Advancing the productivity of science with citizen science and artificial intelligence

L. Ceccaroni, Earthwatch, United Kingdom

J.L. Oliver, University of Sydney, Australia

E. Roger, Commonwealth Scientific and Industrial Research Organisation (CSIRO), Australia

J. Bibby, Industry and Environment, Australia

P. Flemons, Australian Museum, Australia

K. Michael, Arizona State University, United States

A. Joly, Institut national de recherche en sciences et technologies du numérique (INRIA), France

Introduction

Citizen science is a form of scientific inquiry where members of the public engage in scientific investigations, often in collaboration with, or under the direction of, professional scientists and scientific institutions. It supports scientific research and applied sciences through a wide range of activities and across diverse topics. Thanks to advances in communication and computing technologies, the public can collaboratively participate in new ways in citizen science projects. For example, participants submit observations and samples about the environment via eBird, iNaturalist or the EchidnaCSI project, among other platforms. They also engage on line by transcribing historical documents or classifying photographs, audio and video via platforms such as DigiVol or Zooniverse. In other cases, participants collaboratively solve mathematical problems via the Polymath Project, or play online games via Foldit, to inform medical research. The public disseminates project outcomes as well.

To date, the most significant impact of citizen science in accelerating scientific discoveries has been in relation to data collection and processing activities (Bonney et al., 2016). Citizen science continues to gain support and acceptance, delivering positive societal, economic and environmental impacts. Many projects actively support learning about specific topics, increase understanding of science and inform decision making (Bonney et al., 2014). Citizen scientists are involved in projects across scientific domains such as astronomy, chemistry, computer science, environmental science, mathematics, medicine and social science. However, the vast majority of citizen-science projects support the understanding of biodiversity, wildlife, plants and environmental processes (Kullenberg and Kasperowski, 2016).

Intelligence demonstrated by machines, known as artificial intelligence (AI), is widely applied across various scientific domains. Citizen science is no exception, and is increasingly being enhanced by the integration of AI (Ceccaroni et al., 2019). This essay examines the synergy of AI and citizen science to improve the productivity of science. It concludes by exploring future opportunities and considerations in this emerging area, including policy implications.

How citizen science coupled with artificial intelligence can increase the productivity of science

Over the past decade, there has been huge growth in the capabilities and applications of AI in citizen science. These applications can take an unsupervised or supervised machine-learning approach. In the former, data do not have to be annotated accurately by people first. In the latter, which occurs more frequently, data labelled by humans are needed to train the AI algorithms. At present, citizen science systems using AI are advancing science through a variety of mechanisms:

- increasing the speed and scale of data processing
- increasing projects' temporal and geographical scope
- improving the quality of data collected and processed
- supporting learning between humans and machines
- leveraging new data sources
- diversifying engagement opportunities.

These mechanisms are detailed below, and current examples are provided.

Increasing the speed and scale of data processing

Cameras triggered by motion typically capture many photos of moving vegetation rather than the intended animals moving past. Audio, video and other media can often be filtered using similar machine-learning techniques. In this case, AI algorithms can reduce how much data need to be processed by humans. AI is used to filter out false positives in images so that citizen scientists are more likely to see photographs of animals that need identification (Willi et al., 2018). More robust integration of AI and citizen science applied to the ever-growing volume of measures used by ecological studies will lead to more conclusive environmental insights at scale (Tuia et al., 2022). A similar filtering technique is applied in Galaxy Zoo, an online citizen science project. In this project, participants classify types of galaxies based on visible features in satellite photographs. The analysis of large amounts of data is facilitated by image pre-processing performed by AI. Here, the combination of humans and machines, often referred to as human-machine teaming, increases the rate of data processing (Beck et al., 2018).

