Cambridge Forum on AI: Culture and Society (2026), **2**, e12, 1–20
doi:10.1017/cfc.2026.10026

Cambridge Forum

RESEARCH ARTICLE

Deliberating the algorithmic future: Reconfiguring AI ethics through citizen juries

Ana Pop Stefanija and Rob Heyman

imec-SMIT, Vrije Universiteit Brussel, Brussels, Belgium
Corresponding author: Ana Pop Stefanija; Email: ana.pop.stefanija@vub.be

(Received 1 July 2025; revised 23 March 2026; accepted 13 April 2026)

Abstract
Mainstream artificial intelligence (AI) ethics primarily takes vertical approaches: top-down, principle-driven and expert-led frameworks that often exclude the very publics affected by algorithmic systems. This paper explores whether citizen juries (CJs) can function as a horizontal alternative empirical ethics method that emphasizes participation, situated knowledge and epistemic justice. We conducted a citizen jury with Brussels residents to deliberate on the use of an AI system for electricity distribution during brownouts. The jury process combined guidance ethics, speculative fiction, and generative AI tools to scaffold inclusive deliberation. Using affordance theory, we analyze how the citizen jury method structurally mediates certain forms of participation and ethical reasoning. We evaluate its normative value through the lenses of epistemic justice and effective performativity. Our findings show that citizen juries can foster testimonial inclusion and support hermeneutical sense-making, offering a partial corrective to abstract, generic AI ethics. However, we also identify critical constraints: problem-posing was narrowed, alternative epistemologies discouraged, and the outputs risked reproducing the generic performativity forms of vertical ethics. While citizen juries hold promise for doing AI ethics otherwise, their transformative potential hinges on reflexively examining their own methodological affordances and institutional framing.

Keywords: AI ethics; citizen jury; empirical ethics; epistemic justice; horizontal ethics; participatory methods

1. Introduction

As artificial intelligence (AI) systems become increasingly embedded in critical infrastructures and social decision-making, the need to establish ethical oversight intensifies. Yet dominant approaches to AI ethics are often shaped by vertical orientations: there is a focus on top-down principles, abstract values and expert-driven governance mechanisms that rarely engage those most affected by technological systems. These vertical AI ethics approaches have been extensively critiqued for their vagueness, lack of enforceability, susceptibility to ethics washing (Casiraghi 2023; Green 2021; Munn 2022) and "law-conception of ethics" (Rességuier and Rodrigues 2020). What remains underexplored is whether empirical ethics – ethics grounded in horizontal, deliberative bottom-up participation that accounts for the situated, contextual, and lived experience of those it affects – can offer a more inclusive, responsive and epistemically just alternative.

This paper investigates that possibility by experimenting with a citizen jury (CJ) as a methodological and normative intervention in AI ethics. Citizen juries are a well-established method for resolving issues of public interest, informing policymaking and providing guidelines for issue resolution. We frame citizen juries not merely as deliberative exercises, but as artifacts: participatory

infrastructures whose architectures shape who can speak, what can be known, and how ethics is enacted. Unlike vertical approaches that codify norms from above, the jury process invites citizens as non-experts to grapple with value-laden trade-offs at the heart of sociotechnical design. Through this, we explore whether citizen juries can support horizontal AI ethics grounded in situated knowledge, mutual learning, and epistemic pluralism.

Our study builds on an applied case: a citizen jury conducted with a representative sample of Brussels residents, concerning the development and use of a (speculative) AI system to allocate electricity during a brownout scenario. The jury was embedded in a broader research project on the design of future energy infrastructures and involved scholars from social, computer and legal sciences. The strict and highly codified format of the citizen jury was adapted to fit the specifics of the envisioned AI technology by employing approaches and tools such as the guidance ethics approach (Verbeek and Tijink 2020), speculative fiction (Nathan et al. 2007) and Generative AI (GenAI) visualizations to scaffold deliberation. In repurposing the citizen jury as an empirical ethics method, we asked not only *what* citizens concluded, but *how* the method itself shaped what was possible to deliberate, know, imagine and decide.

We start our inquiry by asking: Can we apply mediation frameworks, such as affordances, to a non-material thing like a participatory method? We approach this citizen jury method reflexively, analyzing the affordances of the citizen jury as a research device: what it allows, demands, refuses, and discourages (Davis 2020; Davis and Chouinard 2017) in terms of epistemic practices and participatory agency. We complement this with a normative evaluation along two axes. The first, internal, aims to answer the question: does the jury enable epistemic justice by including non-expert (and in that sense marginalized) knowers and supporting hermeneutical inclusion, where various forms of knowing, experiencing, perspectives and situatedness are taken into consideration and accounted for? The second, the external one, focuses on the question: does it achieve effective performativity by producing outputs that can meaningfully reshape AI development, rather than reaffirming the AI technology under investigation? To be more specific, is the method transformative or affirmative? Throughout, we remain attentive to the tensions, limits, and contradictions of participatory methods – especially the risk that horizontal forms are reabsorbed into top-down, vertical frameworks. Citizen juries hold potential as instruments of horizontal, participatory, and empirical ethics, but only if their design, implementation, and evaluation are critically and rigorously examined. This paper offers such a critical examination.

We do so by first examining citizen juries as artifacts and architectures of participation, and by describing how this form of technological assessment co-shapes moral reasoning. Next, we introduce a more normative framework, in which we evaluate citizen juries as a method for participatory AI ethics through the lens of *affordances* (Davis 2020; Davis and Chouinard 2017). We do this by looking at citizen jury from the perspective of *epistemic justice*, in particular, *testimonial* and *hermeneutical justice*. This framework allows us to further critically evaluate and discuss citizen jury along the lines of *bottom-up* vs *top-down, horizontal* vs *vertical ethics, situated knowledge* vs *view from nowhere, full participation* vs *participation bias, problem posing* vs *issue-setting, performative* vs *transformative ethics, participatory solutionism* vs *ethical solutionism* and the nuances in between.

Our reflexive and critical insights are informed by the outputs produced during the jury process, the collected and analyzed observational data, and our active participation in the conceptualization, preparation and execution of the citizen jury. We use the outputs produced by the jurors (a list of identified and prioritized households, identified and prioritized values, and conditions and principles to enforce these values), the collected observational data (24 hours of transcripts from the jury sessions, notes from observers), survey answers (pre- and post-jury survey), session activities (future scenarios stories, exercises), reflections and elaboration (on the reasoning for particular prioritizations) and the debriefs between the researchers, observers and facilitators. While this article does not focus on the content of the outputs, we analyzed them in order to critically investigate and discuss citizen juries' potential for empirical, horizontal and transformative AI ethics.

The article proceeds in four steps. We begin by situating citizen juries as a participatory method within debates on empirical and horizontal AI ethics. We then motivate and develop our analytical perspective to treat citizen juries as artifacts, as architectures of participation whose design mediates ethical reasoning and epistemic agency. Building on this, we introduce an affordance-based analytical framework that we use to examine how citizen juries enable, constrain, or shape ethical deliberation in practice. This framework is applied reflexively through the heuristic of *the good, the bad and the ugly*, allowing us to critically assess both the emancipatory potential and the structural limitations of citizen juries. Finally, we deepen this evaluation by mobilizing epistemic justice as a normative lens, foregrounding the method's promises, tensions, and failure modes as a tool for genuinely participatory AI ethics.

2. Citizen juries as architectures of participation

A Citizen jury (or assembly, or a council) is a (temporary, randomized and representative) assembly of citizens that convenes in order to discuss and give recommendations on a particular issue (DemocracyNext 2023a; Smith and Wales 2000; Street et al. 2014; Wells et al. 2021). As such, it is a process and not a one-time event. The aim is to find common ground and develop collective recommendations, often to inform policy and decision-making (DemocracyNext 2023a; PeoplePowered 2021). This aim is to create democratic spaces for everyday people to grapple with, discuss and propose solutions to critical issues that are or will affect them in the future, thus establishing them as an equal party in the decision- and policymaking process. It is, therefore, crucial that the entire process is structured in a way that enables and facilitates learning, deliberation, consensus-building, recommendations drafting and final decision-making in the form of recommendations (DemocracyNext 2023a; OECD 2021). As a form of deliberative democracy, citizen juries are founded on the idea that individuals from different backgrounds and without special prior knowledge or expertise are capable, able and should tackle a public (policy) question.

