CITIZEN SCIENCE:
THEORY AND PRACTICE

Collaborative Policy Development for Artificial Intelligence Ethics: A Case Study with the Zooniverse Platform

CASE STUDIES: APPLICATION AND INTEGRATION

ubiquity press

SAMANTHA BLICKHAN
HILLARY BURGESS
KAYLENE CASWELL STOCKING
CHAD HARPER
LUCY FORTSON
TANISHA TATE WOODSON
LEA WITKOWSKY
L. CLIFTON JOHNSON
LAURA TROUILLE

*Author affiliations can be found in the back matter of this article

ABSTRACT

As the use of artificial intelligence and machine learning (AI/ML) has become more prevalent in digital participatory research (DPR), the community is reckoning with increasing attention to the ethical, social, and technical implications of automation, and its impacts on research outputs and volunteer experience. This article details the process and outcome of collaboratively developing platform-level policy for ethical AI/ML-engaged DPR, using Zooniverse as a case study. Zooniverse has engaged more than 3 million volunteers worldwide in 500+ participatory research projects, each led by a different research team. As the use of AI/ML has increased—now in ⅓ of Zooniverse projects—it has sparked a range of reactions within the volunteer community, reflecting broader societal discourse. Volunteers have surfaced concerns and insights on issues like ownership, agency, transparency, and trust. In response, Zooniverse launched an initiative to develop recommendations for running AI/ML-engaged projects. Based around a virtual workshop series focused on the application of AI/ML in DPR, the process convened stakeholders from across the community: platform leadership, project teams, volunteers, ethicists, and experts in related topics. Workshops were preceded by community surveys to gather input from volunteers who did not participate in workshops, in order to ensure discussion agendas were shaped by input from the community. Here, we share the resulting Zooniverse AI ethics framework, collaboratively developed with community input, and which will continue to be refined through an iterative improvement process. We hope this case study provides an example to other platforms and researchers working at the intersection of public participation and AI/ML who wish to address AI/ML integration in their DPR, and incorporate community input into policy development.

CORRESPONDING AUTHOR:
Samantha Blickhan
The Adler Planetarium, US
samantha@zooniverse.org

KEYWORDS:
machine learning; ethics; governance; citizen science platforms; co-creation

TO CITE THIS ARTICLE:
Blickhan, S., Burgess, H., Stocking, K.C., Harper, C., Fortson, L., Woodson, T.T., Witkowsky, L., Johnson, L.C. and Trouille, L. 2026. Collaborative Policy Development for Artificial Intelligence Ethics: A Case Study with the Zooniverse Platform. *Citizen Science: Theory and Practice,* 11(1): 54, pp. 1–14. DOI: https://doi.org/10.5334/cstp.1052

INTRODUCTION: BACKGROUND AND TERMINOLOGY

Consistent with recommendations that vocabulary should be carefully chosen and clearly explained (Eitzel et al. 2017; Ellwood et al. 2023), we use the term "Digital Participatory Research" (DPR) to refer to online projects that involve members of the public as research contributors. This includes work traditionally labeled as online citizen science, but encompasses research disciplines outside the sciences, such as the humanities. We use "AI/ML" as a catch-all, encompassing artificial intelligence (AI: "computers that can emulate human thought") and machine learning (ML: "algorithmic tools and technologies that learn from existing data to solve tasks") (Fortson et al. 2024).

DPR is facilitated through platforms designed to support the activities and benefits of this work (Liu et al. 2021). The growth and popularity of these platforms has coincided with advancement of techniques for processing and analyzing large datasets. These include the coupling of machine-human systems, in which tasks are distributed among researchers, volunteers, and increasingly complex algorithmic processes (Trouille, Lintott, and Fortson 2019; Ceccaroni et al. 2019; Franzen et al. 2021; Fortson et al. 2024; Abdul-Rahman, Zwitter, and Haleem 2025). The tasks of the latter are frequently categorized based on their relationship to the machine component of the effort: (1) assisting or replacing humans in task completion, (2) influencing human behavior and fostering engagement, and (3) improving human insight through pattern identification and decision support (Ceccaroni et al. 2019; Abdul-Rahman, Zwitter, and Haleem 2025). The methods and techniques that underlie these scenarios are varied but fall under the AI/ML umbrella.

Use of AI/ML in DPR is not new (Banerji et al. 2010; Smith et al. 2011; see literature review in Fortson 2021), but it is growing. Authors report benefits such as acceleration of research and automated data validation, while noting challenges including loss of public trust and introduction of ethical questions alongside changes in educational and engagement opportunities (Franzen et al. 2021; Lotfian, Ingensand, and Brovelli 2021; Fortson et al. 2024; Pankiv and Kloetzer 2024).

Outside of DPR, the application of AI/ML to high-risk areas of scientific research in the 2010s (and other domains like healthcare and employment) led to a rise in scholarship examining extant or potential bias and risk, as well as frameworks for adapting existing methods of critical interpretation and intersectionality in the face of new and powerful technologies (Buolamwini and Gebru 2018; Bender et al. 2021; Ferrara 2023; Klein and D'Ignazio 2024). More recently, LLM-powered generative AI (genAI) tools like ChatGPT have seen widespread adoption outside the research community and integration into daily lives through internet and smartphone usage. This ubiquity has shifted the landscape and discourse around the terminology of AI/ML, the technologies it represents, the expansion of risks and challenges, and the urgency for development and adoption of intentional frameworks for integration with DPR. A global survey of more than 48,000 people across 47 countries found that while usage has grown significantly since 2022, distrust in the broad concept of AI and concern about its societal effects have increased in parallel (Gillespie et al. 2025).

While there has been significant focus on data ethics toolkits and data governance frameworks for DPR and citizen science (ECSA 2015; Cooper et al. 2022; Groot and Abma 2022), less attention has been given to the ethical aspects of developing AI/ML models using DPR-produced data, consideration for how humans and machines interact in projects, and/or how the interests of volunteer and research communities can remain aligned in this new context. The increasing role of AI/ML in mediating the relationship between researchers, volunteers, and data raises ethical questions that existing frameworks are not fully equipped to address. As AI/ML systems take on functions such as routing tasks based on inferred skill, scaffolding volunteer learning, or determining when contributions are sufficient (e.g., Østerlund et al. 2024), they move from being the object of volunteer labor to agents that structure it. This shift complicates the epistemic relationship between volunteers and the research process: Volunteers may not know that a machine is assessing their competence or shaping what they are allowed to contribute, nor can they typically scrutinize or contest those judgments. Without transparency about how AI/ML systems are created and applied in DPR, the conditions emerge for the alienation and disengagement that Ceccaroni et al. (2019) warn could follow when volunteers feel their contributions are not properly acknowledged or understood.

This potential for imbalance in power between machines, researchers, volunteers, and project or platform owners in questions of AI/ML systems mirrors longstanding conversations in the DPR field around project and platform design, research question development, and at what stage in the process volunteers are invited to contribute (Shirk et al. 2012). This project draws inspiration from community co-creation and participatory action research methods in its aims to involve all Zooniverse stakeholders, including volunteers, in the governance process, through ensuring that platform policy reflects the needs, interests, and concerns of the volunteers who are helping to create data.

