ARTICLE

Check for updates

Human-machine-learning integration and task allocation in citizen science

Marisa Ponti [1✉] & Alena Seredko[1]

The field of citizen science involves the participation of citizens across different stages of a scientific project; within this field there is currently a rapid expansion of the integration of humans and AI computational technologies based on machine learning and/or neural networking-based paradigms. The distribution of tasks between citizens ("the crowd"), experts, and this type of technologies has received relatively little attention. To illustrate the current state of task allocation in citizen science projects that integrate humans and computational technologies, an integrative literature review of 50 peer-reviewed papers was conducted. A framework was used for characterizing citizen science projects based on two main dimensions: (a) the nature of the task outsourced to the crowd, and (b) the skills required by the crowd to perform a task. The framework was extended to include tasks performed by experts and AI computational technologies as well. Most of the tasks citizens do in the reported projects are well-structured, involve little interdependence, and require skills prevalent among the general population. The work of experts is typically structured and at a higher-level of interdependence than that of citizens, requiring expertize in specific fields. Unsurprisingly, AI computational technologies are capable of performing mostly well-structured tasks at a high-level of interdependence. It is argued that the distribution of tasks that results from the combination of computation and citizen science may disincentivize certain volunteer groups. Assigning tasks in a meaningful way to citizen scientists alongside experts and AI computational technologies is an unavoidable design challenge.

[1] University of Gothenburg, Gothenburg, Sweden. ✉email: marisa.ponti@ait.gu.se

Introduction and background

Over the last few years we have witnessed a large growth in the capabilities and applications of artificial intelligence (AI) in citizen science (CS), a broad term referring to the active engagement of the general public in research tasks in several scientific fields, including astronomy and astrophysics, ecology and biodiversity, archeology, biology, and neuroimaging (Vohland et al., 2021). CS is an expanding field and a promising arena for the creation of human-machine systems with increasing computational abilities, as several CS projects generate large datasets that can be used as training materials for AI streams, such as machine-learning (ML) models and image and pattern recognition (Lintott and Reed, 2013). The integration of humans and AI applications can help process massive amounts of data more efficiently and accurately and monitor the results. Examples of AI applications in CS include image recognition as in iNaturalist (Van Horn et al., 2018) and Snapshot Serengeti (Willi et al., 2019), image recognition and classification to map human proteins in cells and tissues as in Human Atlas (Sullivan et al., 2018), and consensus algorithms to locate or demarcate objects as in Galaxy Zoo (Willett et al, 2013) and the Koster Seafloor Observatory (Anton et al., 2021). The integration of humans and computational technologies opens up new ways of collaboration between the two but also raises questions about the distribution of tasks, and how humans and this type of technologies can complement each other to expand their respective skills. As AI grows "smarter", people become increasingly concerned with being replaced in many domains of activities. Opportunities and risks for work of digital technologies including AI have been a long-standing topic of inquiry across research disciplines and in policy documents. In a recent report for the European Parliament, the Panel for the Future of Science and Technology (STOA) (2021) pointed out that technology impacts the distribution of tasks within jobs; just as technology may help to improve skills and raise the quality of work, it can also result in deskilling and creating low paid and low autonomy work. Importantly, while technology can help to preserve work, it can also impact negatively the qualitative experience of work (Panel for the Future of Science and Technology (STOA), 2021). In this respect, during a discussion panel aimed to initiate a dialog on how citizen scientists interact and collaborate with algorithms held at the 3rd European Citizen Science 2020 Conference, participants expressed concern with the possible negative impact of AI on the qualitative experience of participation in CS. As mentioned during that discussion, the current rapid progress in ML for image recognition and labeling, in particular the use of deep learning through convolutional neural networks (CNN) and generative adversarial networks, may be perceived as a threat to human engagement in CS. If computational technologies can confidently carry out the tasks required, citizen scientists may feel that there is no space for authentic engagement in the scientific process (Ponti et al., 2021). This concern suggests the tension that arises when "designing a human-machine system serving the dual goals of carrying out research in the most efficient manner possible while empowering a broad community to authentically engage in this research" (Trouille et al., 2019, p.1).

Why task distribution matters and aim of the paper. Task distribution between humans and machines has always been a crucial step in the design of human-machine systems and a main topic of research in human-automation interaction (e.g., Dearden et al., 2000; Hollnagel and Bye, 2000; Sheridan, 2000). Considered an "evergreen" topic, task allocation has been covered by a large body of literature in different fields, including cognitive engineering, human factors, and human-computer interaction, but