Citizen juries have been established and used for policymaking on a range of issues, including environmental and climate policies, transportation, strategic and urban planning, health services, public budgeting and water tariffs (see the OECD database [OECD 2023], counting 733 cases until 2023). The citizen methodology and process can also be considered highly codified and structured (see, e.g., the DemocracyNext 2023a guide and the PeoplePowered 2020 Handbook), involving multiple actors with distinct roles and a chronologically ordered set of steps. In Figure 1, we illustrate in detail the planning and design processes in the preparation (pre-jury deliberation) phase, the (internal and external) parties involved, the timeline with each step in the three stages (preparation, execution, analysis), as well as the jury prompt, participant demographics and all the outputs of the jury deliberation. The codified and structured nature of the citizen jury is evident in the assignment of roles and functions to the various actors across all phases of the process: researchers, facilitators and co-facilitators, observers, expert-witnesses and participants.[1]

Not every societal issue is seen as appropriate for this type of public deliberation. In principle, the need for a citizen jury should satisfy the following criteria: it should raise concerns within a broader community in a situation when choices must be made, but there are no clear "right" answers; be related to a genuine problem, preferably an issue that is complex and involves value

[1]While the researchers are de facto the ones initiating the citizen jury process, each CJ process involves what we call "formal organizers" – an external hired party, an organization with experience in citizen participation and deliberative processes – who organize and facilitate the processes. As can be seen in Figure 1, a number of steps are foreseen in the planning and design process. The formal organizers are the ones who: define the jury demographics and the sortition algorithms; recruit and select participants; are in charge of organizational aspects (venue, payments, follow-up and communication with the participants); recruit facilitators and co-facilitators; define and prepare the program of jury activities; recruit expert witnesses; brief and debrief expert witnesses, and similar. Important decisions are made in agreement with the researchers, regarding the definition of the jury prompt (questions), the jury scenario, the expert witnesses and the pre- and post-jury questionnaire (if applicable). The researchers also serve as non-participant observers, too (but are not the only observers).

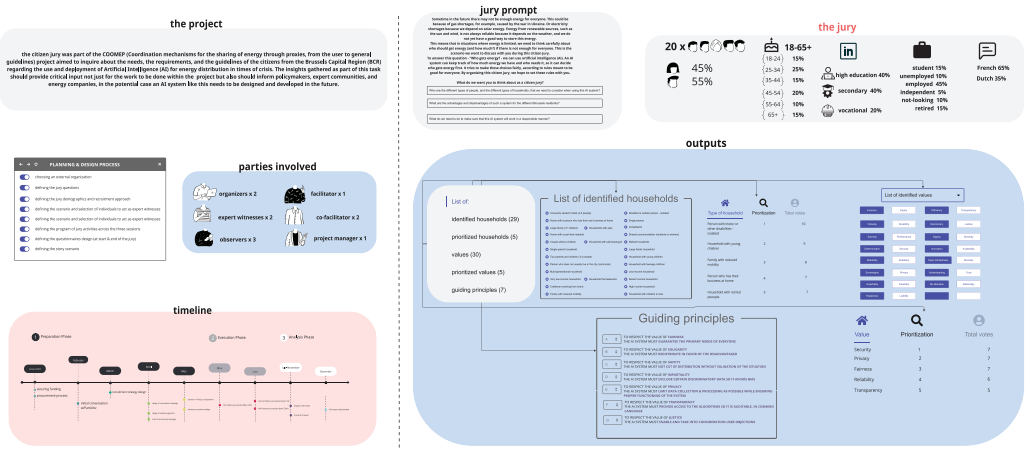

Figure 1. Overview of the citizen jury design, process, and outputs.

judgements; implies trade-offs; there are no straightforward and easy *yes/no* answers to the issue, and is underpinned by values-based dilemmas; and it is not solely technical nor requires (only) a technical solution. As such, it encourages and facilitates discussion of highly relevant and emerging issues by engaging a broader community of non-experts (Ada Lovelace Institute 2021b). The process is initiated and guided by a *jury charge* or *prompt* – a (set of) question(s) that sets the problem and guides the expected outcomes (see Fig. 1 for the specific jury prompt of the citizen jury).

Conceived and designed this way, a citizen jury would be an exemplary participatory and horizontal method of doing empirical AI ethics. However, as we elaborate in the remainder of the article, *it is complicated*. While citizen juries' distinct architecture of participation and affordances allow for inclusive, participatory and legitimate democratic processes, these very elements also come with shortcomings, limitations and risks.

2.1 A citizen jury for a technology-in-the-making

The citizen jury we discuss was undertaken within the COOMEP[2] project. As an interdisciplinary study involving scholars from social, computer and legal sciences, the citizen jury's goal was to inquire about the needs, the requirements, the guidelines and the "redlines" of the citizens from the Brussels Capital Region (BCR) regarding the use and deployment of AI for energy distribution in times of crisis (*brownout*). The main premise of the project was that energy distribution could be uncertain in the future. Following technological advancements, one potential approach to addressing the challenge of making complex decisions about how energy will be distributed during a brownout is to use automated decision-making agents. In our highly hypothetical, speculative and risk-heavy scenario, these complex and complicated decisions will be delegated to programmable agents. Deciding the energy distribution during a brownout entails determining who will be prioritized to receive energy and how much, and who will be deprioritized and potentially left behind. This raises not only ethical but also legal questions, as it might clash with fundamental human rights and other applicable laws and regulations.

To probe this potential future, we asked those most affected by the possible development and deployment of this technology – the residents of the Brussels Capital Region (BCR). How do they envision this decision-making process? What are the guiding principles and the redlines that must not be crossed when designing a sociotechnical system like this? Approximately 20 residents began

[2]COOMEP stands for "Coordination mechanisms for the sharing of energy through proxies, from the user to general guidelines."

drafting this potential future scenario based on the jury scenario and prompt (Fig. 1). We followed the established and codified methodology for deliberative participation, comprising mandatory and sequential phases of learning (knowledge-making), deliberation, and collaborative decision-making.

The calls for a participatory turn and the broader inclusion of citizens in technology development processes are not new. For example, Carugati (2020) has been advocating for citizen participation for algorithmic governance and regulation; McQuillan (2018) for ethical machine learning, and Hildebrandt and Gutwirth (2008) for technology assessment in general (2008). However, when it comes to deliberative participation in AI systems development, there are only a few instances. In the last years, there have been citizen juries organized about ethical AI and fair data-sharing in medical contexts (Information Commissioner's Office 2019), biometrics (Ada Lovelace Institute 2021a), datafication (Kennedy et al. 2022), data governance (Patel et al. 2022; Wong et al. 2023) and regarding the trade-offs between explainability vs accuracy of AI (van der Veer et al. 2021), but the full potential of the method has not been more widely used. In that sense, the citizen jury we elaborate on is one of the very few (if not the only) that addresses a particular (though speculative) AI technology under development.

2.2 (Evaluating) citizen juries as artifacts

Most approaches that deconstruct AI or tech ethics focus on the content of ethics and the inclusion or exclusion of ethical principles (Hagendorff 2020; Munn 2022). The literature does expand on the context or means of production of these guidelines to illustrate how this form of ethics is performative (Kerr et al. 2020), strategic, and often leads to ethics washing (Green 2021; Hagendorff 2020; Munn 2022). We move beyond this debate and study *the materiality of doing ethics* during technology assessment processes. We situate this approach in the *empirical turn* in the philosophy of technology (Verbeek 2006) and the *participatory turn* in technology design (Birhane et al. 2022; Delgado et al. 2023; Gourlet et al. 2024). To this end, we approach ethical assessment methods as artifacts. If we approach ethical deliberation in this way, we can study the mediation and affordances of ethics as a practice. This allows us to move away from the ethics content (*the what*) and toward the material context (*the how*) that mediates this content.

To study the materiality of doing (empirical) ethics, and to critically reflect on the potential of citizen juries for a horizontal, bottom-up approach to AI ethics, we reach out to, and apply, the vocabulary of affordance theory: affordance "refers to the range of functions and constraints that an object provides for, and places upon, structurally situated subjects" (Davis and Chouinard 2017, 241). Using affordances as a descriptive method allows for a description of human–artifact interactions that acknowledges agency on both ends: on the one hand, the artifact shapes actions by increasing or decreasing ease or friction. On the other hand, users have characteristics and goals that co-shape the execution of these actions. Verbeek (2006) proposes to investigate technology because technologies provide material answers to moral questions. We propose to investigate the affordances of technology assessment processes because they provide material answers to how we frame moral reflection.