ZOONIVERSE AS A CASE STUDY

Zooniverse (https://www.zooniverse.org) is the world's largest DPR platform, with more than 500 projects

launched since 2009, and more than 3 million registered volunteers at the time of writing. Zooniverse provides a space where researchers create opportunities for meaningful participation by volunteers who are largely intrinsically motivated; in particular by the feeling that they are contributing to "real" research across disciplines (Dowthwaite et al. 2025). Since the early 2010s, we have been aware of the promise of Zooniverse-produced data as a tool for training models for effective data processing (Banerji et al. 2010; Smith et al. 2011). The number of Zooniverse teams running AI/ML-engaged projects has since increased significantly, with a third of the projects on the platform now including some form of AI/ML.

Applications of AI/ML on Zooniverse range from using machine learning to improve processes for astronomical discovery (House et al. 2023), to speeding up archival digitization efforts for museum specimen labels (Guralnick et al. 2024), to improving species identification in camera trap images (Huebner, Palmer, and Packer 2024), to generating data for training and testing natural language models (Piper, Xu, and Ruths 2024), and fine-tuning LLMs for domain-specific caption generation in historical artworks (Reshetnikov and Marinescu 2025). While success stories of combined human and machine classification efforts in DPR are becoming abundant, examples of projects that effectively justify their methods to an increasingly concerned public are rarer. Rapid technological acceleration and the desire to produce cutting-edge research can lead to insufficient time allocated for teams to pause and consider the risks, challenges, and broader ethical questions of incorporating AI/ML methods into DPR.

Reactions to the application of AI/ML across Zooniverse vary, reflecting broader societal discourse about these technologies. Pre-2022, the Zooniverse community was largely approving of these practices, though the language used to describe them at the time was crucially focused around "machine learning" rather than the current popular practice of using "AI" as an umbrella term for automated processes (Gal et al. 2022). Post-2022, public conversations on Zooniverse's built-in Talk message boards (as well as private emails sent to the Zooniverse) became more mixed, surfacing issues like ownership, agency, transparency, and trust.

We saw this increase in volunteer debate as a sign that deeper discussion was necessary—across our own community and with AI/ML and ethics expert input—and received funding from The Kavli Foundation to support our work. Here, we present our framework for AI/ML-engaged DPR on the Zooniverse platform, which provides new guidance for project teams and describes how it will be incorporated into existing platform policy and practice. We provide an overview of the project history and context, methods, key learnings and recommendations, and discussion of the ongoing challenges of incorporating AI/ML technologies alongside participatory research practices in an ethical, equitable way that reflects the diverse needs of our communities. We hope that sharing the process and lessons learned alongside the project outcomes will benefit other practitioners working at the intersection of DPR and AI/ML, as well as non-DPR research communities who want to create collaborative and equitable ethics frameworks.

THE CONTEXT

Zooniverse's structure as a platform offers an ideal case study. It is a marketplace in which people "do" DPR across disciplines and so represents a broad range of technical requirements (tools, infrastructure) as well as differing methodological approaches.

When a research team wants to launch a public project on Zooniverse, they first develop their project via the do-it-yourself Project Builder (https://www.zooniverse.org/lab). Once the project is built, the team must apply for review, which takes place in two stages: first, an internal review by the Zooniverse team, and second, a beta review wherein testers from the Zooniverse volunteer community try a project and provide feedback.

AI/ML discussion during project review has typically included asking whether the team had explored automating the main goals of their project, and if not, to explain why. Research teams that used AI/ML were expected to explain their methods within the project About page, but this was in the context of providing background information about their project and its goals, and lacked a standard approach for how, or why, to share AI/ML-specific information. The purpose was to prepare research teams to answer the question of "Why can't this task just be automated?" (e.g., justifying the need for human effort, in line with the core ethical tenet from the days of the original Galaxy Zoo project of not wasting volunteers' time). This approach framed many early Zooniverse platform and infrastructure development decisions and even included an explicit statement that volunteers should "not be asked to undertake tasks that could be better undertaken by a computer" (Prestopnik and Crowston 2012).

While it is still essential for DPR project leads to make the best use of volunteer labor, the conversation is shifting, particularly around the purpose of DPR in a time when AI/ML has significantly improved. The assumption that if something *can* be automated then it *should* be assumes that the output and impact of DPR is limited to the data processed and produced. We know that this thinking

overlooks the additional outcomes and impacts of DPR, including serendipitous discovery (Cardamone et al. 2009), participant learning and skill development (Prather et al. 2013; Masters et al. 2016; Aristeidou, Scanlon, and Sharples 2020; Gunson, Murphy, and Brown 2021), and the ways participation can shape attitudes, identities, and forms of civic or community engagement (Ballard, Harris, and Dixon 2018; Van Oudheusden and Abe 2021; Skarlatidou et al. 2024; Phillips, Hebbard, and Karl 2025).

To address the risks, opportunities, challenges, and broader ethical questions that arise as AI/ML becomes increasingly integrated into DPR, we sought the input of experts and stakeholders, in particular from our community of more than 3 million platform volunteers. We used the information and insights gained through a participatory co-development process to create a set of recommendations for researchers using AI/ML methods in DPR, which launched publicly on September 9, 2026. These recommendations have been implemented for the Zooniverse platform, but we believe that they are generally applicable for DPR efforts outside the Zooniverse, as well as in participatory research and other public-facing research spaces more broadly.

THE PROCESS

We designed a multi-layered participatory process that included a variety of entry points for stakeholder engagement, with the intent of prioritizing input from our volunteer community about the ways that AI/ML are used on Zooniverse. We sought to balance breadth of input (ensuring diversity of voices) with discussion-based analysis (enabling the translation of complex topics into actionable recommendations). This was achieved through a nested structure of platform-wide surveys to shape workshop agendas, topic-focused workshops, and smaller working sessions.

We iteratively gathered input through discussions that included project teams (including Zooniverse and non-Zooniverse projects from across disciplines); volunteers (including Zooniverse and other platforms); Zooniverse leadership; doctoral and postdoctoral fellows from Berkeley's Kavli Center for Ethics, Science, and the Public; experts and stakeholders in computer science, data licensing, science communication, and data management; and an external evaluator and facilitator. Engagement points and feedback cycles centered on four thematic streams: 1) transparency and communication best practices: developing guidelines to support researchers in effectively communicating about their AI/ML work; 2) ethical approaches to AI/ML: identifying foundational elements of an ethical approach to ML-engaged DPR; 3) deepening contextual understanding: examining disciplinary differences, task type affordances, and varied stakeholder needs; and 4) downstream data protection: examining licensing frameworks and other tools to balance open data values with concerns about misuse.