Increasing projects' temporal and geographical scope

There is growing awareness of the potential of citizen science (integrated with AI) to expand environmental monitoring programmes. These include projects where solutions depend on large numbers of observations distributed across space and time (McClure et al., 2020). The Pl@ntNet citizen-science platform, for instance, includes tools to identify plants automatically. This has resulted in citizen scientists contributing more accurate data to global repositories and monitoring projects (Bonnet et al., 2020). Similarly, the iNaturalist project include tools to automatically identify most species, respectively. This has enabled the collection of observations at temporal and spatial scales not achievable with traditional science.

Improving the quality of data collected and processed

Several highly successful projects use AI to improve the quality of data collected and processed. Through the global platform eBird, birdwatchers have submitted copious bird observations, which have informed development of species distribution models (Sullivan, et al., 2014). These models have subsequently been applied to improve data quality by automatically filtering out observations of bird species residing outside of the birdwatcher's location (Kelling et al., 2012).

Supporting learning between humans and machines

Citizen scientists can contribute to training AI to solve complex analytical tasks usually carried out by experts. Human-in-the-loop processes are systems built with human supervision at different stages of the project cycle. For example, humans create and label datasets that are then used to train AI algorithms and models, with humans overseeing the models and fine-tuning them. Humans can also test and validate these models, resulting in high-quality AI systems. Several large citizen science projects focused on identifying species, such as iNaturalist, Pl@ntNet and BirdNet, are strongly enhanced by adopting a human-in-the-loop approach. In some cases, these types of human-AI systems can train AI algorithms to recognise species almost as accurately as humans with species expertise (Bonnet et al., 2018). In the online Gravity Spy project, participants identify glitches in visual representations of data from interferometers to assist scientists' search for gravitational waves; AI is used to train newcomers to learn more quickly (Jackson et al., 2020). Such AI integrations make projects more efficient.

Another example is a monitoring project called Penguin Watch, in which humans analyse time-lapse images of penguin colonies (Jones et al., 2020). This analysis by volunteers greatly helps assess the reliability of the AI algorithm used to identify species. It also helps refine it in different conditions (day and night) and for the different species.

In iNaturalist, AI provides participants with immediate feedback, derived from computer-vision models, about the organisms (plants, animals or fungi) in the photographs submitted. This feedback is an opportunity for citizen scientists to learn more about biodiversity, and has the potential to maintain their engagement in the project (Van Horn et al., 2018). Other, more expert members of the community identify species more specifically or validate AI identifications. Such contributions are used to refine the computer-vision algorithms (Van Horn et al., 2018).

Leveraging new data sources

Tapping into non-traditional data sources, such as social media, with the support of AI (data filtering), can vastly enhance the temporal and geographic availability of data and collect real-time information (MacDonald et al., 2015). In the Aurorasaurus project, participants submit observations and verifications of aurora sightings. The project is relatively novel in aggregating observations from both direct submissions through the project website and social media. Several other projects (particularly weather observation projects) are also starting to harvest data from social media platforms such as Twitter to increase the amount of available data for analysis (MacDonald et al., 2015).

Diversifying engagement opportunities

The use of AI offers more ways for participants to take part, and increased engagement provides more information for scientific investigations. Some people enjoy searching through a lot of data to find something uncommon. Participants may, for example, hope to see wildlife captured in photographs from motion-triggered cameras (Bowyer et al., 2015) or hear the calls of a rare bird species (Oliver et al., 2019). In some cases, AI can be trained to quickly perform tasks that might be considered time-consuming or uninteresting to some participants. This allows the citizen science community to engage with tasks that are considered more exciting and challenging (Ceccaroni et al., 2019). In some camera trap projects, AI is used to remove false positives in images. This enables citizen scientists to focus on identifying animals just in the pictures where an animal occurs, saving their time. In the iNaturalist application, AI assists species identification and increases biodiversity knowledge in participants using the platform (Unger et al., 2021).