In *Materializing Morality*, Verbeek's (2006) argument for having ethics of engineering is because technologies can mediate ethical agency: "When technologies coshape human actions, they give material answers to the ethical question of how to act" (p. 361). This line of reasoning has been applied only to technologies, but not to tools for ethical reflection or assessment. Can we apply mediation frameworks, such as affordances, to a non-material thing like a participatory method? The answer is positive. Building on the distinction between the various affordance mechanisms (Davis 2020; Davis and Chouinard 2017), we discuss citizen juries in terms of what they request, demand, allow, encourage, discourage and refuse regarding participation and empirical AI ethics (see Table 1 for examples). We use the affordance framework to describe architectures of participation. What is important to keep in mind is that the mechanisms of affordance of the citizen jury are analytical categories or "stopping points that help describe the intensity with which technological objects [or processes like

Table 1. Affordance mechanisms examples (following the conceptualization by Davis 2020; Davis and Chouinard 2017)

Affordance mechanism	Affordance mechanism definition
Request	This mechanism invites action without a hard enforcement. Example: A speed bump requests that a driver slow down but does not make speeding impossible.
Demand	A strong mechanism in which an artifact makes one path functionally necessary for engagement or progression. The action is effectively mandatory for use. Example: Facebook demands a gender selection during signup – registration cannot proceed without it.
Encourage	The artifact fosters or promotes a specific behavior while suppressing or deprioritizing alternatives. Example: Large dinner plates encourage overeating while small plates encourage the opposite.
Discourage	The artifact introduces barriers to certain actions, making them possible but effortful or normatively penalized. Example: Tinder discourages deep bio-based selection by restricting profile text and emphasizing quick swiping.
Refuse	The artifact makes an action functionally impossible under ordinary use. Refusal is a form of architectural or coded denial. Example: A car that refuses to lock when keys are in the ignition.
Allow	Neutral availability of options without pushing or pulling toward specific outcomes. It permits but does not guide. Example: Snapchat allows users to choose filters and recipients, without pressing them toward a particular configuration.

the citizen jury], facilitate or impede particular lines of action and social dynamics" (Davis 2020, 106). For example, when we talk about requests and demands as affordance mechanisms, we are talking about artifacts that instruct users (Davis 2020). In our case, the artifact is the codified structure of the citizen jury that must be respected. The affordance mechanisms of encourage, discourage and refuse respond to the actions of the user and articulate the extent to which an artifact allows user actions.

Participation methods use a broad spectrum of mechanisms to allow, encourage or forbid certain interactions. These choices often have a direct effect on the type of ethical reasoning allowed in an ethics deliberation. Many deliberative practices are approached as architectures in which specific methods are used to encourage or discourage certain behavior. For example, in brainstorming, there is a general rule of not criticizing ideas; it is forbidden to use "no, but," and it is encouraged to use "yes, and." The rule not to voice a "no, but" is an instruction or an affordance for participants that *encourages* a nonjudgmental perspective toward others. The rule to use "yes, and" encourages participants to listen, consider other participants' ideas and build on these. Further on, when we consider guidance ethics[3] (Verbeek and Tijink 2020), there is much framing regarding what ethical questions are *allowed*. Guidance ethics *demands* reflections that guide an innovation into society. This form of deliberation does not *allow* dissent, such as questioning whether a technology should be used as a solution in the first place. The "architecture of participation" within this method *demands* that this question cannot be asked.

In that regard, while we followed the codified structure of the citizen jury methodology, we also played with its method and research affordances and modified it according to the specificities of our own project. We adapted the strict format of citizen jury to fit the specifics of the envisioned AI technology in development by employing the guidance ethics approach (Verbeek and Tijink 2020), speculative fiction through value scenarios (Nathan et al. 2007), and GenAI for scenario visualizations. These "interventions" in the citizen jury script were possible because the process-architecture of the citizen jury allows for the use of various tools and venues for deliberation. We used the guidance ethics approach to foster a participatory, constructive ethical discussion. The aim of this particular

[3] *Guidance ethics embeds ethical reflection within the design, deployment and use of AI systems across their lifecycle. It adopts an "ethics from within" and bottom-up orientation, involving developers, users and affected publics in identifying context-specific ethical issues. Rather than focusing on harm avoidance alone, it advances a positive ethics approach that translates societal values into concrete design, governance and use practices.*

Figure 2. A few examples of the future scenario visualizations produced by the participants (with DALL-E).

approach is to construct the deliberative processes as a collaborative activity, where all stakeholders, both those directly and indirectly affected by the technology, are both involved and considered in the development, design, and deployment of the technology. The three stages of the approach focus on explaining the envisioned technology; on identifying and mapping the actors involved, the underlying values and the possible effects of the technology; and on the options for action to enforce these values and the conditions for use. In our jury design, these are translated as a list of identified and prioritized households (Fig. 1), a list of identified and prioritized values (Fig. 1), and a list of conditions and principles to enforce these values (Fig. 4). The design of the jury process in this way enabled us, as researchers, to also identify the sociotechnical imaginaries (Jasanoff 2015), the needs, requirements, guidelines and redlines for the technology. By analyzing the process outputs and the interactions captured during the deliberative sessions, we identified the negotiations, frictions and trade-offs that took place. We used speculative fiction (value scenarios) to make "imagining the future" even more tangible for the jurors. We invited them to write a future scenario (utopian or dystopian),[4] imagining what the future would look like if our speculative technology were to become a reality. The element of future-imagining requires both an understanding of how the technology works (including the dis/advantages, and the potential risks and harms), an engagement with lived experiences, and "walking in someone else's shoes" (considering different groups of citizens). This exercise in making the future "imaginable" and "predictable" helped the participants form opinions about the guiding principles and the redlines that should not be crossed, and to identify potential issues and their implications. To make these scenarios even more imaginable, we illustrated them using a GenAI tool (DALL-E) (Fig. 2).

[4]During the third jury session, we instructed the jurors to write a future scenario: either a utopian or dystopian story. We used the jury prompt itself (Fig. 1) as a reminder and gave them the following instructions: "Write a maximum of 15 lines per story. When writing your story, remember to answer the following questions: When? Where? With someone? If yes, with whom? What is happening?"

If we consider the citizen jury as we designed it, we can understand it not only as a process of participatory deliberation but also as an object of research. In that sense, the methods we use and can use are always scripted, and as researchers, we are always constrained by their research affordances (Kennedy et al. 2022; Law 2004; Marres and Gerlitz 2016). Marres and Gerlitz (2016, 22) call this "methodological bias" – as researchers, we navigate around what is afforded through particular methods and tools, and our analytical practices are influenced by the specific configurations of the object and method of research. This includes "sampling techniques, options for analysis" (Marres and Gerlitz 2016, 31), scope and depth of insight, and ultimately of intervention. We need to consider the co-shaping from ethics methods because it allows us to investigate what (Green 2021) calls "ethical solutionism" – when ethical principles and practices are seen and promoted as "the only solution to sociotechnical problems" (219). Green (2021) gives the example of the effort to solve algorithmic discrimination through algorithmic fairness, but by doing that through mathematical definitions of fairness, completely disregarding the "societal" in the equation. In this way, these mathematical/technical solutions do not seek systemic solutions but rather fast-tracked shortcuts, thereby bypassing the structural and systemic conditions that, in fact, cause the majority of algorithmic harms.

What are the implications of an approach like the one we propose? Our approach shifts the question from what the moral values or implications *of a technology are*, to what the moral values and implications that *can arise from a specific mediation are*. This question helps us answer whether this is the type of moral reflection and agency we want to pursue in the first place. This is in line with the arguments of Davis (2020; see also Davis and Chouinard 2017) that shifting the questions from *what* to *how* allows for "diverse subjects and circumstances" (40) because it allows to capture the underlying processes, the nuances, the specificities and the actors involved. By asking "how, for whom, and under what circumstances?," an affordances approach allows and encourages a relational positioning and analysis that takes into consideration both the demands and the inscriptions of the citizen jury methodology and our responses to them. We find the mechanisms and conditions framework of affordances (Davis 2020) particularly productive and widely applicable precisely because of its focus on the relationality and the fluidity of the object (method)–human relation, and their mutual co-constituency. Looking through the analytical lens of affordance theory also allows us to look into the practices that are not *requested* by the codified methodology of the citizen jury but are still *afforded*. This provides an opportunity to go beyond what is established and codified in the jury process, to repurpose it, and to add certain modifications and introduce additional elements and processes (e.g., introduce GenAI as a future-imagining tool, as we did). The repurposing of the affordances theory also allows for the application of more critical and self-reflective lenses, such as epistemic justice, that uncover and capture the nuances of the citizen jury process.

3. A dual-layered evaluation of citizen juries

We evaluate the citizen jury along two dimensions: internally, as a method for enabling deliberation, and externally, in terms of its capacity to reshape AI applications through ethical intervention. Internally, we draw on the concept of *epistemic justice* to assess whether the method enables fair and inclusive knowledge practices. Externally, we adopt the *sociology of expectations* (Kerr et al. 2020) to analyze how the citizen jury constructs visions of technological futures and whether it can meaningfully contest or reconfigure dominant top-down paradigms of AI ethics. Taken together, these perspectives enable a dual-layered evaluation and critique of the citizen jury, examining what it affords its participants and what it can demand from technology.

3.1 Epistemic justice

Citizen jury as a method has two key characteristics: (1) it aims to include a diverse group of citizens, and (2) these citizens are guided through a process to make informed decisions on complex issues.

We therefore want to evaluate the citizen jury as a tool that encourages, allows, demands, or requests epistemic justice, given that the latter is focused on the same goals. Epistemic justice (Fricker 2007) refers to fairness in knowledge practices, particularly in how individuals are treated as *knowers* and contributors to understanding and knowledge. Epistemic justice aims to avoid two types of epistemic injustice: *testimonial injustice*, when someone's knowledge is undervalued due to prejudice (e.g., due to race, gender, class), and *hermeneutical injustice* – when a person lacks the interpretive tools or social recognition to make sense of their experiences due to systemic marginalization. Citizen jury addresses testimonial injustice as a method because it *demands* a diverse group of stakeholders who would otherwise be undervalued. It is one of the core components of this method. A citizen jury aims to reduce hermeneutical injustice through a deliberative process that allows citizens to explore an issue, ask questions and build confidence in discussing a complex issue.