continues to be an important area for research on automation (Janssen et al., 2019). A prominent approach used for years to decide which tasks are better performed by machines or by humans has been the HABA-MABA ("Humans are better at, Machines are better at") list firstly introduced by Fitts (1951). This list contains 11 "principles" recommending the functions that are better performed by machines and should be automated, while the other functions should be assigned to humans. Although researchers differ in what they consider appropriate criteria for task allocation (Sheridan, 2000), the influence of Fitts's principles persists today in the human factors' literature. De Winter and Dodou (2014) conclude that Fitts's list is still "an adequate approximation that captures the most important regularity of automation" (p. 8). Given the primary interest of researchers to optimally distribute tasks between humans and machines to maximize efficiency and speed to achieve a given goal (Tausch and Kluge, 2020), this conclusion would arguably hold. However, as Tausch and Kluge (2020) noted, we need more research on task distribution in order to make decisions that allow not only an optimal allocation but also a rewarding experience for humans. This aspect is important in CS projects as they rely on volunteer engagement and concerns have been raised over the potential of AI to disengage citizen scientists: the use of AI can result in a reduction in the range of possible volunteer contributions or in the allocated tasks becoming either too simple or too complex (Trouille et al., 2019).

While task distribution to participants in citizen science projects has been studied by Wiggins and Crowston (2012), task distribution between experts, citizen scientists, and AI computational technologies (hereinafter also used interchangeably as "computational technologies") does not appear to have been investigated. Therefore, we present a literature review to illustrate the current state of the distribution of tasks between humans and computational technologies in CS. We used an adapted version of the framework developed by Franzoni and Sauermann (2014) to analyze the results and highlight the differences in the nature of the task and the skills contributed by humans and this type of machines to perform those tasks. Through the analysis, we answer the following questions:

1. What tasks do citizen scientists, experts, and computational technologies perform to achieve the goals of citizen science projects?
2. What type of skills do citizen scientists, experts, and computational technologies need to perform their tasks?
3. In which activities do citizen scientists, experts, and computational technologies, respectively, perform their tasks?

We now clarify the terms used in the questions. We use Theodorou and Dignum's (2020) definition of AI computational technologies as systems "able to infer patterns and possibly draw conclusions from data; currently AI technologies are often based on machine learning and/or neural networking-based paradigms" (p. 1). This definition is appropriate in this paper because our review is focused on technologies based on machine learning and/or neural networking-based paradigms. For a proper understanding of the term "task", we refer instead to Hackman's (1969, p. 113) definition of the term as a function assigned to an individual (or a group) by an external agent or that can be self-generated. A task includes a set of instructions that specify which operations need to be performed by an agent concerning an input, and/or what goal is to be achieved. We used Hackman's (1969) conceptualization of tasks as a behavior description, that is, a description of what an agent does to achieve a goal. This conceptualization applies to both humans and machines performing tasks. The emphasis is placed on the reported
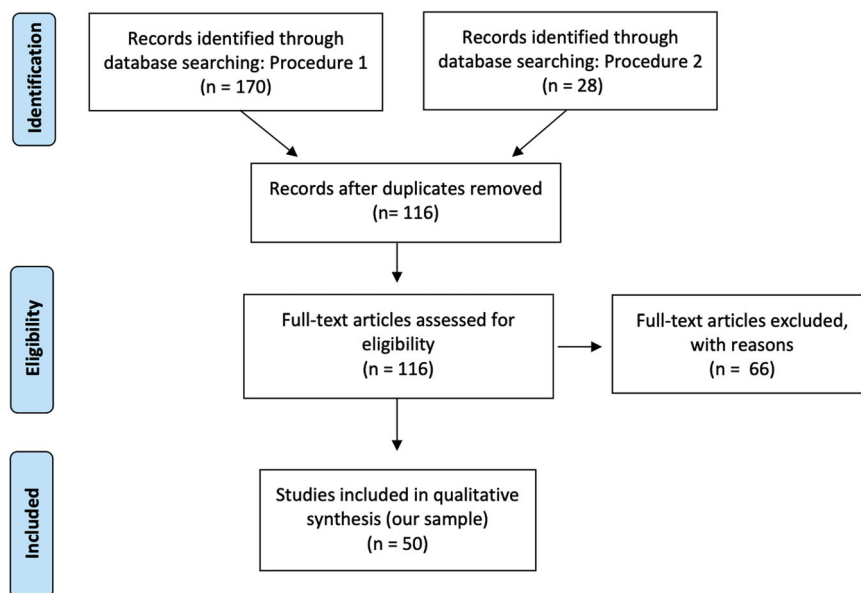
Fig. 1 Prisma flow diagram for paper selection.

behavior of the task performed. Regarding the activities included in our analysis, we discussed the tasks performed within activities of the research process such as data collection, data processing, and data analysis (McClure et al., 2020), as CS projects typically involve these activities. Regarding the term "expert", we used it in a broad sense, to include not only professional scientists but also persons responsible for developing algorithms and running the projects.

The contribution of this paper is threefold: (1) providing the scholars studying CS and human computation with a synthesis of results shedding descriptive insights into the distribution of tasks in CS; (2) potential broader implications for how we think about work in general and how we organize work involving "non-experts" and computational technologies, and 3) point to important questions for future research.