Information flowed through each stream as follows (Figure 1):

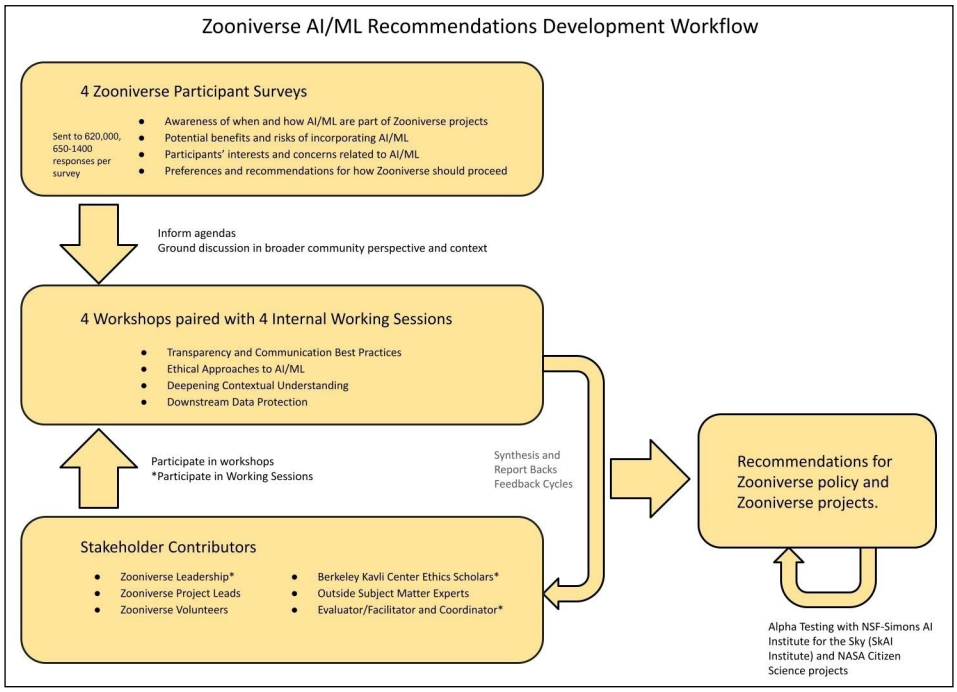

Figure 1 Recommendations development workflow with input cycles and synthesis steps.

- Platform-wide survey: Prior to each workshop, a survey was distributed via the Zooniverse global email list to gather input from the broader volunteer community on the workshop topic. Survey findings were used to shape workshop discussion agendas, to ground discussions in authentic community perspectives.
- Workshop: Each 2.5-hour virtual workshop brought together approximately 15 participants for structured discussion using breakout groups, plenary sessions, and collaborative tools.
- Working session: Following each workshop, a smaller group of approximately seven participants convened for a 2-hour working session to delve deeper into themes and begin translating discussion into specific recommendations.
- Synthesis and report-back: The project leads and Kavli Center Ethics fellows synthesized workshop and working session outputs into summary documents with draft recommendations, which were shared with all workshop participants for review.
- Member-checking: Workshop participants were given a specified timeframe to review report-back documents and provide feedback, corrections, and additions. This feedback was incorporated into refined recommendations.

In small-group working sessions, we discussed key workshop takeaways within and across the four themes to generate draft recommendations for 1) the governance of the Zooniverse platform and 2) project teams/researchers leading projects on the platform. These draft recommendations were reviewed and prioritized by Zooniverse leadership, identifying which could be implemented immediately, and which would require more information or resources to advance.

Recommendations for project teams were tested with projects representing the NASA-Zooniverse project community as well as the NSF-Simons-funded AI Institute for the Sky (SkAI Institute). We used feedback from these testing sessions to further refine the language in the recommendations for clarity and usability.

KEY LEARNINGS

VOLUNTEER SURVEYS

Through the pre-workshop surveys, we learned that members of the Zooniverse community: 1) have varied and sometimes strong feelings about AI/ML; 2) hold varying amounts of knowledge about what AI/ML are, how these technologies work, and their relationship to Zooniverse; and 3) are chiefly concerned with how the use of AI/ML could negatively impact data quality, though additional concerns include environmental impacts, copyright, privacy issues, and for-profit use of volunteer-generated data. Volunteers also expressed concern about the potential for datasets and outputs to be used later for unintended, and even untoward, purposes. Respondents explored the tradeoffs between the principles of open data—optimize for transparency, reuse, innovation, and data protection—with the goal to reduce risk of future undesired uses. If given this binary choice, most members of our community would err toward openness, which aligns with current Zooniverse policy. A recurrent theme was that not all AI/ML is the same (e.g., a ML algorithm versus genAI), and so we consider how different methods and uses come with different risks and benefits. Most respondents felt conservatively about volunteers using external AI tools to assist with completing DPR tasks. Very few respondents reported having used AI to complete tasks while classifying on Zooniverse projects.

Crucially, this information was not gathered as a qualitative study of participant feelings about, or understanding of, AI/ML, and the survey results are not the main focus of this paper. Participants were asked to take part in a voluntary survey in order "to help [the Zooniverse platform] understand participant perspectives on ML/AI and… ensure that the interests, priorities and concerns of Zooniverse project participants are reflected in recommendations for the platform and project leaders." (See Supplemental File 1: Appendix C: Survey preamble for the framing language used in surveys). Survey participants were not subjects of study, but contributors to policy development, providing critical feedback on the current successes and failures of platform- and project-level communications about AI/ML, and sharing what focus areas they would like to see reflected in a Zooniverse AI/ML policy.

WORKSHOPS

In our virtual workshops, rich discussions surfaced nuanced considerations and tradeoffs. Workshop participants were generally supportive of intentional, transparent uses of AI/ML to enhance research or outcomes, except to manipulate volunteer behavior without their consent or awareness. The relationship between responsibility, literacy, transparency, and agency were continuously highlighted. Through group discussion, participants determined that Zooniverse as a platform is responsible for developing and upholding the AI/ML ethics framework, research leads are responsible for interpreting and implementing this framework within their projects, and volunteers are empowered to confirm or question the implementation by voicing feedback or choosing not to participate. Project-specific information must be accessible and interpretable for volunteers to make

informed choices, and while the platform is not responsible for direct individual literacy about AI/ML, research leads can look to Zooniverse to provide support and examples for how best to communicate technical details.

Where possible, we translated workshop discussion outcomes directly into recommendations and actions. In other cases, complexities and tensions surfaced that could not be resolved in a single workshop. These examples highlighted the need for ongoing research, discussion, and iterative development of recommendations and resources at both the platform and project levels. We describe the resulting recommendations, as well as unresolved tensions, in the following sections.

RECOMMENDATIONS

The goal of our project was to develop community-informed recommendations for research teams running Zooniverse projects with AI/ML components. During the process, we determined key thematic principles for this work. We also identified critical governance commitments for the Zooniverse platform that would be necessary for the success of the project-facing recommendations. Both are described below and included in the public framework.

Key principles

Several key principles emerged from the surveys and workshops. They serve as a framework and summarize overarching goals that, if adhered to, will help to address the risks and concerns raised by participants.

1. Build **transparency** into AI/ML-related platform governance, project development, and delivery.
2. Design for volunteer experience and prioritize **agency and consent** for volunteers in the act of selecting projects; consider **benefits to** and **motivations of** volunteers in AI/ML-involved projects.
3. Facilitate **accountability** by demonstrating adherence to clear guidelines as a prerequisite for launching an official Zooniverse project.
4. Commit to ongoing dialogue and project and platform **responsiveness** to evolving attitudes and conversations about AI/ML.
5. Utilize and align with **existing related AI/ML framing resources** and continue to engage with wider communities and stakeholders to **develop and adopt best practices** around AI/ML-engaged participatory research.