Citizen juries, seen through this lens as a technology assessment tool in the pre-development phase of a technology-in-the-making, demand, request, and encourage several critical aspects that make it distinctive from the dominant *AI Ethics as usual* paradigm: testimonial and hermeneutical justice, situated knowledge, horizontal ethics, and transformative instead of performative ethics. The citizen jury approach acknowledges and embraces the fact that everyone 'comes from somewhere' and treats this as an essential and productive component. It intentionally looks for the lived experiences, for the individual perspectives, the contextualized circumstances, and the pressing concerns of the (potentially) affected communities. It looks for the partial visions and the situated and embodied experiences (Haraway 1988). It enables a "breathing space" (Ruckenstein 2024), a "shared space for articulating and grappling with issues at stake in digital society-making" (11) that should safeguard the ability to think critically, reflect, foresee, intervene and develop a technology and a 'future-in-the-making' in accordance with individuals' own values and interests. As such, by insisting on the valuable and legitimate experiential and situated knowledge of different, intersectionally situated individuals, it goes against the predominantly technocratic and expert engagement in institutionalized AI ethics. These top-down, vertical AI ethics-producing practices insist on creating objective and neutral, 'speaking for all' perspectives that should be applicable to all contexts, uses and stakeholders/affected parties. But as Haraway (1988) says, there is no 'view from nowhere'. This is a form of relativism that flattens out all differences, a "way of being nowhere while claiming to be everywhere equally" (Haraway 1988, 584). This is precisely the perspective that has been advocated with vertical AI ethics: who gets to say what matters and who matters, whose values, needs, experiences and contexts are valid and will be taken into consideration is decided by a handful of actors in more powerful positions (technocrats, developers, consultants).

By demanding a process of information-acquisition and knowledge-making, citizen jury also goes against hermeneutical injustice. AI ethics has often been weaponized against those it affects the most, and has been framed as an elitist undertaking inaccessible 'to the masses' since it is considered as *demanding* a specific expert knowledge (either technical, legal, or ethical) in order to understand the technology in question and to get a seat at the proverbial table. On the one hand, the jurors are not chosen because of their professional expertise, nor are they expected to have one. On the contrary, it is their unique situatedness, experiential and tacit knowledge that makes them the appropriate 'experts' to be consulted. On the other hand, while they do lack the professionalized interpretive tools to approach complex subjects (e.g., AI technologies), within the mandatory and codified first phase of the jury process – the learning phase (DemocracyNext 2023b), they are provided with information and knowledge in order to form a better understanding of the issue at hand. Jurors are not expected to become experts, but they are given the necessary basic tools to approach and dissect the problem they are invited to deliberate on. In that sense, the citizen jury is also an empowering process that disturbs the established epistemic power asymmetries and brings hermeneutical recognition of the lived experiences and tacit knowledges of the jurors as members of particular (intersectional) communities.

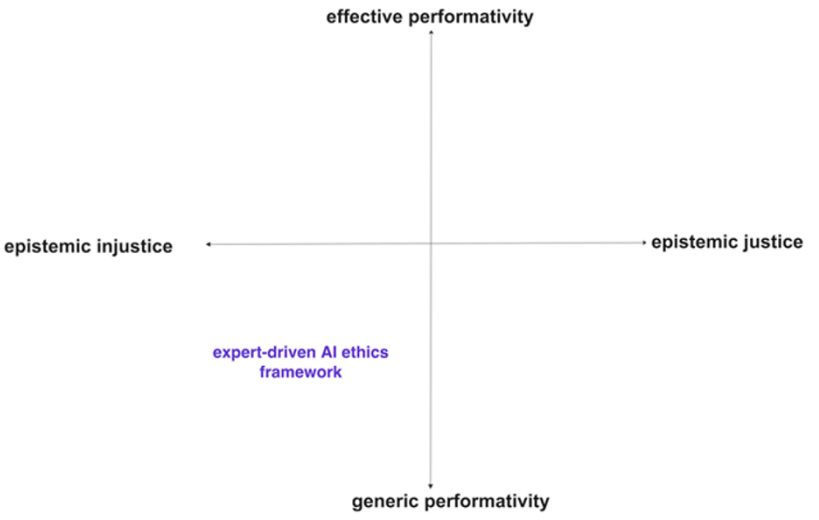

Figure 3. Evaluating and situating AI ethics.

In that regard, as a participatory, empirical and practice-oriented approach that aims to intervene in the development phases of technology, the citizen jury holds the potential to become a *transformative* participatory tool. Building from the bottom up, acknowledging and overcoming epistemic and hermeneutical power asymmetries, citizen jury for AI ethics can be an empowering process and an emancipatory tool. Against institutionalized, vertical AI ethics that produces ethics lists and guidelines untranslatable in practice, a citizen jury for ethics development can be an instrument that could even lead to structural change, enable self-determination, and the cultivation and acquisition of new skills and knowledge. In that sense, ethics here should have an active role, a technology (or world) building role and not (only) a risk-mitigation one.

3.2 Generic and effective performativity

We will use epistemic justice as an internal criterion to evaluate if the citizen jury method delivers on its promise. We also include an external evaluation criterion considering the degree of change the citizen jury can bring to the world. Kerr et al. (2020) refer to MacKenzie (2008) to analyze policy documents' implicit expectations of AI. Two things are important to this paper: first, that expectations statements in documents manage uncertainty in innovation: "Formal mechanisms include foresight and research prioritisation exercises, which are deployed by governments, consultancy firms and companies to rationalise future innovation agendas and investment" (Kerr et al. 2020, 3). AI ethics frameworks can also be formal mechanisms that rationalize AI ethics. The difference between generic and effective performativity is the second reason to include this perspective. The latter refers to how these expectations in statements allow change in technology or social actions. Generic performativity does not create change, whereas effective performativity enables change. This dichotomy seems to be what many AI ethics frameworks are criticized for: being too vague, toothless (Rességuier and Rodrigues 2020), and open to ethics washing (Green 2021; Wagner 2018), which means that, while they declaratively do aim for change, they rarely allow any meaningful change. The integration of both perspectives allows us to situate generic top-down AI ethics and citizen juries in Figure 3, along the axes of epistemic justice/injustice and generic/effective performativity. We will investigate where we can situate which parts of the citizen jury in the quadrants of Figure 3 and revisit it in the last section of the article as Figure 5.

4. The good affordances of a citizen jury

In this section, we outline the citizen jury as a process and identify its affordance mechanisms. This allows us to both define the method and identify these affordances simultaneously.

If we follow the concepts of affordances theory, more specifically the affordance mechanisms (Davis 2020; Davis and Chouinard 2017), we could say that citizen juries *allow* and *encourage* citizens to participate in decision-making processes that will *allow* them to influence the outcomes. Being prescriptive and codified as a methodology (see DemocracyNext 2023a; OECD Guidelines for Citizen Participation Processes 2022; People Powered 2021), the citizen jury also *demands* that a representative sample of a particular, potentially affected by the issue under discussion, community is recruited and participates in the jury (see Fig. 1). In that sense, it *refuses* a random participant selection.

The citizen jury is founded on the premise that once the jurors become more informed about a policy problem, they can contribute to a public-issues deliberation and solution, regardless of their expertise or lack of it. As an artifact, the citizen jury *allows* and *encourages* non-experts to participate equally in the technology development process. To allow for an informed deliberation, the process of knowledge acquisition and learning is crucial within the process, preceding the deliberation process (Ada Lovelace Institute 2021b; DemocracyNext 2023b; Kennedy et al. 2022), and a substantial portion of the jury process is dedicated to it (one-third in our case). To be able to make informed recommendations, it is *requested* that the participants are able to learn about the issue at hand. In that sense, it both *encourages* knowledge-making among the participants and *requests* reflexive practices. What the citizen jury achieves in this way is that, first, it challenges the ingrained assumption that citizens do not care or cannot grasp and engage with complex, expert topics (Ada Lovelace Institute 2021b), and, second, it overcomes known obstacles such as a lack of information, expert vocabulary and/or legalize, as well as time, to get involved in a meaningful way.

The rigidity of the citizen jury's participation structure is evident in how the knowledge and deliberation processes are prescribed. To facilitate knowledge creation, the process requires the involvement of first, expert witnesses, and then information kits. By engaging expert witnesses – members of the expert community who should provide the most relevant information in a balanced manner, focusing on the benefits and risks of using AI proxies and offering diverse perspectives, we promoted an ongoing knowledge-sharing process for our participants. This was further supported by providing an *information kit*, a take-home folder containing relevant sources tailored to the participants' non-expert, non-legal vocabulary. We will revisit the selection and contents of this folder in the "bad and the ugly" section, as it is a complex and somewhat arbitrary task action.