The paper is organized as follows. We first describe the methodology used for collecting and assessing the reviewed papers. We then present the framework used for our analysis of the results. Building on this framework, in the Discussion section we propose a matrix to classify CS projects on the basis of the distribution of tasks between humans and computational technologies. We also reflect on the role of citizen scientists borrowing from a labor perspective. The last section presents conclusions from this study and points to future research.

Methods

Strategies for searching, screening, and assessing the literature. To answer our questions, we conducted an integrative literature review, a subcategory of systematic reviews (Torraco, 2016). Integrative reviews follow the principles of systematic reviews to ensure systematization and transparency in searches, screening, and assessment processes, but they allow more flexibility regarding selection and inclusion of literature. The strategy for searching, screening and assessing the literature followed the systematic approach of the PRISMA reporting checklist (Fig. 1).

The review corpus was sourced from the Web of Science, SCOPUS, and the Association for Computing Machinery (ACM) Digital Library—the single largest source of computer science literature. These three databases are well-established, multi-disciplinary research platforms, including a wide variety of peer-reviewed journals, and they are updated regularly. We used them because they constitute a baseline for the search of published peer-reviewed papers. However, we are aware that, in the case of citizen science publications, the true extent of these publications can be larger as studies can be published in non-peer-reviewed literature sources and would not be referred in these three databases. We did not include preprints.

We employed two search procedures. In the first search procedure (Table 1), we searched for papers containing "citizen science" and "artificial intelligence" or "machine learning" in the title, abstract, and keyword sections. However, after a brief initial scanning of the resulting papers, we added several other search terms to include the most widely used computational technologies, such as supervised learning, unsupervised learning, reinforcement learning, reinforcement algorithm, deep learning, neural network(s), and transfer learning. The search was limited to papers written in English and published until July 2020. We collected a total of 170 papers across the three databases, 99 of which were unique. Table 1 summarizes the search terms used and the number of results per database during search procedure 1.

The initial examination of these papers revealed that the chosen search strategy did not fully cover those focused on citizen science games. The reason for this is that some authors use game titles in the abstracts instead of referring to games as 'citizen science'. Therefore, a second search procedure was employed using the same terms related to AI and ML as in the first procedure, and the titles of 36 citizen science games (Baert, 2019). The search was performed in two steps. First, the game titles 'Neo', 'Turbulence', and 'The Cure' were excluded, because the search generated many false-positive results not related to citizen science games. At the second step, to locate those articles that discuss 'Neo', 'Turbulence', and 'The Cure' games, an additional search was implemented using these titles and 'game' as search terms. The search using the ACM Digital Library did not produce any results, thus we excluded this database from the table. The searches were limited to articles written in English and published until July 2020. Table 2 contains the employed search strings and the number of generated results for both steps of the search procedure. The searches generated 28 results, 20 of which were unique. Out of the remaining 20 papers, 17 were not covered by the search procedure 1.

Exclusion criteria. A single reviewer parsed all the retrieved papers based on their titles, abstracts, keywords, and, if necessary,

Table 1 Search procedure 1.

Database	Search strings	Results
Web of Science	TOPIC: ("citizen science" OR "citizen scientist*") AND TOPIC: ("artificial intelligence" OR "machine learning" OR "supervised learning" OR "unsupervised learning" OR "reinforcement learning" OR "reinforcement algorithm" OR "deep learning" OR "neural network*" OR "transfer learning") Refined by: LANGUAGES: (ENGLISH) AND DOCUMENT TYPES: (ARTICLE) Timespan: All years. Indexes: SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH, ESCI.	83
SCOPUS	(TITLE-ABS-KEY ("citizen science" OR "citizen scientist*") AND TITLE-ABS-KEY ("artificial intelligence" OR "machine learning" OR "supervised learning" OR "unsupervised learning" OR "reinforcement learning" OR "reinforcement algorithm" OR "deep learning" OR "neural network*" OR "transfer learning")) AND (LIMIT-TO (DOCTYPE, "ar")) AND (LIMIT-TO (LANGUAGE, "English"))	86
ACM	[[Abstract: "citizen science"] OR [Abstract: "citizen scientist"] OR [Abstract: "citizen scientists"]] AND [[Abstract: "artificial intelligence"] OR [Abstract: "machine learning"] OR [Abstract: "supervised learning"] OR [Abstract: "unsupervised learning"] OR [Abstract: "reinforcement learning"] OR [Abstract: "reinforcement algorithm"] OR [Abstract: "deep learning"] OR [Abstract: "neural network"] OR [Abstract: "neural networks"] OR [Abstract: "transfer learning"]] Applied filters: Research article, Journals	1
Total		170

Databases, search strings, and result count.

Table 2 Search procedure 2 (citizen science games).

