CITIZEN SCIENCE:
THEORY AND PRACTICE

COLLECTION:
THE FUTURE
OF ARTIFICIAL
INTELLIGENCE AND
CITIZEN SCIENCE

EDITORIAL CONTENT

]u[ubiquity press

Artificial Intelligence and the Future of Citizen Science

LUCY FORTSON
KEVIN CROWSTON
LAURE KLOETZER
MARISA PONTI

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

INTRODUCTION

Artificial Intelligence (AI) and citizen science (CS) are two approaches to tackling data challenges related to scale and complexity. CS by its very definition relies on the joint effort of typically a distributed group of non-expert people to solve problems in a manner that relies on human intelligence. As AI capabilities increasingly augment or complement human intelligence, if not replicate it, there is a growing effort to understand the role that AI can play in CS and vice versa. With this growing interest as context, this special collection, The Future of AI and Citizen Science, illustrates the many ways that CS practitioners are integrating AI into their efforts, as well as identifies current limitations. In this spirit, our editorial briefly introduces the special collection papers to demonstrate and assess some uses of AI in CS; then, we contextualize these uses in terms of key challenges; and conclude with future directions that use AI with CS in both innovative and ethical ways.

To start, it is worth summarizing the data ecosystem presented in Figure 1 in which the integration of AI and CS is occurring. Data collection, data processing, and data analysis (McClure et al 2020) are the main activities undertaken by people participating in CS projects. These activities are typically initiated either in response to the need for widely distributed yet fine-grained spatiotemporal monitoring (Cooper, Shirk, and Zuckerberg 2014), or the ever-increasing demands to process and analyze "big data" (Trouille, Lintott, and Fortson 2019). Figure 1 thus sketches the flow of data from data collection (by machines or humans) to data processing or analysis, and shows how human-in-the-loop (HITL) and machine-in-the-loop (MITL) strategies can be deployed across a wide range of data domains and research disciplines. At the heart of this special collection then, is the idea that AI has the potential to significantly advance the field of CS by accelerating the pace and breadth of data processing, expanding the temporal and geographical reach of projects, enhancing the quality of collected and processed data, harnessing novel data sources, facilitating learning interactions between humans and machines, and broadening the spectrum of engagement opportunities for citizens (Lotfian, Ingensand, and Brovelli 2021; Ceccaroni et al. 2023).

CORRESPONDING AUTHOR:
Lucy Fortson
University of Minnesota, US
lfortson@umn.edu

KEYWORDS:
Artificial intelligence; machine learning; citizen science; community science; human-machine collaboration; AI ethics

TO CITE THIS ARTICLE:
Fortson, L, Crowston, K, Kloetzer, L and Ponti, M. 2024. Artificial Intelligence and the Future of Citizen Science. *Citizen Science: Theory and Practice,* 9(1): 32, pp. 1–9. DOI: https://doi.org/10.5334/cstp.812

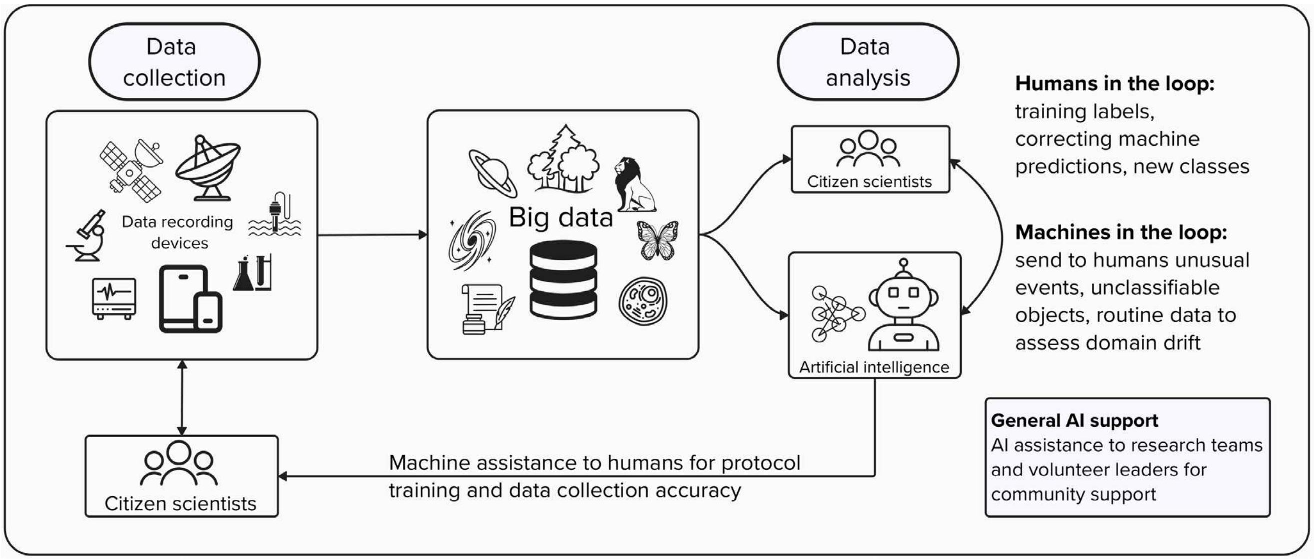

Figure 1 Data flow showing how the combination of citizen science and machine learning is used in both the data collection and analysis of data. Examples of Human-in-the-loop and Machine-in-the-loop actions are listed. Note that artificial intelligence can provide general support as well to the researchers and volunteers who recruit, engage, and sustain a volunteer community within a citizen science ecosystem.