The citizen jury is demanding in one more aspect – as we discussed in Section 2, to ensure and guarantee neutrality and objectivity, an external organization must oversee the execution of the process itself, including recruitment, sortition, facilitation, and the selection of expert witnesses. The role of the organization specialized in citizen participation is to support the deliberative participatory processes, both in terms of organizational aspects and regarding the content and deliberation aspects (e.g., how the sessions will be structured, how the deliberation will take place, how consensus will be reached, and similar).

4.1 Epistemic justice and effective performativity affordances

As summarized in Figure 1, citizen juries afford several mechanisms: demanding, encouraging, and refusing particular forms of engagement, which differentiate them from dominant vertical ethics approaches. These affordance mechanisms align with the requirements for *epistemic justice* (Fricker 2007) and *effective performativity* (Kerr et al. 2020).

Citizen juries actively demand the inclusion of diverse, non-expert participants and encourage structured learning to foster informed deliberation. These design features directly target testimonial injustice by valuing the lived knowledge of intersectionally situated individuals, and reduce

hermeneutical injustice by creating spaces for interpretation and articulation of experience. By insisting that 'everyone comes from somewhere', acknowledging and encouraging situated knowledge (Haraway 1988), this "breathing space" (Ruckenstein 2024) offers an architecture for critical engagement, allowing participants to reflect on and intervene in imaginaries of technological futures. In this sense, the citizen jury exceeds mere generic performativity. It enacts effective performativity by shifting epistemic authority, and it repositions ethics as world-making rather than risk-mitigating activity.

While institutionalized AI ethics often produce abstract, non-binding guidelines detached from affected publics, citizen jury builds from the ground up. It demands knowledge co-production, enables reflexivity among both participants and researchers, and foregrounds ethical pluralism over universality. In doing so, it offers a counter-model to vertical ethics – one that holds potential for structural transformation rather than merely performative inclusion.

5. The bad and the ugly

In the next sections, we will reflect on the critical aspects that emerge from the application of citizen juries in AI ethics research and elaborate on the potential critical points that risk reducing citizen juries to another method for misusing participatory research and AI ethics.

While citizen juries encourage, allow, and request several actions, they also have limitations and challenges. It must be emphasized that the entire process of organizing and executing a citizen jury is lengthy and resource-demanding in terms of time, finances, and engagement of both jurors, researchers and other involved parties (see Fig. 1). Compared with other research and assessment methods, it requires significantly more procedural preparation regarding research and workshop materials. Each session requires six to eight hours of activities, for at least three to five days. This also requires dedication and commitment from the participants and researchers. Although we designed the sessions in such a way that they would interfere the least with the daily activities and obligations of our participants (sessions were scheduled during the weekends) and we financially compensated them for their time, ensuring ongoing participation proved challenging, and the attrition rate (as is often the case with citizen juries and other participatory methods) was high.

5.1 Participatory atrophy

The challenges of ensuring ongoing participation are well known in social science research. Although all recruited participants are initially highly interested, invested, and engaged, the participation rate tends to decline over time. This is especially true for processes that involve longer engagement periods, both in terms of time and session duration. Each of the three citizen jury sessions lasted for eight hours. That required dedication from our participants, both in time and in brain power. Although we provided financial remuneration and ensured that sessions were scheduled during the weekend, participation declined in subsequent sessions. In the first session, out of the 20 selected participants, only 14 attended. In the second session, that number dropped to 12 participants. We finished the last session with nine participants.

5.2 Participation bias

At the level of architecture, the citizen jury *demands* demographic representation via sortition and *encourages* deliberative, informed decision-making. However, this promise of inclusivity is undermined by the way our jury *allows* participation: it relies on a self-selecting pool of citizens who are already predisposed or motivated to engage. In our case, recruitment occurred through online and street campaigns, yielding 100 potential jurors. From this pool, 20 were selected via a stratified sample, using a sortition algorithm.

This design *refuses* engagement from those who do not or cannot self-identify as interested participants. The result is a participatory bias rooted not in bad faith, but in the recruitment method. French and Laver (2009) refer to this as jurors being "biased towards participation" (5). As our jurors reported, their motivations included prior interest in AI or civic processes, which may have biased the process toward already engaged publics. Such self-selection is not unique to citizen juries. It is a recurring limitation in participatory methods, from surveys to focus groups. However, in AI ethics – where participation is often invoked to legitimize technological trajectories – the line between genuine deliberation and *participation washing* becomes dangerously thin. As Birhane et al. (2022) note, participatory language can obscure exclusion when poorly executed or co-opted to legitimize predefined aims.

5.3 Concerned publics and problem-posing bias

Each citizen jury process starts with an emergent or urgent issue that requires engagement from the public and public deliberation. Usually, this is an issue around which a concerned public has been formed, something has been recognized as an issue that affects the citizenry and members of the public come together and form a concerned and relevant public (Hildebrandt and Gutwirth 2007). However, in our case, as a specific application of citizen jury for AI ethics and technology assessment, the relevant public does not exist; it does not emerge organically or predate the jury process; rather, we, as researchers, are artificially constructing it. In that sense, the issue to be tackled does not exist yet. This could be problematic in two ways. First, it could reinforce the corporate paradigm of techno-solutionism and of finding problems to which technical solutions are offered, or of offering technical solutions to purely societal problems. Second, this could be problematic in terms of *issue framing*. Both bring the risks of reinforcing power asymmetries.

In academic research settings, the issue framing is usually established top-down. Something has been recognized as an issue requiring investigation, and the problem statement emerges from there. This process of problem formulation is both crucial and critical because, as Barabas et al. (2020) outline, "normative and epistemological issues" arise during this process (167). We consider this a critical issue since it might lead to biased framing (French and Laver 2009, 15) where the decision of what will be recognized as an issue, what is considered as posing a problem, does not originate from real experiences and needs of the people on the ground but is decided elsewhere. However, if we want a truly participatory process that adheres to the tenets of epistemic justice in particular, we believe that the issue framing and problem-posing should originate with the concerned public, particularly those who will be most affected. This is what Freire (2017) calls "problem-posing pedagogy" – individuals are guided through dialogue, not explanation, to single out "objects of their consideration" (56) that would lead to critical awareness and potentially to a critical intervention or action as a praxis (56). The aim is for people to develop their own power and capacities to perceive critically "*the way they exist* in the world *with which* and in which they find themselves" (Freire 2017, 56), and to identify and raise their own, genuine questions of concern. This approach acknowledges the individual contextual reality (Freire 2017, 77). People know things, and their knowledge is situated – emerging from their experiences, contexts and relations with others and the world around them, including with technology. Building further on Freire's pedagogy of the oppressed, Markham (2019) develops her critical pedagogy approach that acknowledges and nurtures the validity of non-academic knowledge, lived experiences, and practice-based knowledge (Pop Stefanija and Pierson 2024). While individuals are guided by experienced pedagogues (facilitators in our case), the process of recognizing the issues at hand should come from the participants themselves because problem-solving is most effective based on personal experience. This should be reinforced through a knowledge-making process, and here the role of the facilitators in these processes is crucial. They should "dialogue" (Freire 2017) and guide but not impose; they should support but not steer. Knowledge-based deliberation, especially when it comes to AI technology, is essential, but several critical points warrant discussion.

5.4 Delegated epistemic authority and the affordances of expert framing

Our decision to use a citizen jury was grounded in the conviction that ethical deliberation depends on collective knowledge-making. Yet this process is shaped by those *allowed* to define and deliver knowledge: researchers, facilitators, and expert witnesses. Both researchers, experts and facilitators have a 'view from somewhere' (Haraway 1988). Despite our aim for neutrality, our own positionality influenced who was invited to speak, what information materials were distributed, and how deliberation was structured.

Following both the codified architecture of citizen jury and our dedication to ensure objectivity and neutrality in and of the process, we *delegated* much of the jury process to professional facilitators. This delegation *demanded* that they take control of the deliberative architecture, framing how knowledge would be presented and discussed. As Kennedy et al. (2022) argue, facilitators are not neutral; they shape the field of inquiry – "like the methods they deploy, are 'shaped by the social world' and 'help to shape that social world'" (Law et al. 2011, 2 in Kennedy et al. 2022, 397). While this ensured procedural impartiality, it *refused* our direct involvement in key framing decisions. Instead, we delegated authority to the facilitators regarding the decision of how the deliberation will be structured, guided, and take place.

Expert selection further illustrates a constraint in citizen jury's affordance mechanisms. The method *demands* expert input, and we invited two academic voices – a computer scientist and a legal scholar. While well-qualified, this choice discouraged epistemic pluralism by privileging credentialed knowledge over lived experience. Even more, "expertise is neither neutral nor innocent" (Kennedy et al. 2022, 397); experts are shaped by their disciplinary ideology, knowledge repositories and vocabulary, they possess (obfuscated) knowledge enmeshed in professional practices and traditions of thought and worldviews. Community actors, such as members of energy-sharing initiatives, were *allowed* in principle but excluded in practice due to logistical constraints and our assumptions about what counts as expertise.