Database	Search strings	Results
Web of Science	TOPIC: (hexxed OR mobot OR "NeMO-net" OR hewmen OR "Skill Lab: Science Detective" OR "Cancer Crusade" OR "Crowd Water Game" OR mozak OR "Stall Catchers" OR "Colony B" OR "Decodoku Colors" OR decodoku OR "Sea Hero Quest" OR "MalariaSpot Bubbles" OR "Project Discovery" OR mark2cure OR questagame OR nanocrafter OR "Reverse The Odds" OR mequanics OR apetopia OR "Quantum Shooter" OR "Alien Game" OR "Quantum Moves" OR malariaspot OR "Forgotten Island" OR eyewire OR "Quantum Minds" OR nanodoc OR artigo OR phylo OR eterna OR foldit) AND TOPIC: ("artificial intelligence" OR "machine learning" OR "supervised learning" OR "unsupervised learning" OR "reinforcement learning" OR "reinforcement algorithm*" OR "deep learning" OR "neural network*" OR "transfer learning") Refined by: LANGUAGES: (ENGLISH) AND DOCUMENT TYPES: (ARTICLE) Timespan: All years. Indexes: SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH, ESCI.	9
Web of Science	TOPIC: (neo OR turbulence OR "The Cure") AND TOPIC: (game*) AND TOPIC: ("artificial intelligence" OR "machine learning" OR "supervised learning" OR "unsupervised learning" OR "reinforcement learning" OR "reinforcement algorithm*" OR "deep learning" OR "neural network*" OR "transfer learning") Refined by: LANGUAGES: (ENGLISH) AND DOCUMENT TYPES: (ARTICLE) Timespan: All years. Indexes: SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH, ESCI.	2
SCOPUS	(TITLE-ABS-KEY (hexxed OR mobot OR "NeMO-net" OR hewmen OR "Skill Lab: Science Detective" OR "Cancer Crusade" OR "Crowd Water Game" OR mozak OR "Stall Catchers" OR "Colony B" OR "Decodoku Colors" OR decodoku OR "Sea Hero Quest" OR "MalariaSpot Bubbles" OR "Project Discovery" OR mark2cure OR questagame OR nanocrafter OR "Reverse The Odds" OR mequanics OR apetopia OR "Quantum Shooter" OR "Alien Game" OR "Quantum Moves" OR malariaspot OR "Forgotten Island" OR eyewire OR "Quantum Minds" OR nanodoc OR artigo OR phylo OR eterna OR foldit) AND TITLE-ABS-KEY ("artificial intelligence" OR "machine learning" OR "supervised learning" OR "unsupervised learning" OR "reinforcement learning" OR "reinforcement algorithm*" OR "deep learning" OR "neural network*" OR "transfer learning")) AND (LIMIT-TO (DOCTYPE, "ar")) AND (LIMIT-TO (LANGUAGE, "English"))	15
SCOPUS	TITLE-ABS-KEY (neo OR turbulence OR "The Cure") AND TITLE-ABS-KEY (game*) AND TITLE-ABS-KEY ("artificial intelligence" OR "machine learning" OR "supervised learning" OR "unsupervised learning" OR "reinforcement learning" OR "reinforcement algorithm*" OR "deep learning" OR "neural network*" OR "transfer learning")) AND (LIMIT-TO (DOCTYPE, "ar")) AND (LIMIT-TO (LANGUAGE, "English"))	2
Total		28

Databases, search strings, and result count.

full-text reading. Since we aimed to focus on papers reporting on projects implementing integrations of human and computational technologies, as they were more likely to describe how tasks were distributed, we applied several selection criteria to filter out irrelevant papers. We excluded papers that provided no significant focus and discussion on, or application of, computational technologies in CS. Studies excluded from the review included, for example, a comparison of the results of citizen scientist classifications with the results of computational technology's models trained on expert-produced data (e.g., Wardlaw et al., 2018), or an overview of projects on Zooniverse with some description of how ML is used in these projects (Blickhan et al., 2018).

Table 3 presents the exclusion criteria and the number of excluded papers. As a result of the selection process, 50 papers were selected for the review (the list of all the included and excluded papers with reasons for why they were excluded is in Supporting Information 1, available at https://zenodo.org/record/5336431#.YSyzFtOA5EJ).

Analysis and synthesis strategy. The process of searching for relevant studies, filtering, data extraction, and synthesis took place from April 1st to September 30th 2020. To ensure consistency in the reporting, we used a spreadsheet for the 50 articles to note the author(s) and title for each article, the publication year, the source, the research field and type of the CS project.

Table 3 Exclusion criteria for procedure 1 and 2 and number of excluded papers.

Exclusion criteria	Count
Not related to AI computational technologies	9
Not related to citizen science	18
Citizen science and AI computational technologies not examined in combination	12
Citizen science data/procedure are simulated	3
Not related to AI computational technologies applied to CS	8
Irrelevant topics	17
Total	66

In addition, for each article we annotated the aim of the article, the computational technologies used, and the tasks assigned to citizens and experts (this extraction data is in Supporting Information File 1). These annotations provided a preliminary overview of the aspects relevant to address our research questions.