Future applications

Opportunities exist for further growth of AI-supported citizen science. These include developing new AI applications; more accessible ways for non-experts to use AI techniques; and increased private investment in AI, similar to Microsoft's existing investment in "AI for Earth" (Joppa, 2017). Realising these opportunities will likely result in more participants using AI-assisted citizen science applications (Rzanny et al., 2022). It can also lead to including more citizen science data in international data repositories. This is useful because these data are generally more accessible to the public, researchers and policy makers. In the future, AI will be increasingly applied in citizen science. Applications will include autonomous systems of all types, such as drones, autonomous vehicles, and other robotic and remote sensing instrumentation that is integrated with AI. It will also include improvements in mobile applications and hardware, and communication technologies such as wireless broadband networks and cloud computing. All these emerging applications will give rise to new capabilities, particularly in data collection and in the automatic detection and identification of items in images, audio recordings or videos.

In integrating AI and citizen science, risks, traceability, transparency and upgradability of AI algorithms and AI-assisted information systems must be carefully considered (Ceccaroni et al., 2019; Ponti et al., 2021). Traceability is essential to reproduce, qualify and revise the data generated by AI algorithms (e.g. through version control and accessibility of the AI models). Transparency is crucial for understanding and correcting biases in AI models (e.g. by making training data fully accessible). Without appropriate transparency, errors by AI algorithms cannot be understood or, in some cases, even detected. Upgradability – the ability of AI algorithms to be upgraded over time – is necessary to accommodate new inputs and corrections made by experts and citizen scientists.

Additionally, quantifying uncertainty is essential. In the case of citizen science, uncertainty originates from any error or bias in the data collection, classification or processing resulting from AI algorithms (e.g. results, predictions) or participants, and from natural data variance. It is crucial to maintain meta-information on how the data have been treated throughout the data's life cycle. Tracking uncertainty can ensure that the related variables and biases (e.g. errors in an observation map that may affect subsequent decisions) are findable, accessible, interoperable and reusable (Wilkinson et al., 2016). A first step to achieving this, in relation to biodiversity, could be integrating the uncertainty associated with species identification into Darwin Core (i.e. a broadly accepted biodiversity data standard). This information could then be made searchable in biodiversity data applications. The allowable uncertainty in data ultimately depends on how the data are being used. Data quality cannot be reduced to a binary attribute (usable vs. unusable). For example, the construction of a species distribution model can tolerate a certain percentage of error in the input data (Botella et al., 2018). However, a single erroneous observation can severely impact a warning system based on the early detection of certain species (Botella et al., 2018).

Policy considerations

As technology improves, machines will perform more of the heavy data processing and time-consuming aspects of citizen science projects. This provokes several questions. How will citizen scientists be motivated to maintain their involvement in projects? How can they be engaged with learning? How can they be educated? How can their contributions be appropriately attributed and acknowledged? How can their time and effort be rewarded? Finally, how can data exploitation and ownership be managed? (Franzen et al., 2021; Ponti et al., 2021). Without resolving these challenges, the interest and participation in citizen science might decrease. It will be an ongoing challenge to ensure these issues are adequately considered. At the same time, these issues should not hinder or limit citizen science and detract from its appeal. Indeed, AI could attract more people to citizen science because some (youth, for example), especially curious about AI, might be drawn to the domain.

Policy makers should dedicate resources to generating creative ideas on how AI could help advance science productivity with citizen science around the following issues:

- **Expansion of the range of science project types that can use citizen science**. To date, the area has been primarily dominated by projects on biodiversity, wildlife, plants and environmental processes. Typically, these research domains have a more extended history of readily engaging the public. AI's contribution to these areas has henceforth evolved the most.
- **Best practice guidance for scientists, technologists and broader groups so they can adopt a citizen science approach**. Guidance is especially needed for breaking complex research projects into discrete tasks that citizen scientists can then undertake. AI could assist in this partitioning of tasks.
- **Validation of citizen science contributions by quantifying the accuracy of output**. AI could help ensure adherence to the scientific method and assist in quality and impact assessment, whose metrics continue to be challenging for citizen science projects to report on (Wehn et al., 2021). Improved reporting measures could help alleviate long-running concerns over data quality that remain prevalent in citizen science and science more broadly.
- **Proper application of AI**. Joppa (2017) suggests that, for every problem, two questions should be asked: "How can AI help solve this?" and "How can we facilitate its application?" An additional question should also be asked: "How can we ensure that each use of AI in citizen science carefully considers risks, traceability, transparency and upgradability?"