EXPLORING THE CONTRIBUTIONS TO THE SPECIAL COLLECTION

We first give a brief overview of the papers in the Special Collection (see Figure 2 for detailed summaries of the collection contributions). Most employ AI to identify or locate specific classes of objects in images. Pennington et al. seek to identify viruses in cellular biology images; Nelson et al. and Chan et al. wish to better identify harmful mosquitoes and snails respectively; Sharma et al. would like to improve the accuracy in identifying bees in the UK; similarly, Huebner et al. seek to improve the identification of rare animal species in the Serengeti; and Meisner et al. would like to accelerate the identification of brown dwarfs, a difficult-to-detect astronomical object.

Other papers use a particular science objective as a means to study new ways to integrate AI with citizen science (CS). Østerlund et al. use a gravitational wave project to study how machines and humans can best learn from each other; Sankar et al. use a project identifying cloud types on Jupiter to learn whether AI can use lack of consensus in volunteer responses to learn something about novel relationships within the dataset; and Mantha et al. use a galaxy morphology project to study how combining human and machine anomaly detection strategies can efficiently find scientifically interesting objects in vast datasets. Nelson et al., Chan et al., and Sharma et al. also describe the development of custom apps that integrate AI to help improve data collection and analysis steps taken by citizen scientists. Pankiv et al. explores the use of AI in another custom app to investigate how the use of machine learning (ML) classification models affects the learning of novice birders, directly exploring the potential role of AI in participant learning. Duerinckx et al. describes a holistic approach to engaging citizens in learning about AI while co-creating potential projects that use AI to solve issues relevant to the public.

DEPLOYING ARTIFICIAL INTELLIGENCE TECHNOLOGY IN CITIZEN SCIENCE

To better understand the work presented in this collection, some background in the technology is helpful. The concepts of AI and ML were first introduced in the 1950s and tend to be used interchangeably, but there are significant differences. Briefly, AI is a more general term that describes computers that can emulate human thought, resulting in actions taking place in the real world. ML is a subset or a building block of AI, referring specifically to the algorithmic tools and technologies that learn from existing data to solve tasks such as pattern recognition or decision-making. Unless we are referring to specific ML techniques, we use the term AI in the context of this collection to refer to the broader range of techniques and applications, such as image classification, anomaly detection, feature extraction, and text summarization.

CS projects have deployed a diversity of AI techniques. The primary technology applied in the papers in this collection is deep learning, a subset of ML involving architectures that incorporate multiple layers of processing. Convolutional neural networks (CNNs) in particular are highly effective at extracting patterns directly from pixel data, providing revolutionary capabilities for handling large