The framework

We used an adapted version of the conceptual framework developed by Franzoni and Sauermann (2014) to characterize citizen science projects with respect to two main dimensions: (a) the nature of the task outsourced to a crowd, and (b) the skills that crowd participants need to perform the task. We generalize this framework in order to map the tasks performed not only by the crowd but also by experts and AI computational technologies. We now describe the two dimensions in more details.

Nature of the task. Franzoni and Sauermann describe this dimension in an aggregate sense at the level of the crowd, with each individual making distinct contributions to that task by performing specific subtasks. They subsume under "nature of the task" two related attributes: the complexity of the task and the task structure. In the context of CS projects, Franzoni and Sauermann define *task complexity* as the degree of interdependency between the individual subtasks that participants perform when contributing to a project. In the simplest case, tasks are independent of each other. Task outputs can be pooled together in the end to generate the final output. In addition, contributors can perform their tasks independently. The authors made the example of Galaxy Zoo, where the correct classification of an image does not depend on the classifications of other images. However, when dealing with complex and interdependent tasks, the best solution to one subtask depends on other subtasks, so contributors must consider other contributions when working on their own subtask.

Franzoni and Sauermann define *task structure* as the degree to which the overall task outsourced to participants is well-structured, with clearly defined subtasks, or ill-structured, with specific subtasks not clearly defined from the start. Ill-structured tasks are said to provide fewer opportunities for the division of labor. The two attributes are typically highly correlated and we follow Franzoni & Sauermann by using a single dimension ranging from independent/well-structured to interdependent/ill-structured.

Skills need to perform the task. Franzoni and Sauermann distinguish three types of human skills: (a) common skills held in the general population, e.g., recognizing the shape of a plant; (b) specialized and advanced skills that are less common and not related to a specific scientific domain, e.g., specialized knowledge of certain technical tools, and (c) expert skills that are specifically related to a given scientific or technological domain, e.g., expert knowledge of mathematics.

This distinction is applicable to humans in general, including citizens and professional scientists. To be able to apply this framework to computational technologies as well, we expand this categorization to include different types of machine skills, in particular classification and prediction. We define them as the abilities to perform a task that computational technologies learn from training. We characterize these skills as being generated through programs of actions—consisting of goals and intentions – delegated by developers to technologies (Latour, 1994).

This framework forms a two-dimensional space that can categorize the tasks performed by citizen scientists, experts, and computational technologies, and the skills needed by humans and computational technologies to perform their tasks. We link these two dimensions to three main research activities/stages in the research process in which citizen scientists, experts, and computational technologies perform their tasks. The three activities include: data collection, data processing, and data analysis. We define them here as follows. Data collection refers to acquisition and/or recording of data. Data processing refers to actions aimed to organize, transform, validate, and filter data in an appropriate output form for subsequent use (for example, for a ML training model). Data classification and data validation are considered data processing actions. Data analysis refers to actions performed on data to describe facts, detect patterns, develop explanations, test hypotheses, and predict the distribution of certain items. Modeling species is considered here a data analysis action. We refer to data modeling as a process of analyzing data—for example, on species distribution—and their correlated variables —for example, bioclimatic variables—to identify areas where sensitive species exist or not.

The choice of the two dimensions is one aspect we should take into consideration. Our argument is that the nature of task and skill are fundamental, not just for mapping the distribution of tasks, but also to understand how tasks are allocated to citizens versus experts versus computational technologies. By mapping an array of papers, we expect to observe certain characteristics. For example, we can expect that tasks performed by computational technologies will be on the right side, but it would also be interesting to see if these technologies are mostly doing well-structured tasks and at what level of interdependence, while humans also do ill-structured tasks.

Results

The results section is divided into two parts. First, we provide a descriptive overview of the dataset, including some basic characteristics of the reviewed publications and the fields of the reported citizen science projects. Second, we organize the results around the three review questions of the study, concerned with: the nature of the task performed by citizens, experts, and computational technologies, the skills needed by both humans and computational technologies to perform theirs tasks, and the activities in which citizens, experts, and computational technologies, respectively, perform their tasks. The type of skills were mostly inferred from the description of the tasks, as the reviewed papers generally do not include explicit statements about this dimension. (Supporting Information 2 provides an overview of the tasks and the relative references).

Overview of the dataset. The reviewed papers were published between 2011 and 2020, with the majority (35 out of the 50 papers) published between 2018 and 2020 (Supporting Information 1 contains the list of the 50 papers). This increasing interest in combining AI computational technologies and citizen

science (CS) is also evident from the growing diversity of research fields with which the described CS projects are associated. The review demonstrated a considerable variety of citizen science projects ($n = 42$) with some papers reporting on using data from several projects. The three main areas that attract the most attention across the whole timespan are astronomy and astrophysics (e.g., Galaxy Zoo, Gravity Spy, and Supernova Hunters), biology (e.g., EteRNA, EyeWire, and Project Discovery), and ecology and biodiversity (e.g., eBird, Bat Detective). However, starting from 2017, we observe a larger variation including archeology (e.g., Heritage Quest, and field expeditions), neuroimaging (Braindr.us), seismology (MyShake app), and environmental issues (recruiting volunteers to measure the quality of air or water). Table 4 shows the distribution of the reviewed papers per research area from 2011 to 2020.