Internal actions

The Zooniverse team are implementing actions that facilitate or precede the recommendations for project teams and bring this new guidance into existing Zooniverse policy and practice (see https://www.zooniverse.org/about/ai-ethics#internal-actions). These include creating and implementing structures, guidance, and support for project teams to learn about and act on recommendations (e.g., adapting the internal project review process with AI/ML-specific review criteria); modeling implementation of the recommendations; and developing resources to support volunteers in making informed decisions about whether to participate in projects. The Zooniverse created a new webpage on the platform to host the recommendations and supplementary material: https://www.zooniverse.org/about/ai-ethics. The framework has also been incorporated into existing Zooniverse policy of internal project review, to ensure that teams who want to launch public projects are engaging with the recommendations. The team has committed to an annual review of the framework (including the recommendations), to ensure that the contents reflect the current needs of our communities as well as the state of AI/ML technology.

Recommendations for project teams

The below recommendations (in italics) launched on September 9, 2026. In this section, a brief discussion follows each, to tie back to the key principles and relevance to the broader DPR community. Additional discussion is provided at the end of the list, and in the following section.

1. Design your project with volunteer agency (choice) in mind. *This requires transparency and clear communication for volunteers to make informed decisions about whether they want to take part in your project. If you are incorporating AI/ML at any stage of your work (pre- or post-processing, human-in-the-loop classification systems, etc.), you need to provide specific details about how these processes work, and why they are necessary. For example, saying "we will use the resulting data to train a model" is insufficient detail.*

At the most basic level, volunteer agency is exercised in the decision whether to participate in a project. This choice should be made with access to relevant and understandable information. At a more granular level, a project may include separate workflows or features, some with AI/ML components and others without. Clear and transparent labeling of options supports individual choice.

2. Show participants where/when/why AI/ML is incorporated in your project. *Some workflows within a project may be AI/ML involved/assisted and others not. These should be clearly labeled in the Workflow Name field, or Workflow Description area, so volunteers can choose accordingly (if they have a preference). If you used AI/ML to pre-process your data, describe what your quality control metrics are (including if this is your reasoning for running a validation workflow on Zooniverse).*

Enabling volunteers to opt in or out of any activity is predicated on providing sufficient information for an

individual to make an informed choice. Integration with AI/ML is not a decision driver for all volunteers, but it is for some. Disclosure about the presence of AI/ML and the explicit roles that volunteers and machines play in furthering research goals allow prospective volunteers to assess projects against their own values and motivations.

3. Consider the consequences of optimization. *When considering "offloading" easy tasks to AI/ML and having humans do harder/more complex tasks, be mindful that more advanced workflows may narrow who is able/qualified to participate, unless there is an onramp. Training tasks may be needed to gain proficiency to do tasks that AI/ML cannot do. Any onramp process or training workflow should be clearly labeled.*

This recommendation operationalizes the need for transparency, learning opportunities, and clarity around the purpose and meaning of time spent completing tasks on the platform. Throughout our process, there was consensus across stakeholder groups that volunteer behavior should not be influenced by lack of information or misinformation. Clearly labeling training tasks ensures that skill-building activities are not mistaken for "real" research contributions.

4. Retain context. *For projects inviting volunteers to engage with or review machine-generated output, consider whether and how context can impact the task—including both the quality of data and the quality of the volunteer experience. For example, if a task asks volunteers to review the quality of machine transcriptions but only includes a portion of the document (e.g., a single word or phrase), including a link to the full image or document will ensure that volunteers can continue to interact with the source material and its meaning, if desired. This information can be included within a subject's metadata.*

This recommendation encourages teams to think about the importance of context in a DPR task. How much, and what kind of, information do volunteers need to enjoy and be successful in the task they are being asked to complete? Does incorporating AI/ML impact volunteers' access to this information or prevent them from having a satisfying experience? This recommendation encourages teams to consider project design but also acknowledges that volunteers come to DPR with varied motivations. These can include the desire to engage with subjects and source data beyond what is available or required for the task at hand. If incorporating AI/ML into a workflow reduces a volunteer's ability to seek out additional information about a subject, teams should consider how they might provide that information.

5. Follow up. *Provide feedback on the degree of success achieved as AI/ML models are trained. The method will be determined by the project team, and may include a Talk post, email newsletter, etc.*

For AI/ML–integrated projects, the research goal may be the development of models for a specific purpose. The reasoning and end purpose of this development should be clear, with regular communication about progress toward success. Project teams should communicate the success (or failure) of model development and report back to volunteer communities as they would about any other project output.

6. Create pathways for feedback. *Any project with an AI/ML component should have a route for volunteers to share their feedback with project team members. For example, teams may wish to create a dedicated Talk board to describe the use of AI/ML, respond to questions, and provide updates on any training or outputs.*

This recommendation addresses accountability, in particular the responsibility of project teams to respond to questions volunteers may have about AI/ML components of a project. Project-specific Talk boards provide a forum for a project's research team and volunteers to share updates, provide feedback, and ask and answer questions.

7. Address quality. *Describe how you are evaluating and addressing any potential impact on data quality within your project due to AI/ML usage.*

In our pre-workshop surveys, some Zooniverse volunteers raised concerns about the proliferation of bias and error in AI/ML, and the importance of human oversight in ensuring that the resulting data quality is high. Teams can address these concerns by providing detail around their quality control methods. For example, in a workflow where volunteers are asked to review machine output, project teams should explain why this output cannot be used on its own, and why they believe, or how they established, that human review will increase the output quality.

8. When disclosing your intended use of project results/data, consider how your intended use could shift over time. *Volunteers may choose to take part in your project based on the project goals and intended uses of results/data indicated in the About page. Later, what are the chances that you will expand those uses?*

This recommendation encourages project teams to be transparent with volunteers about potential (re)use of a project's results, whether data or models. For example, is the model you are training so granular that it will not be applicable outside the specific use case you are training it for? Or will it be more broadly reusable? If the latter, teams should consider what kind of reuse that may include, and how best to share it as part of their explanation of research. This can function as a potential metric for risk but could also help to provide a positive example to volunteers of the wider impact of their contributions.

9. Consider both the opportunities and potential risks of running an AI/ML-engaged project. *We encourage you to understand potential risks—alongside opportunities—before launching projects. Engaging with vetted tools and published scholarship may help develop this understanding, such as Ceccaroni et al. (2019), which presents a survey of risks and opportunities related to AI use in citizen science, and recommendations for risk mitigation.*

This recommendation focuses on accountability and responsibility. Even if experimenting with new or unfamiliar technologies, teams are responsible for the outcomes of their work, regardless of intent. Researcher literacy around both the benefits and risks of AI/ML is key to ensuring intentional design and decision-making.

CONNECTING TO BROADER ETHICAL FRAMEWORKS

This framework fills a necessary gap in the field by providing AI/ML-specific interpretations and extensions of best practices in DPR and citizen science writ-large (ECSA 2015; Resnik, Elliott, and Miller 2015; Cooper et al. 2022). Designing a project with intention and understanding the motivations of potential volunteers (Raddick et al. 2013; Jackson et al. 2015; Dowthwaite et al. 2025), addressing data quality (Kosmala et al. 2016; Burgess et al. 2017; Downs et al. 2021) and data governance (Anhalt-Depies et al. 2019; Göbel et al. 2019; Bowser et al. 2020; de Sherbinin et al. 2021), and committing to return of results (Iacovides et al. 2013; de Vries, Land-Zandstra and Smeets 2019; Wolf 2020) are all dimensions of designing and running a successful and ethical DPR project. The introduction of AI/ML can amplify the pitfalls of inattention to these dimensions, and policies should be explicit about how new technologies fit into existing governance models. This approach builds on expertise, existing principles, and guidance, with the participation of those who would use and be impacted by the outcomes, to develop and implement actionable policy. The incorporation of these recommendations into the Zooniverse project review process connects policy to practice—teams will be required to engage with these recommendations if they want their projects to publicly launch on Zooniverse. As one workshop participant noted, "If structured and communicated effectively, these recommendations can help to accelerate important research while enhancing the literacy of these technologies for [project volunteers]. Getting it wrong could potentially further erode the public's trust in research, including science and scientific institutions."