This narrowing of epistemic input risks undermining the citizen jury's potential for testimonial and hermeneutical justice. What the jury *encouraged* – deliberation – was shaped by what it structurally *discouraged:* diverse epistemic sources. The architecture thus encoded a form of delegated epistemic authority that limits the scope of deliberation, even as it *demands* diversity. If citizen juries are to function as transformative ethical tools, their affordances must be critically examined – not only in what they demand from participants, but in what they allow, discourage, or render invisible through selection of expert witnesses and background knowledge.

5.5 Shrinking deliberation: TRL and project constraints

The architecture of our citizen jury was shaped not only by methodological choices but also by the structural affordances and constraints of the research project itself. One of the most significant limitations was the constrained field of possible deliberation: *refusing* certain lines of inquiry and *discouraging* others, due to the predetermined focus on AI and the system's low technology readiness level (TRL).

Early in the planning process, we instructed facilitators to *discourage* jurors from asking whether AI should be used at all to address energy distribution during brownout scenarios. This explicit framing decision was not neutral. It *refused* the problem-posing dimension central to participatory and emancipatory ethics. Although we recognized the legitimacy of such foundational questions, our positioning within a collaborative, AI-focused research project *demanded* a narrower focus. As one juror astutely observed in a post-session questionnaire: "Your process has a lot of bias on a lot of different levels. The main one is no debate about if AI is necessary, [or] desirable to use … at all." This might lead to co-optation of "the voice of participants to achieve predetermined aims" (Birhane et al. 2022, 2) and of "participation-washing" (Sloane et al. 2022).

This narrowing may take place by focusing (solely) on 'consensus building'. Yet, consensus is one of the key characteristics and aims of the citizen jury as a participatory methodology and process. However, as Büchner et al. (2025) outline, if we focus too much on consensus, we fail to take full advantage of the potential of the tensions, negotiations, conflicts, and dissent as valuable "opportunities for critical reflection, change, and improvement" (3). And while within our citizen jury design, we encouraged diverse perspectives and lived experiences throughout the process, we were still narrowing the field of deliberation by focusing, ultimately, on achieving consensus. And we see this in the outputs of our process: a list of guiding principles, values, and prioritized households. This might read as what Casiraghi (2023) calls the politicization of ethics – "what is concerning is the exclusion, from ethical debates, of people and groups interested in discussing ends rather than how to develop ethical frameworks and principles to achieve, ex-post, predefined ones" (Casiraghi 2023, 28).

This form of deliberative narrowing became especially evident in a high-stakes exercise in which jurors were asked to prioritize between two fictional household types during a brownout. Though selected at random, the exercise *demanded* jurors make difficult trade-offs based on thin contextual prompts. What began as a hypothetical task quickly turned into an emotionally and ethically complex deliberation, with jurors invoking solidarity, vulnerability, and equity as guiding principles. Yet, deciding who gets electricity and who does not surfaced and exposed the inherent tension between normative deliberation and technological abstraction. Participants recognized the limits of this design fiction: a system tasked with encoding such values would likely *refuse* nuance and exacerbate harm through binary choices. The final implicit reflection was that some ethical dilemmas might be too complex to resolve by AI systems.

The speculative nature of the technology itself amplified this insight. The proposed system was still in a conceptual phase; no dataset, algorithmic model, or deployment plan existed. While, on the one hand, this is a positive development since it engages participants in the early stages of development and thus counters a known controversy in participation, where participation mainly serves as "ex-post attempts to fix problems that AI applications cause" (Sloane 2024). On the other hand, this is narrowing the deliberative space by discouraging hermeneutical justice. Participants repeatedly asked detailed questions about how the system would function, what data it would use, and how decisions would be made, but we could not provide definitive answers. In such a low-TRL context, the citizen jury *allowed* only abstract engagement. Without lived experience to draw from, jurors relied on speculative imagination, supported by future scenarios, visual storytelling, and guided exercises. While these tools helped surface latent values and imaginaries, the process ultimately *discouraged* concrete recommendations. Instead of formal outputs like opinions or policy recommendations, we settled on co-created narratives and a list of guiding principles.

Ironically, this list resembled the kind of abstract principles typically generated in top-down ethics exercises (see Fig. 4). The only difference was its provenance: rather than being imposed by institutional actors, it emerged from a participatory process. Yet this raises a critical question: if citizen juries designed for epistemic justice and transformative ethics are reduced to producing symbolic artifacts due to project constraints, are we not at risk of participatory performativity? Or of technosolutionism? This is the risk we face if we only discursively aim for transformative AI ethics, but performatively end up selecting and integrating values to be 'embedded' in the technology design at the expense of neglecting and oversimplifying the fact that they emerge from complex societal issues. These societal issues are often seen as *too* complex and requiring broader sociopolitical engagement and actions, or are seen as too abstract to be addressed (Wang and Blok 2025), hence, offering what seems like easier (technological) solutions.

6. Discussion

The citizen jury is a promising tool that holds great potential for enabling representative and deliberative participation. When transposed to the field of technology development, especially in the

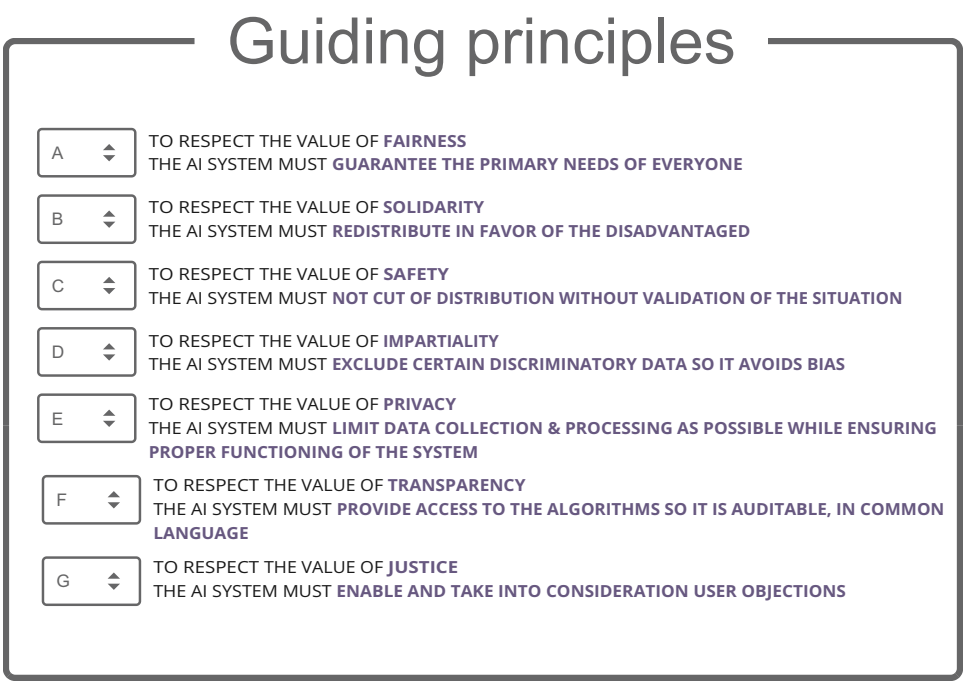

Figure 4. List of guiding principles.

pre-design and conceptual phases, it proves to be a useful tool. It overcomes several limitations and shortcomings related to the traditional participation methods like focus groups, interviews, and surveys, especially when the technology is not yet developed and remains in the abstract, hard-to-envision phase.

In that sense, we could discuss the affordances of the citizen jury as a technology assessment process during the pre-development stage in terms of horizontal empirical ethics. Citizen juries demand a deliberative, public, and representative process of decision-making. They not only allow but also encourage and actively request deliberation processes where the situated knowledge of affected communities is taken into consideration. Citizen juries request that the lived experiences, circumstances, and contexts of different 'belongings' are being acknowledged and accounted for. The situatedness of each participant is not a bug, but a feature of the process that allows us to capture the messy nature of reality and the varied ways in which a technology (even only an envisioned one) impacts them differently and to various degrees. In that sense, this application of the citizen jury is empowering and transformative. Empowering for the participants themselves, since it enables not only their inclusion in the process, but also knowledge-making opportunities. It should be empowering and transformative for the communities they belong to, too, since it advocates for the recognition of their very own needs and accounts for their situated contexts. The question remains whether this was achieved or if it is achievable at all for the latter.