Citizen scientists: nature of the tasks performed, skills needed, and activities. The two main categories of tasks performed by citizen scientists are collecting data and classifying observations. Other tasks include generating new taxonomies, validating the algorithm classification results, solving in-game puzzles, and going through training.

Table 5 provides an overview of the main tasks performed by citizen scientists and the skills they need.

Data collection. This refers to a set of tasks widely assigned to citizen scientists in the areas of ecology, biodiversity, and environmental monitoring. Delegating the collection of data

to volunteers allows researchers to map geographical distributions of species and spatial variation in unprecedented scope and detail, which is especially relevant when monitoring by researchers is not feasible or efficient enough. The most common types of data contributed by volunteers include photos of plants or animals, accompanied by some context information (such as location and date/time of observation), and sometimes by a description (e.g., Derville et al., 2018; Capinha 2019). Less common types are videos and audio recordings (e.g., Zilli et al., 2014; Hardison et al., 2019). These observations were often accompanied by species classification, as citizen scientists were asked to submit observations of a particular species (e.g., Jackson et al., 2015). Alternatively, volunteers submitted observations that they classified with the help of an instrument, e.g., the eBird app (Curry et al., 2018), where the mobile app suggestions were used but no photo attached. Several papers also reported on citizens sending a specimen to researchers, e.g., bee trap nests (Kerkow et al., 2020; Everaars et al., 2011).

Another type is relatively passive data collection that does not require analysis on the part of the citizens. Lim et al. (2019) and Adams et al. (2020) reported on projects aimed at sampling air quality: volunteers were equipped with AirBeam sensors and asked to sample several routes by walking or cycling there. In Winter et al. (2019), an Android app was presented that allowed for identifying and classifying charged papers in camera image sensors. The only task outsourced to the citizens in this case was installing the app.

Overall, the reported tasks involve a low-level of complexity as taking good photos of animals or plants, or collecting quality data of air pollution do not depend on the data collected by other volunteers. Regarding the background skills to perform these tasks, citizen scientists seem to need general/common skills, required in routine tasks. However, they can be required to have some training, which was sometimes done face-to-face if citizens were asked to collect specific types of data in the field (Hardison et al., 2019). In other cases, this training occurred mainly online, as citizens went through guidelines prepared by project authors (Keshavan et al., 2019). The training could also be guided, facilitated, and assessed using ML algorithms (Zevin et al., 2017). While we may suggest that all citizen scientists go through some kind of training, not all of the reviewed papers included related information.

Table 4 Distribution of the reviewed papers per research area from 2011 to 2020.

Research area	2011–2020
Archeology	1
Astronomy and Astrophysics	16
Biology	4
Ecology, Biodiversity and Conservation	23
Environment	3
Neuroinformatics, Neuroimaging, Medicine	1
Recording wildlife	1
Seismology	1
Total	50

Table 5 Tasks performed by citizen scientists and skills needed across the activities—examples from the literature.

	CS project	Specific tasks performed	Nature of the tasks		Skills
			Task complexity	Task structure	
Data collection	Marine mammal observation network	Collect and annotate photos of plants and animals (Derville et al., 2018).	Low-level	Well-structured	Common skills (taking pictures)
	AirBeam	Passive sensing (Adams et al., 2020).	Low-level	Well-structured	Common skills (installing a sensor)
Data processing		Classify species (Jackson et al., 2015).	Low-level	Well-structured	Common skills (identify and count objects)
	Gravity Spy	Generate new taxonomies	Low-level	Well-structured	Specialized skills (identify new classes of objects)
	EyeWire	Map 3D structures of retinal neurons (Kim et al., 2014).	Low-level	Well-structured	Specialized skills (visualization and manipulation skills)
		Validate automatically detected archeological sites (Lambers et al., 2019)	Low-level	Well-structured	Common skills (identify objects)
Data analysis	EteRNA	Solve two-dimensional puzzles (Lee et al., 2014)	Low-level	Well-structured	Specialized skills (visualization and manipulation skills)