Authors	Goal	Human task	Machine task	ML architecture and training
Chan et al.	Identify occurrence of disease transmitting snail species in remote regions of Africa.	Image collection of target species.	Object detection	Transfer learning from COCO data set onto five YOLOv4 object detection models (CNN-based) that were trained within the Google Colab Pro virtual environment.
Duerinckx et al.	Educate citizens on AI through co-creation of issues that then deploy AI solutions.	Design research questions and project goals; prioritize and test projects.	Various	Various
Huebner et al.	Improve the classification accuracy on rare species by judiciously combining AI and humans.	Validate machine predictions with high confidence scores requiring only minimal votes; classify species in images with low machine confidence.	Predict whether image has an animal or not; predict whether an animal in an image is one of 56 possible species classes.	CNN model trained using ResNet-18 architecture and Tensorflow.
Mantha et al.	Investigate the combination of human and machine proposed anomalies in large astronomical data sets to most efficiently find novel objects of scientific interest.	Pick images of galaxies that look odd or interesting; comment on interesting images with hashtags.	Provide list of objects ranked by anomaly score as determined by machine.	A generative CNN comprising a Wasserstein GAN with gradient penalty (WGAN-GP) produces anomaly scores, followed by an encoder to produce the feature space on which anomaly scores were evaluated.
Meisner et al.	Accelerate identification of rare astronomical object (brown dwarfs) in large data set by using AI to preselect images shown to volunteers.	Identify true brown dwarf candidates from a set of those proposed by AI.	Classify brown dwarf candidates in a large astronomical data set.	A custom Recurrent Neural Network architecture using 3-dimensional convolutional layers and 2-dimensional convolutional Long Short-Term Memory layers. Note that this ML model was developed by a project citizen scientist.
Nelson et al.	Incorporate AI into the GLOBE Observer citizen science data ecosystem in ways which can both complement and support the citizen scientists.	Collect and upload images of target species (plants, animals); validate a machine curated dataset; ideation of new projects.	Database ingest including AI photoscreening; data extraction including AI classification of uploaded images; provision to volunteers of AI-driven feedback and GeoAI data enrichment.	Amazon Rekognition which uses a combination of a pre-trained CNN and a region proposal network along with a classification algorithm (Fastest-RCNN) for detecting objects in an image.
Østerlund et al.	Investigating dynamic human and machine co-learning in CS projects by allowing humans to augment machine learning, and AI to augment human learning.	Identify specific patterns in spectrograms that could represent glitches in the detectors of gravitational waves; provide examples of unknown classes to and purify results from a similarity search.	Score the likelihood that a glitch belongs to individual known classes.	CNN with fusion and label smoothing; semi-supervised clustering (similarity search).
Pankiv et al.	Compare the learning of novice birders with and without the use of machine learning classification models.	Identify bird species with or without AI assistance.	Identify bird species.	CNN
Pennington et al.	Accelerate annotation and segmentation of volumetric bio-images.	Locate and classify viruses in bio-imaging data.	Segment viruses on images followed by class-wise counts.	3D CNN, specifically a 3D attention U-Net for segmentation.
Sankar et al.	Apply AI to discern the difference between intrinsic variability of a data set and the uncertainty in the response of citizen scientists to annotating images of cloud structures on Jupiter.	Classify specific cloud structures based on attributes such as color or shape.	Learn and map information between image-level features in the dataset and corresponding classification labels to identify and diagnose class-wise confusion.	A convolutional variational auto-encoder (VAE) conditioned with a classifier trained on labels provided by volunteers; followed by application of an attention technique using scorecam to visually locate features within an image.
Sharma et al.	Explore the integration of image recognition as a 'dialogic AI partner' in citizen science projects developed for identifying bee species in the UK.	Identify bumble bee species, with or without AI assistance.	Identify bumble bee species, with or without human assistance.	1. CNN Inception V3 model trained on the inaturalist dataset, and then used transfer learning to optimise for the task of bee species classification. 2. A Bayesian method to combine human and machine classifications.

Figure 2 This table presents a summary of the ways in which each paper in the special collection uses machine learning along with the task(s) asked of the citizen scientists and the overall goal of the work.

volumes of images. In Supplemental File 1, we provide a more detailed overview of the key concepts in modern deep learning using examples drawn from this collection. Therein, we discuss specific AI techniques in the context of deployment strategies in CS projects.

STRATEGIES IN COMBINING ARTIFICIAL INTELLIGENCE AND CITIZEN SCIENCE

Figure 3 provides an overview of the roles that humans can take in a project versus the roles that machines can play, and it also includes the supporting technologies needed to combine AI and CS. As shown at the bottom right of Figure 3, an ML model can be trained in a supervised manner (typically with human-provided labels), an unsupervised manner (with no labels provided, the machine learns patterns in the data often through clustering on features inherent in the dataset), or a semi-supervised manner (using labeled data to infer label information for unlabeled data in the dataset). A model can also be trained using transfer learning from a pre-trained model potentially from a completely different domain. Training must balance improved model performance against the need for large sets of training data and the difficulty of obtaining labels for supervised learning, while avoiding biases or gaps in the training data that will propagate to the model output. Furthermore, it is important to recognize that a machine model is just part of a system. Additional work is needed to preprocess images or other inputs for training and to connect a deployed model to input data and to use its output.

Many of the contributed papers describe custom ML architectures (represented in Figure 3, *top right*) based on "off-the-shelf" AI technologies (published papers with public code available). A few describe the use of cloud-based AI technologies and services such as Amazon's Rekognition package or Google's Colab. Authors used a range of tools to carry out the CS element (corresponding to Figure 3, *bottom left*). These tools can play a distinct role in what choices are available for project managers to incorporate AI into their projects, and which tools are used depends on which CS activities are paramount to the project. For example, the GLOBE Observer project (Nelson et al.) focuses on the collection and analysis of Earth system and environmental data, and has taken a holistic approach to integrating AI