That brings us to the question of performativity vs transformativity of AI ethics. As a horizontal, deliberative, and participatory AI ethics, it should. Citizen juries have the potential to transform the field of AI ethics as institutionalized, corporate, and co-opted into a process that takes place bottom-up, considers and acknowledges the situatedness and contextual knowledge of various affected communities. It also transforms and empowers the jurors themselves and potentially brings societal and institutional changes. However, there are still risks to be considered. Due to the specific way

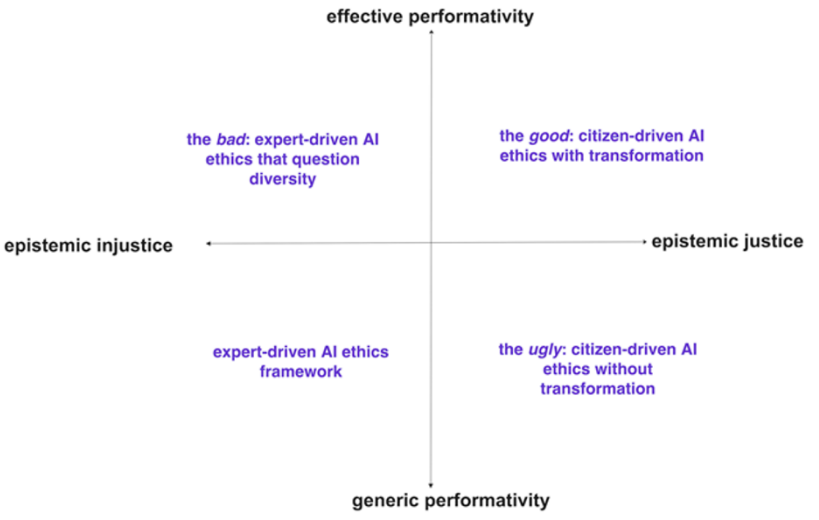

Figure 5. Evaluating and situating the citizen jury as empirical AI ethics.

in which our citizen jury process was framed and designed, we risked falling into the performativity trap – our jurors were instructed to produce a list of guiding principles for the speculative technology (Fig. 4). This is one of the characteristics of institutionalized AI ethics, focused on risk mitigation, but with low chances for implementation in practice. In that sense, the critical question emerges regarding what exactly the envisioned and possible outcome of the jurors' engagement is. Otherwise, we risk falling into the trap of participatory and ethical solutionism. If we revisit Figure 3 based on our reflexive assessment through the heuristic of the good, the bad and the ugly across the axes of epistemic justice/injustice and performativity, we can better situate the affordances of the citizen jury. Figure 5 illustrates how we critically assess both the emancipatory potential and the structural limitations of citizen juries.

7. Conclusion

This paper examined whether transformative empirical AI ethics can be achieved through the participatory method of a citizen jury. Our approach reframed ethical assessments and participatory methods as artifacts in their own right. This method's affordance mechanisms mediate participants', facilitators', and researchers' behavior, and shape the outcomes of deliberative processes. Using epistemic justice as an internal criterion, we examined whether citizen jury facilitates testimonial inclusion and reduces hermeneutical injustice. As an external measure, we employed the spectrum of generic vs effective performativity (Kerr et al. 2020) to assess whether citizen jury merely performs ethics symbolically or enables ethical transformation in technology development. Together, these two lenses – internal and external – allowed us to evaluate whether citizens participated and what kind of ethics their participation enabled.

The citizen jury method affords several key mechanisms that enable horizontal, participatory AI ethics: it demands inclusive deliberation, encourages the integration of experiential knowledge, and allows structured reflection and future imagination. However, these same mechanisms also afford certain limitations. They permit framing biases in guiding questions, allow narrow definitions of expertise, and constrain the scope of the jury's output. In our case, the technology's speculative status and the research project's institutional constraints discouraged foundational problem-posing, such as whether AI should be used at all, and ultimately refused lived experience as a form of expertise.

Our citizen jury operated within the epistemic and practical limitations of a low TRL. The technology's conceptual status prevented jurors from engaging it through lived experience, forcing them to rely on speculative imagination. While this surfaced rich value reflections, it also meant the jury's output leaned toward abstract principles – ironically reproducing the artifacts of vertical AI ethics that the citizen jury was meant to critique.

Nonetheless, the citizen jury as an empirical research method revealed critical insights: it illustrated the tensions between participatory methods and their institutional framing, and highlighted the conditions under which ethics can move from symbolic to structural. Repeating this process with a more mature technology might open different affordances that enable experiential grounding, expand the range of valid knowledge contributions, and encourage truly transformative deliberation. Citizen juries hold promise as empirical instruments for doing AI ethics otherwise. But realizing this potential depends on reflexively engaging not only with who participates, but also with how the method itself constrains what can be deliberated, known, imagined and decided.

Acknowledgements. The authors would like to thank, first and foremost, the Brussels residents who participated in the citizen jury. We are grateful for all their dedication, time, discussions, and willingness to learn and to share with us their concerns, ideas, and solutions. We would also like to acknowledge and thank Léa Rogliano (FARi Brussels) and Simone Casiraghi (LSTS, VUB) for their involvement in the organization, note-taking, and support during the citizen jury session. Finally, our gratitude to the anonymous reviewers and to the Guest Editors for their valuable comments and suggestions.

Funding statement. The authors disclosed receipt of the following financial support for the research and authorship of this article: the research was conducted as part of the project "COOMEP: Coordination mechanisms for the sharing of energy through proxies, from the user to general guidelines," funded by Innoviris. Brussels and the FARi-AI for the common good Institute (grant number CPC-FARI 2022).

Competing interests. The authors declared no potential conflicts of interest with respect to the research, authorship and/or publication of this article.

References

Ada Lovelace Institute. 2021a. "The Citizens' Biometrics Council." March 2021. https://www.adalovelaceinstitute.org/project/citizens-biometrics-council/.

Ada Lovelace Institute. 2021b. "How to Make a Citizens' Biometrics Council." Accessed April 1, 2021. https://www.adalovelaceinstitute.org/feature/how-to-make-citizens-biometrics-council/.

Barabas, C., J. B. R. Colin Doyle, and K. Dinakar. 2020. "Studying Up: Reorienting the Study of Algorithmic Fairness around Issues of Power."In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, 167–176. New York, NY, USA, FAT* '20, January 27. https://doi.org/10.1145/3351095.3372859.

Birhane, A., W. Isaac, V. Prabhakaran, M. Díaz, M. C. Elish, I. Gabriel, and S. Mohamed. 2022. "Power to the People? Opportunities and Challenges for Participatory AI." *Equity and Access in Algorithms, Mechanisms, and Optimization* 1–8. https://doi.org/10.1145/3551624.3555290.

Büchner, S., I. Zakharova, and J. Jarke. 2025. "Participation as Imposition: Problem Assessments and Recommendations for Enabling Participatory Research: Position Paper." Partizipation als zu mutung, April 21.

Carugati, F. 2020. "A Council of Citizens Should Regulate Algorithms." Wired. https://www.wired.com/story/opinion-a-council-of-citizens-should-regulate-algorithms/.

Casiraghi, S. 2023. "Anything New under the Sun? Insights from a History of Institutionalized AI Ethics." *Ethics and Information Technology* 25 (2): 1–14. https://doi.org/10.1007/s10676-023-09702-0.

Davis, J. L. 2020. *How Artifacts Afford: The Power and Politics of Everyday Things. Design Thinking, Design Theory*. Cambridge, MA, USA: MIT Press.

Davis, J. L., and J. B. Chouinard. 2017. "Theorizing Affordances: From Request to Refuse." *Bulletin of Science, Technology & Society* 36 (4): 241–248. https://doi.org/10.1177/0270467617714944.

Delgado, F., S. Yang, M. Madaio, and Q. Yang. 2023. "The Participatory Turn in AI Design: Theoretical Foundations and the Current State of Practice." EAAMO '23:Equity and Access in Algorithms, Mechanisms, and Optimization, Boston MA USA, October 30–November 01, 1–23. New York, NY, USA: ACM.

DemocracyNext. 2023a. "Assembling an Assembly Guide." DemocracyNext. https://assemblyguide.demnext.org/.

DemocracyNext. 2023b. "During the Assembly. Assembling an Assembly Guide." DemocracyNext. https://assemblyguide.demnext.org/during-the-assembly.

Freire, P. 2017. *Pedagogy of the Oppressed*. London, UK: PENGUIN.

French D and M. Laver. 2009."Participation Bias, Durable Opinion Shifts and Sabotage through Withdrawal in Citizens' Juries" *Political Studies* 57 (2): 422–450. https://doi.org/10.1111/j.1467-9248.2009.00785.x

Fricker, M. 2007. *Epistemic Injustice: Power and the Ethics of Knowing*. 1st ed. Oxford: Oxford University Press.

Gourlet, P., D. Ricci, and M. Crépel. 2024. "Reclaiming Artificial Intelligence Accounts: A Plea for A Participatory Turn in Artificial Intelligence Inquiries." *Big Data & Society* 11 (2): 1–21. https://doi.org/10.1177/20539517241248093.

Green, B. 2021. "The Contestation of Tech Ethics: A Sociotechnical Approach to Technology Ethics in Practice." *Journal of Social Computing* 2 (3): 209–225. https://doi.org/10.23919/JSC.2021.0018.