Data processing. The second popular set of tasks performed by citizen scientists is related to image analysis and includes classifying images into predefined categories, describing objects by choosing all relevant categories from a predefined list, as well as identifying and counting objects. The research fields setting up citizen science projects to outsource these tasks to volunteers include astronomy and astrophysics, ecology and biodiversity, archeology, biology, and neuroimaging. The tasks are performed in web interfaces: the majority of the projects run on the Zooniverse platform, but there are also separate initiatives such as the Braindr.us website (Keshavan et al., 2019) and the Project Discovery implemented in the Eve online game (Sullivan et al., 2018). Allocating classification tasks to citizen scientists is often related to the extremely large size of currently available datasets that makes expert classification unfeasible. The projects leverage human ability for pattern recognition and benefit from the scope of citizen science projects. The resulting classifications constitute training datasets for computational analysis. Citizen scientists classify objects from images into predefined categories. It can be a binary classification task, e.g., citizens decided whether a supernova candidate is a real or a 'bogus' detection (Wright et al., 2017; Wright et al., 2019). Alternatively, there could be a larger number of categories. For example, four studies reported on the Gravity Spy project, where users were presented with spectrograms and asked to classify glitches into predefined categories according to their morphology (Bahaadini et al., 2018; Crowston et al., 2020; Jackson et al., 2020; Zevin et al., 2017). Another task performed by citizen scientists was about describing an object in an image using a set of predefined characteristics. Examples include describing circumstellar debris disk candidates (Nguyen et al., 2018); classification of protein localization patterns in microscopy images (Sullivan et al., 2018); and morphological classification of galaxies (Jiménez et al., 2020; Kuminski et al., 2014; Shamir et al., 2016). Lastly, the projects benefiting from citizen scientists identifying and counting objects asked citizens to identify and locate animals of particular species (Bowley et al., 2019; Torney et al., 2019); mark potential archeological sites (Lambers et al., 2019), and identify and locate Moon craters (Tar et al., 2017), and interstellar bubbles on images (Beaumont et al., 2014; Duo and Offner 2017). Kim et al. (2014) reported on the EyeWire game project, where players contribute to mapping 3D structures of retinal neurons by coloring the area that belongs to one neuron and avoiding coloring other neurons on a 2D slice image. These types of task involve a low-level of complexity as the correct classification of an object does not depend on the classifications of other objects. To perform this type of classification, citizen scientists need common skills, such as identifying and counting objects. More specialized skills, such as good observation skills, can be required to perform tasks related to generating new taxonomies of objects. Coughlin et al. (2019) discussed that Gravity Spy project volunteers did not only classify spectrograms into already known classes of glitches but also suggested new classes, being aided by computational clustering of morphologically similar objects. Citizen scientists also performed validation of algorithm classification (Kress et al., 2018), or object detection results (Lambers et al., 2019). An example is the Leafsnap project where citizens submitted photos of leaves, and if the shape-matching algorithm did not classify the plant with high enough probability, citizens were offered several options to choose from (Kress et al., 2018). In the field of archeology, citizen scientists and heritage managers and/or academic researchers participated in field expeditions to validate discovered archeological objects detected by algorithms in remotely sensed data (Lambers et al., 2019).

Data analysis. In the reviewed papers on games, citizens can perform tasks that differ substantially from the tasks performed in other projects (such as classification). An example is reported by Koodli et al. (2019) and Lee et al. (2014), who discussed the EteRNA project, where players solve two-dimensional puzzles to design sequences that can fold into a target RNA structure. Lee et al. reported that EteRNA volunteer players outperformed previous algorithms in discovering RNA design rules.

Experts: nature of the tasks performed, skills needed, and activities. Tasks performed by experts are the most varied. They include collecting and processing the original data before it is presented to volunteers or algorithms, creating the gold-standard datasets, processing and curating the data collected or classified by citizen scientists, and preparing the training datasets for ML models. Several tasks are related to recruiting, training, and supporting volunteers. Finally, researchers are involved in the evaluation and validation of results. It is important to note that some tasks performed by researchers may not be discussed in papers in detail, since they occur naturally in every project, or because they may not be relevant for the discussion. Therefore, this section outlines only those tasks that are discussed in sufficient detail.

Table 6 provides some examples of the main tasks performed by experts and the skills they need.

Data collection. Several studies on biodiversity reported on researchers collecting observation data of species occurrence in the field (Derville et al., 2018; Jackson et al., 2015; Zilli et al., 2014). Researchers also obtained pre-classified data from external sources, such as the records of ladybirds sourced from the UK Biological Records Centre (Terry et al., 2020). These observations together with observational data collected by citizen scientists were further used to train and test computational technologies. When ML methods were used to predict species distribution or environmental conditions (e.g., coral bleaching), researchers were also responsible for sourcing data related to the characteristics of the environment. Examples of such data were mean temperature and precipitation (Capinha, 2019; Jackson et al., 2015), and geospatial data including roads and types of land usage (Lim et al., 2019).

Researchers involved in the development of citizen science projects were responsible for recruiting, training, and supporting volunteers. In those projects where volunteers were asked to collect data in a specific location (e.g., air quality measurements along certain routes, or coral bleaching measures on specific beaches), researchers recruited volunteers and performed face-to-face training (Adams et al., 2020; Hardison et al., 2019; Kumagai et al., 2018). When citizen participation was not bound to a particular space, volunteers received written guidelines (Bowley et al., 2019; Torney et al., 2019; Wright et al., 2017). Supporting user motivation and engagement was another task performed by researchers. Examples include ensuring that volunteers were involved in real classification tasks that led to the advancement of the project (Crowston et al., 2020). In projects that required volunteers to collect observations in the field, researchers followed up on citizens' contributions (Jackson et al., 2015; Kerkow et al., 2020), and provided online support and feedback (Lambers et al., 2019).