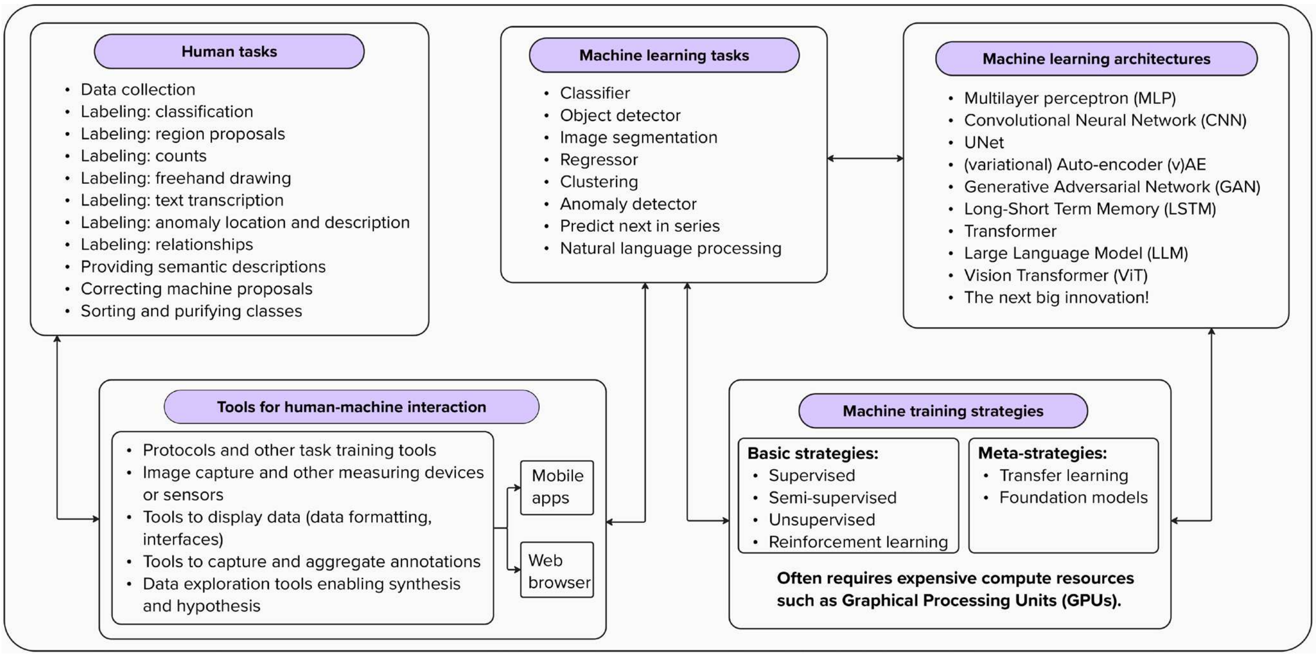

Figure 3 Interconnections between tasks that citizen scientists (humans) and machines each can carry out as well as the technology support ecosystem for these processes to take place. Different research goals will necessitate different combinations of tasks and human-machine interactions as well as different machine learning (ML) architectures which in turn require different machine training strategies. Note that one could also start with a specific ML strategy such as transfer learning, which would dictate a subset of useful architectures for a given ML task. This would then define the human-machine interaction needed to initiate the required human task. Supplemental File 1 provides brief descriptions of the ML side of the figure.

and CS by building out AI components in all three modes of their platform: data collection, data processing, and data analysis. One of the largest and longest-running CS efforts, the eBird project, has integrated AI into their Merlin mobile app, and is explored by Pankiv et al. Six papers (representing biomedicine, ecology, and astrophysics) used the Zooniverse platform, the largest general-purpose data analysis CS platform, which has facilitated the integration of ML into CS projects (Fortson 2021; Trouille, Lintott, and Fortson 2019). Other examples of large CS platforms that integrate ML are iNaturalist (Van Horn et al. 2018) and SciStarter (Zaken et al. 2021). Although these larger platforms provide common tools and services for CS projects, teams may want to develop their own project infrastructure integrating AI based on, for example, funding, specificity of need, or control of project content and goals.

SCIENTIFIC DOMAINS AND ARTIFICIAL INTELLIGENCE

Scientifically, the papers in this collection are concentrated in ecology and biodiversity (five papers), astronomy and astrophysics (four papers), and biomedicine (one paper), with one paper not specifying a domain. The characteristics of each domain and its data affect the possibility to use AI. In cellular biology or neuroimaging, the critical issue is that the biomedical field suffers from vast amounts of complex data; it is particularly difficult to obtain sufficient labels to train AI. The Pennington et al. contribution seeks to accelerate label gathering by combining two CS tasks — classification as well as localization of viruses—to train a model to count the viral load in cell images.

In conservation ecology and wildlife research, there has been real progress in adopting AI into CS. AI has been used to determine whether an animal is present in an image, help classify species by contributing a vote weighted on the ML confidence in the classification, and count animals (Willi et al. 2019; Green et al. 2020; Torney et al. 2019). A common challenge in these areas is an initial lack of training data due to monitoring locations with limited resources or class imbalance in the training set due to underlying differences in the prevalence of species. However, when ML models produce inaccurate classifications or encounter unfamiliar scenarios, human-in-the-loop (HITL) strategies can enhance the quality of datasets used for automation. This also serves to validate and refine initial ML outputs. In this context, the Huebner et al. contribution shows that HITL processes are critical to obtaining highly accurate classifications for rare or visually similar species that pose a challenge for AI. Because these projects often ask volunteers to collect data submitted through mobile apps, there is an interest in deploying AI within the app to assist the citizen scientist with species identification. The Sharma et al. contribution uses the BeeWatch app to investigate a more complex "dialogic" process that enables real-time dialog between the humans and AI, leading to improved identification accuracy of bumble bee species in the United Kingdom.