Hagendorff, T. 2020. "The Ethics of AI Ethics: An Evaluation of Guidelines." *Minds and Machines* 30 (1): 99–120. https://doi.org/10.1007/s11023-020-09517-8.

Haraway, D. 1988. "Situated Knowledges: The Science Question in Feminism and the Privilege of Partial Perspective." *Feminist Studies* 14 (3): 575–99. https://doi.org/10.2307/3178066.

Hildebrandt, M., and S. Gutwirth. 2007. "(Re)presentation: PTA Citizens' Juries and the Jury Trial." *Utrecht Law Review* 3 (1): 24–40. https://doi.org/10.18352/ulr.35.

Hildebrandt, M., and S. Gutwirth. 2008. "Public Proof in Courts and Jury Trials: Relevant for pTA Citizens' Juries?" *Science, Technology, & Human Values* 33 (5): 582–604. https://doi.org/10.1177/0162243907306701.

Information Commissioner's Office. 2019. "Project ExplAIn Interim Report." *ICO*. https://ico.org.uk/about-the-ico/research-and-reports/project-explain-interim-report/.

Jasanoff, S 2015. "Future Imperfect: Science, Technology, and the Imaginations of Modernity." In *Dreamscapes of Modernity: Sociotechnical Imaginaries and the Fabrication of Power*, edited by S. Jasanoff, 1–33. University of Chicago Press.

Kennedy, H., R. Steedman, and R. Jones. 2022. "Researching Public Trust in Datafication: Reflections on the Deliberative Citizen Jury as Method." In *New Perspectives in Critical Data Studies: The Ambivalences of Data Power*, edited by A. Hepp, J. Jarke, and L. Kramp, 391–414. Transforming Communications – Studies in Cross-Media Research. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-96180-0.

Kerr, A., M. Barry, and J. D. Kelleher. 2020. "Expectations of Artificial Intelligence and the Performativity of Ethics: Implications for Communication Governance." *Big Data & Society* 7 (1): 1–12. https://doi.org/10.1177/2053951720915939.

Law, J. 2004. *After Method: Mess in Social Science Research*. London: Routledge. https://doi.org/10.4324/9780203481141.

MacKenzie, D. A. 2008. *An Engine, Not a Camera: How Financial Models Shape Markets*. First MPI Press paperback ed. Inside Technology. Cambridge, MA: MIT Press.

Markham, A. N. 2019. "Critical Pedagogy as a Response to Datafication." *Qualitative Inquiry* 25 (8): 754–760. https://doi.org/10.1177/1077800418809470.

Marres, N., and C. Gerlitz. 2016. "Interface Methods: Renegotiating Relations between Digital Social Research, STS and Sociology." *Sociological Review* 64 (February): 21–46. https://doi.org/10.1111/1467-954X.12314.

McQuillan, D. 2018. "People's Councils for Ethical Machine Learning." *Social Media + Society* 4 (2): 1–10. https://doi.org/10.1177/2056305118768303.

Munn, L. 2022. "The Uselessness of AI Ethics." *AI and Ethics* 3: 869–877. https://doi.org/10.1007/s43681-022-00209-w.

Nathan, Lisa P., Predrag V. Klasnja, and Batya Friedman. 2007. "Value Scenarios: A Technique for Envisioning Systemic Effects of New Technologies." *CHI '07 Extended Abstracts on Human Factors in Computing Systems*. New York, NY, USA: CHI EA. https://doi.org/10.1145/1240866.1241046.

OECD. 2021. *Evaluation Guidelines for Representative Deliberative Processes*, 94. Paris: OECD Publishing. https://www.oecd.org/governance/innovative-citizen-participation/.

OECD. 2023. "OECD Deliberative Democracy Database (2023)." Airtable. Accessed June 26, 2025. https://airtable.com/appP4czQlAU1My2M3/shrX048tmQLl8yzdc/tblm3C6n7vM6vPSCz/viw2EEqp8wOhaHACk.

OECD Guidelines for Citizen Participation Processes. 2022. "OECD." Accessed September 23, 2022. https://www.oecd.org/en/publications/oecd-guidelines-for-citizen-participation-processes_f765caf6-en.html.

Patel, R., A. Peppin, and N. Alnemr. 2022. "The Rule of Trust: Findings from Citizens' Juries on the Good Governance of Data in Pandemics." Ada Lovelace Institute. https://www.adalovelaceinstitute.org/report/trust-data-governance-pandemics/.

Pop Stefanija, A., and J. Pierson. 2024. "Repairing What's Not Broken – Algorithm Repair Manual as Reflexivity Device." *Convergence: The International Journal of Research into New Media Technologies* 30 (6): 1871–1892. https://doi.org/10.1177/13548565241261978.

People Powered. 2021. How to Run a Citizens' Assembly. Department for Digital, Culture, Media and Sport (DCMS); Ministry of Housing, Communities, and Local Government (MHCLG). https://www.peoplepowered.org/resources-content/howtorunacitizensassembly.

Rességuier, A., and R. Rodrigues. 2020. "AI Ethics Should not Remain Toothless! A Call to Bring Back the Teeth of Ethics." *Big Data & Society* 7 (2): 1–5. https://doi.org/10.1177/2053951720942541.

Ruckenstein, M. 2024. "Collaborative Explorations as Breathing Spaces for Digital Futures." *Dialogues on Digital Society* 1 (2): 131–147. https://doi.org/10.1177/29768640241308332.

Sloane, M. 2024. "Controversies, Contradiction, and 'Participation' in AI." *Big Data & Society* 11 (1): 1–5. https://doi.org/10.1177/20539517241235862.

Sloane, M., E. Moss, O. Awomolo, and L. Forlano. 2022. "Participation Is Not a Design Fix for Machine Learning." In *Proceedings of the 2nd ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization*, 1–6. EAAMO'22. New York, NY, USA: Association for Computing Machinery. https://doi.org/10.1145/3551624.3555285.

Smith, G., and C. Wales. 2000. "Citizens' Juries and Deliberative Democracy." *Political Studies* 48 (1): 51–65. https://doi.org/10.1111/1467-9248.00250.

Street, J., K. Duszynski, S. Krawczyk, and A. Braunack-Mayer. 2014. "The Use of Citizens' Juries in Health Policy Decision-Making: A Systematic Review." *Social Science & Medicine* 109: 1–9. https://doi.org/10.1016/j.socscimed.2014.03.005. May.

Veer, S. N. V. D., L. Riste, S. Cheraghi-Sohi, D. L. Phipps, M. P. Tully, K. Bozentko, S. Atwood, A. Hubbard, C. Wiper, M. Oswald, and N. Peek. 2021. "Trading off Accuracy and Explainability in AI Decision-Making: Findings from 2 Citizens' Juries." *Journal of the American Medical Informatics Association* 28 (10): 2128–2138. https://doi.org/10.1093/jamia/ocab127.

Verbeek, P.-P. 2006. "Materializing Morality: Design Ethics and Technological Mediation." *Science, Technology, & Human Values* 31 (3): 361–380. https://doi.org/10.1177/0162243905285847.

Verbeek, P.-P., and D. Tijink. 2020. "Guidance Ethics Approach: An Ethical Dialogue about Technology with Perspective on Actions." ECP: Platform voor de Informatie Samenleving. https://begeleidingsethiek.nl/publicaties/guidance-ethics-approach/.

Wagner, Ben. 2018. "Ethics as an Escape from Regulation: From 'Ethics-Washing' to Ethics-Shopping?" In *Being Profiled: Cogitas Ergo Sum: 10 Years of Profiling the European Citizen*, edited by Emre Bayamlioğlu, Irina Baraliuc, Liisa Janssens, and Mireille Hildebrandt, 146. Amsterdam: Amsterdam University Press. https://doi.org/10.2307/j.ctvhrd092.18.

Wang, H., and V. Blok. 2025. "Why Putting Artificial Intelligence Ethics into Practice Is Not Enough: Towards a Multi-Level Framework." *Big Data & Society* 12 (2): 1–14. https://doi.org/10.1177/20539517251340620.

Wells, R., C. Howarth, and L. I. Brand-Correa. 2021. "Are Citizen Juries and Assemblies on Climate Change Driving Democratic Climate Policymaking? an Exploration of Two Case Studies in the UK." *Climatic Change* 168 (1–2): 1–22. https://doi.org/10.1007/s10584-021-03218-6.

Wong, Y. N., R. Jones, R. Das, and P. Jackson. 2023. "Conditional Trust: Citizens' Council on Data-Driven Media Personalisation and Public Expectations of Transparency and Accountability." *Big Data & Society* 10 (2): 1–13. https://doi.org/10.1177/20539517231184892.

Cite this article: Pop Stefanija, A. and Heyman, R. 2026. Deliberating the algorithmic future: Reconfiguring AI ethics through citizen juries. *Cambridge Forum on AI: Culture and Society*, **2**, e12, 1–20. https://doi.org/10.1017/cfc.2026.10026