177. doi:10.1007/s43681-024-00430-9.
[73] E. Bietti, From Ethics Washing to Ethics Bashing: a View on Tech Ethics from within Moral Philosophy, in: Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency, 2020. doi:https://doi.org/10.1145/3351095.3372860.
[74] E. Keymolen, Trustworthy tech companies: talking the talk or waling the walk?, AI and Ethics 4 (2024) 169–177. doi:https://doi.org/10.1007/s43681-022-00254-5.
[75] K. Hofer, D. Kaufmann, Actors, arenas and aims: A conceptual framework for public participation, Planning Theory 22 (2022) 357 – 379. URL: https://api.semanticscholar.org/CorpusID:253533431.
[76] M. Hisschemöller, R. Hoppe, Coping with Intractable Controversies: The Case for Problem Structuring in Policy Design and Analysis, Knowledge and Policy 8 (1995) 40–60. URL: https://doi.org/10.1007/BF02832229. doi:10.1007/BF02832229.
[77] N. Nelson, S. Wright, Participation and Power, in: N. Nelson, S. Wright (Eds.), Power and Participatory Development: Theory and Practice, Intermediate Technology Publications, London, 1995, pp. 1–18.
[78] J. Newig, Does Public Participation in Environmental Decisions Lead to Improved Environmental Quality? Towards an Analytical Framework, Communication, Cooperation, Participation (International Journal of Sustainability Communication) 1 (2007) 51–71.
[79] J. Gaventa, Finding the Spaces for Change: A Power Analysis, IDS Bulletin 37 (2006) 23–33.
[80] K. Crawford, Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence, Yale University Press, New Haven, CT, 2021.
[81] C. O'Neil, Weapons of Math Destruction, Penguin Books, Harlow, England, 2016.
[82] A. Malizia, F. Paternò, Why Is the Current XAI Not Meeting the Expectations?, Commun. ACM 66 (2023) 20–23. URL: https://doi.org/10.1145/3588313. doi:10.1145/3588313.
[83] H.-F. Cheng, L. Stapleton, R. Wang, P. Bullock, A. Chouldechova, Z. S. Wu, H. Zhu, Soliciting Stakeholders' Fairness Notions in Child Maltreatment Predictive Systems, in: Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems, ACM, 2021, pp. 1–17. doi:10.1145/3411764.3445665.
[84] D. Martin, G. Hayes, Ethnographic and Participatory Methods in Information Systems Research: Accessing Marginalized Communities, International Journal of Human-Computer Interaction 30 (2014) 393–408.

[85] M. K. Lee, D. Kusbit, A. Kahng, J. T. Kim, X. Yuan, A. Chan, D. See, R. Noothigattu, S. Lee, A. Psomas, A. D. Procaccia, E. Y. Huang, WeBuildAI: Participatory Framework for Algorithmic Governance, Proceedings of the ACM on Human-Computer Interaction 3 (2019) 1–35. doi:10.1145/3359283.

[86] A. Fung, Putting the Public Back into Governance: The Challenges of Citizen Participation and Its Future, Public Administration Review 75 (2015) 513–522. URL: https://doi.org/10.1111/puar.12361. doi:10.1111/puar.12361.

[87] D. T. K. Ng, J. K. L. Leung, S. K.-W. Chu, M. S. Qiao, Conceptualizing AI literacy: An exploratory review, Comput. Educ. Artif. Intell. 2 (2021) 100041. URL: https://api.semanticscholar.org/CorpusID:244514711.

[88] H. R. Lee, S. Šabanović, W.-L. Chang, S. Nagata, J. Piatt, C. Bennett, D. Hakken, Steps toward participatory design of social robots: mutual learning with older adults with depression, in: Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction (HRI '17), ACM, 2017, pp. 244–253. doi:10.1145/2909824.3020237.

Declaration on Generative AI

The authors have employed Chat-GPT4*omni* in order to: Sentence Polishing.