The fields of astronomy and astrophysics have also been incubators for exploring how best to combine the strengths of humans and machines (Fortson 2021). For instance, the Meisner et al. contribution deploys a sophisticated AI model to find objects that show movement between stacks of astronomical images. Importantly, the AI model used was developed by one of the project citizen scientists (Caselden et al. 2020). The critical issue in recent and upcoming astronomical surveys is the absolutely vast amount of data that will be produced by next-generation telescopes: billions of galaxy images versus the one million processed by volunteers in the original Galaxy Zoo (Lintott et al. 2008). While recent work (Walmsley et al. 2023) deploys an AI model that learns in near-real-time which images need human intervention, the sheer volume of data raises concerns about limiting opportunities for serendipitous discoveries, a hallmark of human involvement in these projects (e.g., Green Peas in Cardamone et al. 2009). The Mantha et al. contribution combines an AI anomaly detection technique trained on Galaxy Zoo images with the ability of citizen scientists to notice odd things in images to optimize the probability of finding scientifically interesting anomalies in large astrophysical data sets. In a similar vein, the Sankar et al. contribution uses the complexity inherent in images of Jupiter's clouds to explore a novel combination of humans and machines to determine whether there are undiscovered, scientifically interesting correlations between cloud features.

While the distribution of domains in this collection echoes the findings outlined in a recent literature review (Ponti and Seredko 2022), we note an increasing number of other domains integrating AI into CS, such as archaeology (e.g., Heritage Quest – Verschoof-van der Vaart et al 2020), seismology (e.g., MyShake app – Kong et al. 2018) and emergency awareness or disaster relief (e.g., Citizen Science Solution Toolkit – Bono et al. 2023; and the Planetary Response Network – Simmons et al. 2022). New AI technologies have led to a sizeable increase in Digital Humanities projects. For example, the Lives of Literary Characters recently used large language models (LLMs) along with "citizen readers" to learn how fictional social systems inform real-world social growth (Piper et al. 2024). And optical character recognition (OCR) trained on data curated by volunteers has been used to convert imaged labels from natural history collections into digital text (Guralnick et al. 2024).

GEOGRAPHICAL CONSIDERATIONS FOR ARTIFICIAL INTELLIGENCE AND CITIZEN SCIENCE

CS projects are concentrated in Europe, North America, and Australia, and the projects included in this collection are no exception. Yet, particularly in countries where data collection resources are limited, CS can be an important source of indigenous and traditional forms of knowledge for promoting sustainable development and addressing climate change on a national and regional level (Masselot et al. 2023; Reyes-García et al. 2022; Eicken et al. 2021). Furthermore, in developing economies, the lack of connectivity, storage, and processing infrastructure, as well as prerequisite human labor and expertise, often hinder the availability of data for AI (e.g., so-called "data deserts"). However, there is growing attention to both CS and AI in the Global South, including Africa, India, Southeast Asia, Latin American, and the Caribbean. There are now opportunities to integrate AI with CS within these regions, even though early development has primarily been concentrated in the North. Excellent examples of this trend are represented in this collection. First, the Chan et al. contribution explores the minimum data requirements, in terms of both quantity and quality, that local residents need to meet to train an acceptable model for identification of disease-bearing snails in Sub-Saharan Africa (SSA). Taking into consideration technical limitations for these countries, such as the high computational power required by the ML, the choice was made to deploy the model as a web-based application hosted in Belgium. This allows for residents to upload their images and receive prompt feedback on whether they have detected a target snail species. The Nelson et al. contribution describes how to facilitate CS in the Global South through the GLOBE Observer project. This project uses a trained ML model to identify mosquito species uploaded by citizen scientists, helping local communities in Africa monitor these dangerous malaria vectors.

CHALLENGES OF COMBINING HUMANS AND MACHINES

Looking across the papers, we identify a common set of challenges in applying AI for CS. A first challenge is identifying which technologies are appropriate. The choice of ML tasks that fit the problem dictates the machine training strategy, which in turn dictates the most appropriate ML architecture to use and the data needed (refer back to Figure 3). At the same time, the ML task may call for human input, for example, if the ML training strategy requires supervision or reinforcement, or if an unsupervised strategy requires evaluation. How volunteers interact with the machine is dictated by the form in which the machine information is provided. For example, projects that need volunteers to provide image labels for ML training will have a different task structure than one in which images already have a machine proposal that may need correcting.

A related challenge is that AI technology is developing extraordinarily quickly. Constraints such as limited resources, expertise, and infrastructure can make employing cutting-edge technology impractical. However, since CS facilitates human validation of ML, the most sophisticated models may be unnecessary, which can be particularly advantageous for groups lacking computational resources. Using smaller or older models may also be more environmentally friendly.

The next challenge is to design tasks that are enjoyable for volunteers to complete while still enhancing the results of automated approaches (Guralnick et al. 2024). Finding roles that citizen scientists can play meaningfully alongside experts and AI technologies is essential. Concerns have been raised over the potential of AI to disengage citizen scientists. For instance, the use of AI can reduce the range of possible volunteer contributions or make their tasks either too simple or too complex (Trouille, Lintott, and Fortson 2019). Delegating interactions to AI might be efficient for science teams but distancing for volunteers.