DISCUSSION

A number of shared concerns about the use of AI/ML in DPR projects emerged from the surveys and workshops. We believe some of these concerns can be substantially mitigated by adopting ethical principles such as the ones we propose. However, others are more difficult and reflect genuine ethical tensions that may not be fully resolvable. Ongoing and future work is critical for meaningful ethical engagement that goes beyond identifying easy points of consensus to engaging directly with the most difficult questions about what the integration of increasingly capable AI/ML into research means for public participation. In this section, we highlight some of these outstanding concerns, in the hope that future work can build from them.

VOLUNTEER DATA LIFECYCLE

We identified transparency about AI/ML usage in DPR projects as a key ethical principle. Researchers may not be aware of all possible uses for resulting data at a project's outset, especially uses that happen after data collection, such as during data analysis or training AI/ML models on DPR-produced labels. This complicates meaningful transparency: If projects are constrained by their initial stated uses of AI/ML it could harm the future research impact of the project and of volunteer contributions by extension. However, identifying new uses after project launch means that some volunteers may contribute to projects that use AI/ML in ways they would not have initially agreed to. This problem is exacerbated by the open sourcing of collected data after project completion, which is the current standard for Zooniverse projects as well as many DPR- and citizen science–specific ethical frameworks (Cooper et al. 2022). For example, after data is open sourced it could be used for any purpose, including for commercial and/or non-scientific training of AI/ML models.

One option would be to allow volunteers to withdraw their contributions from the dataset, or decline to have them open sourced, but this places a significant burden on DPR platforms and project leads and could interfere with reproducibility. Open access to data, especially from publicly funded research, is generally touted as a gold standard, except where sensitivities of intellectual property or privacy come into play, or where there is risk of harm (e.g., in some conservation projects that do not want to disclose locations of endangered species, as discussed in Anhalt-Depies et al. 2019). But AI/ML has changed the conversation. The scale of web scraping creates significant risk of monetization or other uses that do not align with the values of Zooniverse, a project team, or its volunteers. Current licensing structures fall short of providing meaningful protections because they rely on individuals acting in good faith for the benefit of the commons. New options are emerging like Creative Commons' CC Signals framework, which allows license holders to assert preferences around how datasets can be (re)used by machines (Hardinges, Pearson, and Ross 2025).

There is a critical need for continued conversation across DPR platforms and sub-disciplines as we face common questions brought about by the new scale and speed of technology.

This topic received significant attention during our process, with Workshop 4 dedicated entirely to Downstream Data Protection. While these tensions were not resolved by the end of this project, the discussions provided a strong basis for ongoing research and discussion within our community, and the Zooniverse leadership team will continue to engage with possible solutions going forward. Ultimately, we learned that addressing this emergent challenge is beyond the scope of Zooniverse alone. Gated data access has been proposed as one method of mitigation, that is, data are available but one must request access and agree to terms of use beforehand, but such an approach requires infrastructure and administration capacity that does not yet exist at scale.

POWER STRUCTURES

Project leads and project volunteers have different motivations for engaging with DPR, which can lead to tensions that may be exacerbated by AI/ML use. For example, using AI/ML to personalize the DPR task and make it more engaging might enhance participation rates but limit the scientific contribution a volunteer is able to make in a given amount of time. Establishing transparency about AI/ML usage means that potential volunteers have the power to withhold participation or choose to participate in other projects when they disagree with a project's stated or planned AI/ML use. Ongoing feedback and accountability mechanisms could give volunteers a more nuanced way to communicate concerns, but this places an additional burden on project leads. Furthermore, without a mechanism in place for volunteers to retract their submitted data, volunteers' power to walk away lasts only until the end of data collection. One potential area for more research is in the impact of AI/ML on models of participation and contribution, in particular around who has decision-making power and whether these approaches can strengthen or harm co-creation and power-sharing models.

WEIGHING ASSOCIATION FALLACY AGAINST JUSTIFIABLE CONCERN

Many DPR volunteers have ethical concerns about AI/ML more broadly, for example, because of its emerging effects on employment, human autonomy, and the environment. Media coverage that centers on concerns primarily with LLMs or genAI—with limited nuance when it comes to differential risks across ML applications (Chakravorti 2024)—may amplify this distrust. There are also concerns about the future of AI/ML, such as the risks associated with deploying increasingly powerful systems without appropriate safeguards, or the societal impacts, for example, on critical thinking skills due to cognitive offloading (Gerlich 2025). The terms AI and ML encompass a wide range of rapidly evolving practices and relative associated risks, but the language of differentiation, particularly for public understanding, does not yet exist.

Currently, AI/ML on Zooniverse is used in a way that does not directly entail many or all of these broader concerns. For example, volunteers may harbor anxieties shaped by genAI, which only maps loosely to most AI/ML use cases actually being implemented—typically, supervised classification models trained on volunteer-labeled data. Similarly, a project can train a small ML classifier on astrophysical images without interacting with the genAI ecosystem, and consuming about as much electricity as playing a computer game for a couple of hours. Because of this, it is tempting to try to "educate away" these broader AI/ML concerns by teaching volunteers about AI/ML on Zooniverse as distinct from what concerns them. However, it is well documented that education alone is ineffective at building public trust in scientists doing the right thing (Sturgis and Allum 2004; Wynne 2006; National Academies of Sciences, Engineering, and Medicine 2017). By adopting AI/ML, DPR is unavoidably part of a broader societal conversation about the appropriate usage and regulation of these new technologies: DPR does not exist in a vacuum. Our recommendations can help research teams navigate complex conversations, but engaging with broader concerns should be an anticipated part of choosing to incorporate these methods.

CONCLUSIONS AND NEXT STEPS

The Zooniverse platform has the greatest ability to guide and shape projects in the lead-up to launch through the project review process. While some mechanisms for support and accountability remain once a project is live or has concluded, these are more limited in scope and more difficult to enforce consistently. Implementation of these recommendations will be integral to a project's eligibility to launch. By making the framework public and accessible to volunteers, we envision a material shift in both expectations and practices around the use of AI/ML in Zooniverse projects, as well as increased trust through requesting and incorporating feedback during the process of developing policy. We recognize the ongoing conflicts between creating additional labor and/or barriers to entry

for project teams, and setting standards and processes that encourage intentional, ethical design alongside a positive experience for volunteers and a continued commitment to project and data quality.