Conclusion

Citizen science at local, national and global scales represents an opportunity for a shift in the ability to inform scientific inquiries, enrich lives and engage diverse communities in science. As citizen science grows, new technologies will likely proliferate, supporting people in learning, exchanging information and solving problems collaboratively. With these new technologies making data more accessible and interpretable, new opportunities for synergies between citizen science and AI are likely to emerge. This essay has described how AI, coupled with citizen science, can enhance the productivity of science.

The new technologies, which will integrate AI into citizen science and facilitate automation, also come with potential risks. Project leaders will need to consider these risks and how best to mitigate them to ensure transparency and positive outcomes. The success of this integration, in terms of increasing the scientific and public benefit and enhancing the productivity of science, will require continued investment. It will also demand consideration in areas such as ethics, motivations and attribution for diverse groups of participants, system development, system optimisation, data quality and impact assessment.

References

Beck, M.R. et al. (2018), "Integrating human and machine intelligence in galaxy morphology classification tasks", *Monthly Notices of the Royal Astronomical Society*, Vol. 476/4, pp. 5516-5534, https://doi.org/10.1093/mnras/sty503.

Bonnet, P. et al. (2020), "How citizen scientists contribute to monitor protected areas thanks to automatic plant identification tools", *Ecological Solutions and Evidence*, Vol.1/2, p. e12023, https://doi.org/10.1002/2688-8319.12023.

Bonnet, P. et al. (2018), "Plant identification: Experts vs. machines in the era of deep learning", in *Multimedia Tools and Applications for Environmental & Biodiversity Informatics*, Springer, Cham.

Bonney, R. et al. (2016), "Can citizen science enhance public understanding of science?", *Public Understanding of Science*, Vol. 25/1, pp. 2-16, https://doi.org/10.1177/0963662515607406.

Bonney, R. et al. (2014), "Next steps for citizen science", *Science*, Vol. 343/6178, pp. 1436-1437, https://doi.org/10.1126/science.125155.

Botella, C. et al. (2018), "Species distribution modeling based on the automated identification of citizen observations", *Applications in Plant Sciences*, Vol. 6/2, p. e1029, https://doi.org/10.1002/aps3.1029.

Bowyer, A. et al. (2015), "Mundane images increase citizen science participation", presentation to 2015 Conference on Human Computation & Crowdsourcing, San Diego, https://doi.org/10.13140/RG.2.2.35844.53121

Ceccaroni, L. et al. (2019), "Opportunities and risks for citizen science in the age of artificial intelligence", *Citizen Science: Theory and Practice*, Vol. 4/1, p. 29, http://doi.org/10.5334/cstp.241.

Franzen, M. et al. (2021), "Machine learning in citizen science: Promises and implications" in *The Science of Citizen Science*, Springer, Cham.

Jackson, C. et al. (2020), "Teaching citizen scientists to categorize glitches using machine learning guided training", *Computers in Human Behavior*, Vol. 105/106198, https://doi.org/10.1016/j.chb.2019.106198.

Jones, F.M. et al. (2020), "Processing citizen science- and machine-annotated time-lapse imagery for biologically meaningful metrics", *Scientific Data*, Vol. 7/102, https://doi.org/10.1038/s41597-020-0442-6.

Joppa, L.N. (2017), "The case for technology investments in the environment", 19 December, *Nature*, www.nature.com/articles/d41586-017-08675-7.

Kelling, S. et al. (2012), "eBird: A human/computer learning network for biodiversity conservation and research", in *Proceedings of the Twenty-Fourth Innovative Applications of Artificial Intelligence Conference*, Vol. 26/2, AAAI Press, Palo Alto, https://doi.org/ 10.1609/aaai.v26i2.18963.