Balancing efficiency and speed with other goals, such as volunteer engagement, learning, and development, is crucial and depends significantly on the overarching goal of the research team (Fortson 2021; Pankiv et al. in this collection). This balance becomes especially important when addressing long-standing challenges in CS, such as diversity and inclusion (Cooper et al. 2021). For instance, while the integration of AI in CS enables speedier data processing and improves efficiency, it can also create a valuable learning opportunity for participants—or reduce learning. The point is disputed and likely depends on the design of the project. The concept of ML supporting human learning is central to the co-learning process presented by Østerlund et al. in this collection. In the Gravity Spy project, the AI model plays an active role in guiding volunteers through progressively complex tasks that align with their evolving Zone of Proximal Development (the sweet spot between what a learner can do on their own and what they can achieve with guidance from a more knowledgeable person). Staying in the zone allows volunteers to gradually expand their knowledge and skills by tackling challenges that are neither too easy nor too difficult. At the same time, the humans are finding new classes of objects that can in turn improve the machine learning. In contrast, the Pankiv et al. paper shows that use of the Merlin AI assistant for eBird resulted in less domain learning by novice birders. Understanding these different outcomes is an important task for future research.

Progress in understanding how to introduce AI has been made with a study by Gal et al. (2022), which found that, when given a choice, Galaxy Zoo volunteers preferred by a wide margin to work alongside machines and that specific motivational statements were important to this engagement. The Meisner et al. contribution draws similar conclusions from their survey, though underscores the ethical tightrope that needs to be walked in terms of how much a project must divulge about the use of AI or not in recruitment materials; too much and you may risk losing participants, not enough and the project isn't being transparent about the use of AI.

The Duerinckx et al. contribution speaks to a major challenge for the use of AI in CS, which is the lack of understanding of AI models among the general population. By including the public in the co-creation step of defining which citizen science projects with integrated AI should be developed, the study demonstrates an example of democratization of AI innovation, empowering citizens to understand the potential uses of CS coupled with AI.

CONCLUSION: ARTIFICIAL INTELLIGENCE AND THE FUTURE OF CITIZEN SCIENCE

With the 2024 Nobel Prize in Physics being given to key architects of modern ML (John Hopfield and Geoffrey Hinton), the field of AI has clearly made its mark on society. But what does the future bring for AI and grand challenges in science, and thus CS? There are at least four areas where large amounts of effort by researchers in AI (including those at most of the major tech companies) will likely have an impact on CS: anomaly detection, LLMs, foundation models, and trustworthy, ethical, and explainable AI.

As discussed previously, anomaly detection is relevant for HITL strategies to detect scientifically interesting or rare objects; further research in this area is needed to best integrate with CS. LLMs can support research teams or volunteers with leadership roles for actions such as project recruitment, communication, and retention. Fertile research areas include the impact of in-project messaging on retention or learning, or how LLMs can be harnessed as one element of the task structure. Further, LLMs might be deployed to provide answers to frequent questions of the community of volunteers, explain key underlying concepts or principles, give feedback on individual and collective progress, or share summaries of scientific papers and findings. AI support may thus alleviate time-consuming

CHALLENGE	DESCRIPTION
AI opacity	AI models can be complex, making it difficult to understand their decision-making and identify potential biases.
Public trust	People need to understand how AI is used to build trust and encourage participation.
Citizen science without (meaningful tasks for) citizens	Fear that AI systems could reduce engagement due to task simplification, or replace humans in performing certain tasks.
Data reliability	Volunteer collected data may be inaccurate. AI can help to ensure high-quality data necessary for accurate results but trust needs to be built.
Data ownership	AI requires data for training. Multiple parties may have ownership claims to the data. Clear agreements are needed to prevent conflicts and ensure ethical data handling.
Risk of exclusion	Some countries/institutions/communities might be unable to apply AI in CS due to lack of internet connectivity or appropriate devices.
Limited resources	Projects with limited resources may lack both the ability to implement and maintain AI systems effectively, as well as the expertise in using AI ethically.
Environmental impact	Training sophisticated AI models demands huge computational power with significant energy consumption and greenhouse gas emissions, as well as often environmentally damaging extraction of rare-earth metals for training hardware (e.g., Graphical Processing Units [GPUs]).

Table 1 Key ethical considerations for using artificial intelligence (AI) in citizen science (CS) projects.

communication tasks and allow scientists to focus on the most significant interactions with the volunteers.

LLMs are just one example of so-called foundation models, which use extraordinarily large amounts of diverse data to train, with the intention that this generic model can then be readily transferred to highly specific cases. The provision of open access foundation models for a range of domains important to CS (e.g., astronomy, ecology, environment, health) could reduce barriers to the inclusion of AI in CS projects. At the same time, it is imperative that the provenance of the large amounts of data used in the development of foundation models complies with open data practices, especially if those data (including annotations) are produced via CS methods. This leads not only to research in AI, but necessitates formulation of policy surrounding the development of AI that includes the voices of those impacted by it—the citizens.