These efforts are not occurring in isolation. Across citizen science and adjacent open science communities, there is growing attention to the ethical, social, and technical implications of AI/ML—from principles of data stewardship and protection (e.g., Collective Benefit, Authority to Control, Responsibility, and Ethics [CARE] Principles for Indigenous Data Governance, and Findable, Accessible, Interoperable, Reusable [FAIR] data framework) to emerging guidelines around responsible AI (e.g., the Organisation for Economic Co-operation and Development [OECD] AI Principles). Parallel work in in-person and community-based citizen science is also grappling with questions of consent, data governance, and the appropriate role of automation in shaping participant experience and contribution. Situating Zooniverse's recommendations within this broader landscape both reinforces their relevance and highlights the need for continued cross-platform and cross-disciplinary dialogue (e.g., European Citizen Science Association [ECSA] working groups).

Ultimately, these recommendations are intended not as a comprehensive solution, but as a starting point. While developed in the context of Zooniverse, many are applicable across DPR contexts and broader research efforts, in particular for researchers communicating with the public about their own AI/ML use, or with respect to many of the ongoing challenges included in the Discussion section. We created these recommendations as a response to the growing methodological uptake of AI/ML in DPR, the increasing awareness and concerns about the risk of different forms of AI/ML, and in direct response to questions raised by our volunteer community. We hope this work contributes to a wider effort to foster transparency, informed consent, collaboration, and shared understanding between researchers and volunteers that facilitates meaningful contributions to impactful research.

SUPPLEMENTARY FILE

The supplementary file for this article can be found as follows:

- **Supplemental File 1.** Appendix A: Project timeline; Appendix B: Workshop participant solicitation and selection; Appendix C: Survey preamble; Appendix D: Preparatory materials for workshops. DOI: https://doi.org/10.5334/cstp.1052.s1

ETHICS AND CONSENT

All participants (workshop attendees and survey respondents) were contributors to the process and not subjects of study. Survey responses were used for internal policy development and not for academic study, which is why they are not included here. The workshop participants listed in the Acknowledgements section are named with permission.

ACKNOWLEDGEMENTS

The authors would like to thank the Zooniverse community members who participated in surveys, and who gave us feedback during the process. We would also like to recognize the following people for their participation in the virtual workshops that led to the development of these recommendations:

Abigail Cecilia Zesati
Andrea Grover
Andrew Piper
Antonio Pasqua
Bruce Catter
Caroline Nickerson
Carrie Seltzer
Chris Lintott
Damian Sheils
Gwen Shafer
Harman Kaur
Harry Smith
Hasret Balcioglu
Hayley Roberts
Julia Brown
Julia Parrish
Katina Michael
Kirstie Whitaker
Lea Shanley
Lindsay House
Liz Dowthwaite
Lu Cheng
Lynn Arneill-Brown
Marisa Ponti
Mark Basham
Muireann Nic Corcrain
Michael Corey
Monica Granados
Pen-Yuan Hsing
Peter Mason
Ramana Sankar

Rob Guralnick
Rosemary Johnson
Sadie Coffin
Sallie Taylerson
Sandy Harris
Sara Brumfield
Sarah Kirn

FUNDING INFORMATION

This work is supported by The Kavli Foundation, award number SS-2025-GR-05-3064. LF acknowledges partial support from NSF grant IIS 2334033.

COMPETING INTERESTS

The authors have no competing interests to declare.

AUTHOR CONTRIBUTIONS

SB designed the project, with support from HB and LT. HB led coordination of project activities with support from SB, TTW, KCS, and CH. SB and HB wrote the first draft with support from KCS and CH. LF, LT, LW, TTW, and LCJ helped with writing and editing. All authors approved the final version of the manuscript.

AUTHOR AFFILIATIONS

Samantha Blickhan orcid.org/0000-0002-3775-5744
The Adler Planetarium, US
Hillary Burgess orcid.org/0009-0002-3628-9427
Independent Scholar, US
Kaylene Caswell Stocking orcid.org/0000-0003-3031-1769
UC Berkeley Kavli Center for Ethics, Science, and the Public, US
Chad Harper orcid.org/0000-0003-0356-9270
UC Berkeley Kavli Center for Ethics, Science, and the Public, US
Lucy Fortson orcid.org/0000-0002-1067-8558
University of Minnesota, US
Tanisha Tate Woodson orcid.org/0009-0000-7273-5143
Catalyst Consulting, US
Lea Witkowsky orcid.org/0000-0001-5690-7267
UC Berkeley Kavli Center for Ethics, Science, and the Public, US
L. Clifton Johnson orcid.org/0000-0001-6421-0953
The Adler Planetarium; Northwestern University, US
Laura Trouille orcid.org/0000-0002-1113-4122
The Adler Planetarium, US

REFERENCES

Abdul-Rahman, G., Zwitter, A., and **Haleem, N.** (2025) A systematic literature review on the role of artificial intelligence in citizen science. *Discover Artificial Intelligence*, 5(1), 165. DOI: https://doi.org/10.1007/s44163-025-00437-z

Anhalt-Depies, C., Stenglein, J.L., Zuckerberg, B., Townsend, P.A., and **Rissman, A.R.** (2019) Tradeoffs and tools for data quality, privacy, transparency, and trust in citizen science. *Biological Conservation*, 238: 108195. DOI: https://doi.org/10.1016/j.biocon.2019.108195

Aristeidou, M., Scanlon, E., and **Sharples, M.** (2020) Learning outcomes in online citizen science communities designed for inquiry. *International Journal of Science Education, Part B*, 10(4): 277–294. DOI: https://doi.org/10.1080/21548455.2020.1836689

Ballard, H.L., Harris, E.M., and **Dixon, C.G.H.** (2018) Science identity and agency in community and citizen science: evidence & potential. Commissioned for the Committee on Designing Citizen Science to Support Science Learning. Available at: https://www.mspnet.org/library/33673.html [Last accessed 24 September 2026].

Banerji, M., Lahav, O., Lintott, C.J., Abdalla, F.B., Schawinski, K., Bamford, S.P., Andreescu, D., Murray, P., et al. (2010) Galaxy Zoo: Reproducing galaxy morphologies via machine learning. *Monthly Notices of the Royal Astronomical Society*, 406(1): 342–353. DOI: https://doi.org/10.1111/j.1365-2966.2010.16713.x

Bender, E M., Gebru, T., McMillan-Major, A., and **Schmitchell, S.** (2021) On the dangers of stochastic parrots: Can language models be too big?. *FAcct '21: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, pp. 610–623. DOI: https://doi.org/10.1145/3442188.3445922

Bowser, A., Cooper, C., de Sherbinin, A., Wiggins, A., Brenton, P., Chuang, T.R., Faustman, E., et al. (2020) Still in need of norms: The state of the data in citizen science. *Citizen Science: Theory and Practice*, 5(1): 18. DOI: https://doi.org/10.5334/cstp.303

Buolamwini, J. and **Gebru, T.** (2018) Gender shades: Intersectional accuracy disparities in commercial gender classification. *Proceedings of the 1st Conference on Fairness, Accountability and Transparency*, pp. 77–91. Available at: https://proceedings.mlr.press/v81/buolamwini18a.html [Last accessed 24 September 2026].