Kullenberg, C. and D. Kasperowski (2016), "What is citizen science? A scientometric meta-analysis", *PLOS ONE*, Vol. 11/1, p. e0147152, https://doi.org/10.1371/journal.pone.0147152.

MacDonald, E.A. et al. (2015), "Aurorasaurus: A citizen science platform for viewing and reporting the aurora", *Space Weather*, Vol. 13/9, pp. 548-559, https://doi.org/10.1002/2015SW001214.

McClure, E.C. et al. (2020), "Artificial intelligence meets citizen science to supercharge ecological monitoring", *Patterns*, Vol. 1/7, p. 100109, https://doi.org/10.3389/fmars.2022.918104.

Oliver, J.L. et al. (2019), "Listening to save wildlife: Lessons learnt from use of acoustic technology by a species recovery team", in *Proceedings of the 2019 Designing Interactive Systems Conference (DIS'19), 23-28 June, San Diego*, pp. 1335-1348, https://doi.org/10.1145/3322276.3322360.

Perry, T. et al. (2022), "EchidnaCSI: Engaging the public in research and conservation of the short-beaked echidna", *Proceedings of the National Academy of Sciences*, Vol. 119/5, p. e2108826119, https://doi.org/10.1073/pnas.2108826119.

Ponti, M. et al. (2021), “Can't we all just get along? Citizen scientists interacting with algorithms”, *Human Computation*, Vol. 8/2, pp. 5-14, https://doi.org/10.15346/hc.v8i2.128.

Rzanny, M. et al. (2022), “Image-based automated recognition of 31 Poaceae species: The most relevant perspectives”, *Frontiers in Plant Science*, Vol. 12, 26 January, https://doi.org/10.3389/fpls.2021.804140.

Sullivan, B.L. et al. (2014), “The eBird enterprise: An integrated approach to development and application of citizen science”, *Biological Conservation*, Vol. 169, pp. 31-40, https://doi.org/10.1016/j.biocon.2013.11.003.

Tuia, D. et al. (2022), “Perspectives in machine learning for wildlife conservation”, *Nature Communications*, Vol. 13/1, pp. 1-15, https://doi.org/10.1038/s41467-022-27980-y.

Unger, S, et al. (2021), “iNaturalist as an engaging tool for identifying organisms in outdoor activities”, *Journal of Biological Education*, Vol. 55/5, pp. 537-547, https://doi.org/10.1080/00219266.2020.1739114.

Van Horn, G. et al. (2018), “The iNaturalist species classification and detection dataset”, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,* Institute of Electrical and Electronic Engineers, Piscataway, https://authors.library.caltech.edu/87114/.

Wehn, U. et al. (2021), “Impact assessment of citizen science: State of the art and guiding principles for a consolidated approach”, *Sustainability Science*, Vol. 16/5, pp. 1683-1699, https://doi.org/10.1007/s11625-021-00959-2.

Wilkinson, M.D. et al. (2016), “The FAIR Guiding Principles for scientific data management and stewardship”, *Scientific Data*, Vol. 3/1, pp. 1-9, https://doi.org/10.1038/sdata.2016.18.

Willi, M. et al. (2018), “Identifying animal species in camera trap images using deep learning and citizen science”, *Methods in Ecology and Evolution*, Vol. 10/1, pp. 80-91, https://doi.org/10.1111/2041-210X.13099.

From:

Artificial Intelligence in Science

Challenges, Opportunities and the Future of Research

Access the complete publication at:
https://doi.org/10.1787/a8d820bd-en

Please cite this chapter as:

Ceccaroni, Luigi*, et al.* (2023), “Advancing the productivity of science with citizen science and artificial intelligence”, in OECD, *Artificial Intelligence in Science: Challenges, Opportunities and the Future of Research*, OECD Publishing, Paris.

DOI: https://doi.org/10.1787/69563b12-en