We close then with the point that CS projects should strive to use AI in a transparent, democratic, and trustworthy manner promoting ethical ("moral AI") for the public good. Table 1 lists several challenges for volunteer participants and CS project organizers in achieving this goal. Particular attention is needed for the Global South due to the disadvantaged positions of these countries in the race for AI development along with the threats of "data colonialism" (Baezner and Robin 2018). We need more empirical studies to explore how AI can be effectively and equitably implemented. Research should focus on tailoring AI to local contexts, involving local communities in decision-making, and ensuring long-term sustainability and local innovation.

While the future of artificial intelligence and citizen science is bright, it is critical that all practitioners address the inherent challenges through clear communication, open processes, and public education to ensure that AI in citizen science is used ethically and effectively for the public good.

SUPPLEMENTARY FILE

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

- **Supplemental File 1.** Overview of deep-learning technology with example use-cases from the special collection. DOI: https://doi.org/10.5334/cstp.812.s1

ACKNOWLEDGEMENTS

Lucy Fortson would like to acknowledge partial support from the US National Science Foundation under grant IIS 2006894, NASA award #80 NSSC20M0057 and a Leverhulme Foundation Visiting Professorship grant VP2-2022-016. Kevin Crowston is partly supported by funding from the National Science Foundation of the United States, grants 21-29047 and 21-06865.

AUTHOR CONTRIBUTIONS

Fortson generated Figures 1, 2, and 3. All four authors contributed equally to other contributions including conception, writing, and editing.

COMPETING INTERESTS

The authors have no competing interests to declare.

AUTHOR AFFILIATIONS

Lucy Fortson orcid.org/0000-0002-1067-8558
University of Minnesota, US

Kevin Crowston orcid.org/0000-0003-1996-3600
Syracuse University, US

Laure Kloetzer orcid.org/0000-0001-6703-8562
University of Neuchâtel, CH

Marisa Ponti orcid.org/0000-0003-4708-4048
University of Gothenburg, SE

REFERENCES

Baezner, M., Robin, P. (2018) Cyber Sovereignty and Data Sovereignty (Report). ETH Zurich. DOI: https://doi.org/10.3929/ethz-b-000314613

Bono, C., Mülâyim, M.O., Cappiello, C., Carman, M.J., Cerquides, J., Fernandez-Marquez, J.L., Mondardini, M.R., et al. (2023) A Citizen Science Approach for Analyzing Social Media with Crowdsourcing. *IEEE Access*, 11, pp. 15329–15347. DOI: https://doi.org/10.1109/ACCESS.2023.3243791

Caselden, D., Colin, G., Lack, L., Marocco, F., Kirkpatrick, J., and **Meisner, A.** (2020) WISE image cubes, neural nets, and moving objects with SMDET. *Bulletin of the American Astronomical Society*, 52 (1). DOI: https://ui.adsabs.harvard.edu/abs/2020AAS...23527418C

Cardamone, C., 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, pp. 1191–1205. DOI: https://doi.org/10.1111/j.1365–2966.2009.15383.x

Ceccaroni, L., Oliver, J.L., Roger, E., Bibby, J., Flemons, P., Michael, K., and **Joly, A.** (2023) Advancing the productivity of science with citizen science and artificial intelligence | OECD iLibrary, in: *Artificial Intelligence in Science Challenges, Opportunities and the Future of Research*, pp. 148–154. DOI: https://doi.org/10.1787/69563b12-en

Cooper, C.B., Hawn, C.L., Larson, L.R., Parrish, J.K., Bowser, G., Cavalier, D., Dunn, R.R., et al. (2021) Inclusion in citizen science: The conundrum of rebranding. *Science* 372, pp. 1386–1388. DOI: https://doi.org/10.1126/science.abi6487

Cooper, C.B., Shirk, J., and **Zuckerberg, B.** (2014) The invisible prevalence of citizen science in global research: migratory birds and climate change. *PloS ONE* 9, e106508. DOI: https://doi.org/10.1371/journal.pone.0106508

Eicken, H., Danielsen, F., Sam, J.-M., Fidel, M., Johnson, N., Poulsen, M.K., Lee, O.A., et al. (2021) Connecting top-down and bottom-up approaches in environmental observing. *BioScience* 71, pp. 467–483. DOI: https://doi.org/10.1093/biosci/biab018

Fortson, L. (2021) Chapter 10- From Green Peas to STEVE: citizen science engagement in space science, in: Kaminski, A.P. (Ed.), *Space Science and Public Engagement*. Elsevier, pp. 185–219. DOI: https://doi.org/10.1016/B978-0-12-817390-9.00009-9

Gal, Y. (Kobi), Segal, A., Kamar, E., Horvitz, E., Lintott, C., and Walmsley, M. (2022) A new workflow for human-AI collaboration in citizen science, in: *Proceedings of the 2022 ACM Conference on Information Technology for Social Good, GoodIT '22*. Association for Computing Machinery, New York, NY, USA, pp. 89–95. DOI: https://doi.org/10.1145/3524458.3547243

Green, S.E., Rees, J.P., Stephens, P.A., Hill, R.A., and **Giordano, A.J.** (2020) Innovations in camera trapping technology and approaches: the integration of citizen science and artificial intelligence. *Animals* 10, p. 132. DOI: https://doi.org/10.3390/ani10010132