Burgess, H.K., DeBey, L.B, Froehlich, H.E., Schmidt, N., Theobald, E.J., Ettinger, A.K., HilleRisLambers, J., et al. (2017) The science of citizen science: Exploring barriers to use as a primary research tool. *Biological Conservation*, 208: 113–120. DOI: https://doi.org/10.1016/j.biocon.2016.05.014

Cardamone, C.N., Schawinski, K., Sarzi, M., Bamford, S.P., Bennert, N., Urry, C.M., Lintott, C., et al. (2009) Galaxy Zoo

Green Peas: Discovery of a class of compact extremely star-forming galaxies. *Monthly Notices of the Royal Astronomical Society*, 399(3): 1191–1205. DOI: https://doi.org/10.1111/j.1365-2966.2009.15383.x

Ceccaroni, L., Bibby, J., Roger, E., Flemons, P., Michael, K., Fagan, L., and **Oliver, J.L.** (2019) Opportunities and risks for citizen science in the age of artificial intelligence. *Citizen Science: Theory and Practice*, 4(1): 29. DOI: https://doi.org/10.5334/cstp.241

Chakravorti, B. (2024) AI's trust problem. *Harvard Business Review*, 3 May 2024. Available at: https://hbr.org/2024/05/ais-trust-problem [Last accessed 16 November 2026].

Cooper, C. B., Rasmussen, L., and **Jones, E.** eds. (2022) A toolkit for data ethics in the participatory sciences. Citizen Science Association. Available at: https://participatorysciences.org/resources/data-ethics/ [Last accessed 24 September 2026].

de Sherbinin, A., Bowser, A., Chuang, T.R., Cooper, C., Fanielsen, F., Edmunds, R., Elias, P., et al. (2021) The critical importance of citizen science data. *Frontiers in Climate*, 3. DOI: https://doi.org/10.3389/fclim.2021.650760

de Vries, M., Land-Zandstra, A., and **Smeets, I.** (2019) Citizen scientists' preferences for communication of scientific output: A literature review. *Citizen Science: Theory and Practice*, 4(1): 2. DOI: https://doi.org/10.5334/cstp.136

Downs, R. R., Ramapriyan, H.K., Peng, G., and **Wei, Y.** (2021) Perspectives on citizen science data quality. *Frontiers in Climate*, 3. DOI: https://doi.org/10.3389/fclim.2021.615032

Dowthwaite, L., Spence, A., Lintott, C., Miller, G., Sprinks, J., Blickhan, S. and **Houghton, R.** (2025) Exploring the relationship between basic psychological needs and motivation in online citizen science. *ACM Transactions on Social Computing*, 8(1–2): 1–36. DOI: https://doi.org/10.1145/3702210

ECSA (European Citizen Science Association) (2015) Ten principles of citizen science. Berlin. DOI: https://doi.org/10.17605/OSF.IO/XPR2N

Eitzel, M.V., Cappadonna, J.L., Santos-Lang, C., Duerr, R.E., Virapongse, A., West, S.E., Kyba, C.C.M.K., et al. (2017) Citizen science terminology matters: Exploring key terms. *Citizen Science: Theory and Practice*, 2(1): 1. DOI: https://doi.org/10.5334/cstp.96

Ellwood, E.R., Pauly, G.B., Ahn, J., Golembiewski, K., Higgins, L.M., Ordeñana, M.A., von Konrat, M. (2023) Citizen science needs a name change. *Trends in Ecology & Evolution*, 38(6): 485–489. DOI: https://doi.org/10.1016/j.tree.2023.03.003

Ferrara, E. (2023) Fairness and bias in artificial intelligence: A brief survey of sources, impacts, and mitigation strategies. *Sci*, 6(1): 3. DOI: https://doi.org/10.3390/sci6010003

Fortson, L. (2021) From green peas to STEVE: Citizen science engagement in space science. *Space Science and Public Engagement*, Elsevier, 185–219. DOI: https://doi.org/10.1016/B978-0-12-817390-9.00009-9

Fortson, L., Crowston, K., Kloetzer, L., Ponti, M. (2024) Artificial intelligence and the future of citizen science. *Citizen Science: Theory and Practice*, 9(1): 32. DOI: https://doi.org/10.5334/cstp.812

Franzen, M., Kloetzer, L., Ponti, M., Trojan, J., Vicens, J. (2021) Machine learning in citizen science: Promises and implications. In: Vohland, K., et al. *The Science of Citizen Science*. Springer, Cham. DOI: https://doi.org/10.1007/978-3-030-58278-4_10

Gal, R., Alaluf, Y., Atzmon, Y., Patashnik, O., Bermano, A.H., Chechik, G., Cohen-Or, D. (2022) An image is worth one word: Personalizing text-to-image generation using textual inversion. *arXiv*. DOI: https://doi.org/10.48550/arXiv.2208.01618

Gerlich, M. (2025) AI tools in society: Impacts on cognitive offloading and the future of critical thinking. *Societies*, 15(1): 6. DOI: https://doi.org/10.3390/soc15010006

Gillespie, N., Lockey, S., Ward, T., Macdade, A., & Hassed, G. (2025) Trust, attitudes and use of artificial intelligence: A global study 2025. The University of Melbourne and KPMG. DOI: https://doi.org/10.26188/28822919

Göbel, C., Nold, C., Berditchevskaia, A., Haklay, M. (2019) How does citizen science 'do' governance? Reflections from the DITOs project. *Citizen Science: Theory and Practice*, 4(1): 31. DOI: https://doi.org/10.5334/cstp.204

Groot, B., and **Abma, T.** (2022) Ethics framework for citizen science and public and patient participation in research. *BMC Medical Ethics*, 23(1): 23. DOI: https://doi.org/10.1186/s12910-022-00761-4

Gunson, B., Murphy, B.L., and **Brown, L.J.** (2021) Knowledge mobilization, citizen science, and education. *Journal of Community Engagement and Scholarship*, 13(3): 36–55. DOI: https://doi.org/10.54656/VWXA9015

Guralnick, R., LaFrance, R., Denslow, M., Blickhan, S., Bouslog, M., Miller, S., Yost, J., et al. (2024) Humans in the loop: Community science and machine learning synergies for overcoming herbarium digitization bottlenecks. *Applications in Plant Sciences*, 12(1): e11560. DOI: https://doi.org/10.1002/aps3.11560

Hardinges, J., Pearson, S., and **Ross, R.** (2025) From human content to machine data: Introducing CC signals. *Creative Commons*. Available at: https://creativecommons.org/wp-content/uploads/2025/06/Human-Content-to-Machine-Data_Final.pdf [Last accessed 24 September 2026].

House, L.R., Gebhardt, K., Finkelstein, K., Cooper, E.M., Davis, D., Ciardullo, R., Farrow, D.J., et al. (2023) Using Dark Energy Explorers and machine learning to enhance the Hobby–Eberly telescope dark energy experiment. *The Astrophysical Journal*, 950(2): 82. DOI: https://doi.org/10.3847/1538-4357/accdd0

Huebner, S.E., Palmer, M.S., and **Packer, C.** (2024) Human supervision is key to achieving accurate AI-assisted wildlife

identifications in camera trap images. *Citizen Science: Theory and Practice*, 9(1): 38. DOI: https://doi.org/10.5334/cstp.752

Iacovides, I., Jennett, C., Cornish-Trestrail, C., and **Cox, A.L.** (2013) Do games attract or sustain engagement in citizen science?: A study of volunteer motivations. *CHI EA '13: CHI '13 Extended Abstracts on Human Factors in Computing Systems*, 1101–1106. DOI: https://doi.org/10.1145/2468356.2468553

Jackson, C.B., Østerlund, C., Mugar, G., Hassman, K.D., and **Crowston, K.** (2015) Motivations for sustained participation in crowdsourcing: Case studies of citizen science on the role of Talk. *2015 48th Hawaii International Conference on System Sciences*, Kauai, HI, 1624–1634. DOI: https://doi.org/10.1109/HICSS.2015.196

Klein, L., and **D'Ignazio, C.** (2024) Data feminism for AI. *FAcct '24: Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency*, 100–112. DOI: https://doi.org/10.1145/3630106.3658543

Kosmala, M., Wiggins, A., Swanson, A., and **Simmons, B.** (2016) Assessing data quality in citizen science. *Frontiers in Ecology and the Environment*, 14(10): 551–560. DOI: https://doi.org/10.1002/fee.1436

Liu, H.Y., Dörler, D., Heigl, F., and **Grossberndt, S.** (2021) Citizen science platforms. In: Vohland, K., et al. *The Science of Citizen Science*. Springer, Cham. DOI: https://doi.org/10.1007/978-3-030-58278-4_10

Lotfian, M., Ingensand, J., and **Brovelli, M.A.** (2021) The partnership of citizen science and machine learning: Benefits, risks, and future challenges for engagement, data collection, and data quality. *Sustainability*, 13(14): 8087. DOI: https://doi.org/10.3390/su13148087

Masters, K., Oh, E.Y., Cox, J., Simmons, B., Lintott, C., Graham, G., Greenhill, A., and **Holmes, K.** (2016) Science learning via participation in online citizen science. *Journal of Science Communication*, 15(3): A07. DOI: https://doi.org/10.22323/2.15030207

National Academies of Sciences, Engineering, and **Medicine.** (2017) Communicating science effectively: A research agenda. DOI: https://doi.org/10.17226/23674

Østerlund, Carsten, Crowston, K., Jackson, C.B., Wu, Y., Smith, A.O., Katsaggelos, A.K. (2024) Supporting human and machine co-learning in citizen science: Lessons from Gravity Spy. *Citizen Science: Theory and Practice*, 9(1): 42. DOI: https://doi.org/10.5334/cstp.738

Pankiv, K., and **Kloetzer, L.** (2024) Does using artificial intelligence in citizen science support volunteers' learning? An experimental study in ornithology. *Citizen Science Theory and Practice*, 9(1): 36. DOI: https://doi.org/10.5334/cstp.733

Phillips, T. B., Hebbard, C., and **Karl, R.** (2025) Measuring science identity in informal settings through citizen science: It's complicated. *International Journal of Science Education, Part B*, 15(3): 526–543. DOI: https://doi.org/10.1080/21548455.2025.2488405

Piper, A., Xu, M., and **Ruths, D.** (2024) The social lives of literary characters: Combining citizen science and language models to understand narrative social networks. *Proceedings of the 4th International Conference on Natural Language Processing for Digital Humanities*, Miami, FL: 472–482. DOI: https://doi.org/10.18653/v1/2024.nlp4dh-1.45

Prather, E.E., Cormier, S., Wallace, C.S., Lintott, C.J., Raddick, M.J., Smith, A. (2013) Measuring the conceptual understandings of citizen scientists participating in Zooniverse projects: A first approach. *Astronomy Education Review*, 12(1). DOI: https://doi.org/10.3847/AER2013002

Prestopnik, N.R., and **Crowston, K.** (2012) Citizen science system assemblages: Understanding the technologies that support crowdsourced science. *Proceedings of the 2012 iConference*, ACM, New York: 168–176. DOI: https://doi.org/10.1145/2132176.2132198

Raddick, M.J., Bracey, G., Gay, P.L., Lintott, C.J., Cardamone, C., Murray, P., Schawinski, K., et al. (2013) Galaxy Zoo: Motivations of citizen scientists. *arXiv*, DOI: https://doi.org/10.48550/arXiv.1303.6886

Reshetnikov, A., and **Marinescu, M.C.** (2025) Caption generation in cultural heritage: crowdsourced data and tuning multimodal large language models. *Proceedings of the 1st Workshop on Language Models for Underserved Communities (LM4UC 2025)*, Albuquerque, New Mexico: 42–50. DOI: https://doi.org/10.18653/v1/2025.lm4uc-1.6

Resnik, D.B., Elliott, K.C., and **Miller, A.K.** (2015) A framework for addressing ethical issues in citizen science. *Environmental Science & Policy*, 54: 475–481. DOI: https://doi.org/10.1016/j.envsci.2015.05.008

Shirk, J.L., Ballard, H.L., Wilderman, C.C., Phillips, T., Wiggins, A., Jordan, R., McCallie, E., et al. (2012) Public participation in scientific research: A framework for deliberate design. *Ecology and Society*, 17(2). https://www.jstor.org/stable/26269051.

Skarlatidou, A., Haklay, M., Hoyte, S., van Oudheusden, M., and **Bishop, I.J.** (2024) How can bottom-up citizen science restore public trust in environmental governance and sciences? Recommendations from three case studies. *Environmental Science & Policy*, 160. DOI: https://doi.org/10.1016/j.envsci.2024.103854

Smith, A. M., Lynn, S., Sullivan, M., Lintott, C.J., Nugent, P.E., Botyanszki, J., Kasliwal, M., et al. (2011) Galaxy Zoo Supernovae*. *Monthly Notices of the Royal Astronomical Society*, 412(2): 1309–1319. DOI: https://doi.org/10.1111/j.1365-2966.2010.17994.x

Sturgis, P., and **Allum, N.** (2004) Science in society: Re-evaluating the deficit model of public attitudes. *Public Understanding of Science*, 13(1): 55–74. DOI: https://doi.org/10.1177/0963662504042690

Trouille, L., Lintott, C.J., and **Fortson, L.** (2019) Citizen science frontiers: Efficiency, engagement, and serendipitous discovery with human–machine systems. *Proceedings of the National Academy of Sciences*, 116(6): 1902–1909. DOI: https://doi.org/10.1073/pnas.1807190116

Van Oudheusden, M., and **Abe, Y.** (2021) Beyond the Grassroots: Two trajectories of 'citizen sciencization' in environmental governance. *Citizen Science: Theory and Practice*, 6(1): 13. DOI: https://doi.org/10.5334/cstp.377

Wolf, S.M. (2020) Return of results in participant-driven research: Learning from transformative research models. *Journal of Law, Medicine & Ethics*, 48: 159–166. DOI: https://doi.org/10.1177/1073110520917042

Wynne, B. (2006) Public engagement as a means of restoring public trust in science – Hitting the notes, but missing the music? *Community Genetics*, 9(3): 211–220. DOI: https://doi.org/10.1159/000092659.

TO CITE THIS ARTICLE:
Blickhan, S., Burgess, H., Stocking, K.C., Harper, C., Fortson, L., Woodson, T.T., Witkowsky, L., Johnson, L.C. and Trouille, L. 2026. Collaborative Policy Development for Artificial Intelligence Ethics: A Case Study with the Zooniverse Platform. *Citizen Science: Theory and Practice,* 11(1): 54, pp. 1–14. DOI: https://doi.org/10.5334/cstp.1052

Submitted: 12 May 2026 **Accepted:** 23 October 2026 **Published:** 25 November 2026

COPYRIGHT:
© 2026 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See https://creativecommons.org/licenses/by/4.0/.

Citizen Science: Theory and Practice is a peer-reviewed open access journal published by Ubiquity Press.