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*, e11560. DOI: https://doi.org/10.1002/aps3.11560

Kong, Q., Inbal, A., Allen, R.M., Lv, Q., Puder, A. (2018) Machine learning aspects of the MyShake global smartphone seismic network. *Seismological Research Letters*, 90, pp. 546–552. DOI: https://doi.org/10.1785/0220180309

Lintott, C.J., Schawinski, K., Slosar, A., Land, K., Bamford, S., Thomas, D., Raddick, M.J., et al. (2008) Galaxy Zoo: morphologies derived from visual inspection of galaxies from the Sloan Digital Sky Survey. *Monthly Notices of the Royal Astronomical Society*, 389, pp. 1179–1189. DOI: https://doi.org/10.1111/j.1365-2966.2008.13689.x

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

Masselot, C., Jeyaram, R., Tackx, R., Fernandez-Marquez, J.L., Grey, F., and **Santolini, M.** (2023) Collaboration and performance of citizen science projects addressing the sustainable development goals. *Citizen Science: Theory and Practice*, 8. DOI: https://doi.org/10.5334/cstp.565

McClure, E.C., Sievers, M., Brown, C.J., Buelow, C.A., Ditria, E.M., Hayes, M.A., Pearson, R.M., et al. (2020) Artificial intelligence meets citizen science to supercharge ecological monitoring. *Patterns* 1, 100109. DOI: https://doi.org/10.1016/j.patter.2020.100109

Piper, A. (2024) CR4Interact: citizen readers for character interactions dataset. DOI: https://doi.org/10.5683/SP3/QMIARS

Ponti, M. and **Seredko, A.** (2022) Human-machine-learning integration and task allocation in citizen science. *Humanities and Social Sciences Communications – Nature*, 9, pp. 1–15. DOI: https://doi.org/10.1057/s41599-022-01049-z

Reyes-García, V., Tofighi-Niaki, A., Austin, B.J., Benyei, P., Danielsen, F., Fernández-Llamazares, Á., Sharma, A., et al. (2022) Data sovereignty in community-based environmental monitoring: Toward equitable environmental data governance. *BioScience*, 72, pp. 714–717. DOI: https://doi.org/10.1093/biosci/biac048

Simmons, B.D., Lintott, C., Reece, S., Allen, C., Miller, G.R.M., Yore, R., Jones, D., et al. (2022) Disaster, infrastructure and participatory knowledge: The Planetary Response Network. *Citizen Science: Theory and Practice*, 7. DOI: https://doi.org/10.5334/cstp.392

Torney, C.J., Lloyd-Jones, D.J., Chevallier, M., Moyer, D.C., Maliti, H.T., Mwita, M., Kohi, E.M., et al. (2019) A comparison of deep learning and citizen science techniques for counting wildlife in aerial survey images. Methods in Ecology and Evolution, 10, pp. 779–787. DOI: https://doi.org/10.1111/2041-210X.13165

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

Van Horn, G., Mac Aodha, O., Song, Y., Cui, Y., Sun, C., Shepard, A., Adam, H., et al. (2018) The INaturalist species classification and detection dataset, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. pp. 8769–8778. DOI: https://doi.org/10.1109/CVPR.2018.00914

Verschoof-van der Vaart, W.B., Lambers, K., Kowalczyk, W., and **Bourgeois, Q.P.J.** (2020) Combining deep learning and location-based ranking for large-scale archaeological prospection of LiDAR data from The Netherlands. ISPRS International Journal of Geo-Information, 9, p. 293. DOI: https://doi.org/10.3390/ijgi9050293

Walmsley, M., Géron, T., Kruk, S., Scaife, A.M.M., Lintott, C., Masters, K.L., Dawson, J.M., et al. (2023) Galaxy Zoo DESI: Detailed morphology measurements for 8.7M galaxies in the DESI Legacy Imaging Surveys. Monthly Notices of the Royal Astronomical Society, 526, pp. 4768–4786. DOI: https://doi.org/10.1093/mnras/stad2919

Willi, M., Pitman, R.T., Cardoso, A.W., Locke, C., Swanson, A., Boyer, A., Veldthuis, M., and **Fortson, L.** (2019) Identifying animal species in camera trap images using deep learning and citizen science. Methods in Ecology and Evolution, 10, pp. 80–91. DOI: https://doi.org/10.1111/2041-210X.13099

Zaken, D.B., Gal, K., Shani, G., Segal, A., and **Cavalier, D.** (2021) Intelligent recommendations for citizen science. *Proceedings of the AAAI Conference on Artificial Intelligence*, 35 (17), pp. 14693–14701. DOI: https://doi.org/10.1609/aaai.v35i17.17726

TO CITE THIS ARTICLE:
Fortson, L, Crowston, K, Kloetzer, L and Ponti, M. 2024. Artificial Intelligence and the Future of Citizen Science. *Citizen Science: Theory and Practice,* 9(1): 32, pp. 1–9. DOI: https://doi.org/10.5334/cstp.812

Submitted: 28 October 2024 **Accepted:** 28 October 2024 **Published:** 09 December 2024