The tasks performed by experts regarding data collection generally require specialized skills to train citizens to use bespoke technologies like in the case of sampling a toxic microalga (Hardison et al., 2019), or to source data with certain environmental characteristics (e.g., Jackson et al., 2015).

Table 6 Tasks performed by experts and skills needed across the activities—examples from the literature.

	CS project	Specific tasks performed	Nature of the tasks		Skills
			Task complexity	Task structure	
Data collection	UK Ladybird	Obtain pre-classified data (Terry et al., 2020).	Medium-level	Well-structured	Expert skills
		Train citizens (Hardison et al., 2019)	Low-level	Well-structured	Specialized skills
		Source data with certain characteristics (Capinha, 2019).	Low-level	Well-structured	Expert skills
Data processing	Detection of *Karenia Brevis*	Train and test ML models (Crowston et al., 2020)	High-level	Well-structured	Expert skills
	Braindr.us	Create the gold-standard dataset (Keshavan et al., 2019)	High-level	Well-structured	Expert skills
	UK Ladybird	Classify data from citizen scientists (Terry et al., 2020)	Medium-level	Well-structured	Expert skills
	HeritageQuest	Assist citizens in validating objects (Lambers et al., 2019)	Medium-level	Well-structured	Specialized skills/ Expert skills
	Supernova Hunters	Decide on the number of volunteer votes to generate a final classification label for an image to be used to train ML algorithms (Wright et al., 2019)	Low-level	Well-structured	Expert skills
Data analysis	Pl@ntNet	Evaluate predictive accuracy of results (Botella et al., 2018)	High-level	Well-structured	Expert skills
	eBird	Compare the performance of different ML methods and statistical models to predict species distribution (Curry et al., 2018)	High-level	Well-structures	Expert Skills

Data processing. The data provided by citizen scientists, be it observations or classifications, was processed and curated by researchers. The observations provided by citizen scientists, such as cicada call recordings or ladybird recordings, were classified by field experts to be further used by an ML algorithm (Terry et al., 2020; Zilli et al., 2014). In some projects, original data obtained from cameras or sensors was preprocessed by researchers to be further presented to citizen scientists. For example, the audio recordings from bat observations were split into short sound clips and converted to spectrograms for the Bat Detective project (Mac Aodha et al., 2018), while in the Serengeti Wildebeest Count project, images from trap cameras were filtered to remove the empty ones and thereby reduce the number of images for citizen scientists to classify (Torney et al., 2019).

Other related tasks included processing citizen scientist contributions (e.g., returned bee nests) for future analysis (Everaars et al., 2011; Kerkow et al., 2020); deciding on the number of volunteer votes required before the final classification label for an image was generated and used to train or test ML algorithms (Sullivan et al., 2018; Wright et al., 2019; Wright et al., 2017); and choosing a limited amount of volunteer-produced data for training an algorithm (Koodli et al., 2019). Preparing the training dataset for ML also included tasks such as generating pseudo-absences when the information provided by volunteers only indicates presences observed (Jackson et al., 2015); generating synthetic observations of bubbles in dust emission to improve ML classification (Duo and Offner, 2017); or augmenting the training dataset by transforming existing images to increase the accuracy of ML classification (Dou and Offner, 2017). Performing initial training of algorithms, or calibrating and fine-tuning machine-learning performance involve a high-level of complexity because they depend on the classifications of data done by citizens or experts.

Expert classifications refer to the so called "gold-standard", which is a quality dataset approved as the most accurate and reliable of its kind and that could be used to measure accuracy and reliability of algorithm results. The development of a gold-standard can be considered a high-level complexity task, as it usually relies on multiple experts agreeing on classifying certain topics with a high degree of certainty. Expert classifications were used to perform the initial training of the algorithm (Crowston et al., 2020; Jackson et al., 2020), to calibrate and fine-tune the machine-learning performance (Beaumont et al., 2014; Jiménez et al., 2020), or to provide the testing set for computational classification methods (Crowston et al., 2020; Tar et al., 2017). Expert classifications were also included in the guidelines for volunteers (Keshavan et al., 2019), used to assess the accuracy of citizen scientists' classifications and give feedback to volunteers (Jackson et al., 2020; Zevin et al., 2017), as well as to weight each citizen scientist's vote in the final label based on how much their labels corresponded to the gold-standard set (Keshavan et al., 2019).

Data analysis. Experts were involved in the evaluation of results generated by ML using citizen data. A low error level demonstrated the viability of involving citizen scientists to produce training data for ML. For example, researchers evaluated the predictive accuracy of species distribution models based on the automated identification of citizen observations using CNN (Botella et al., 2018), and the climatic niche of invasive mosquitoes using a support vector machine (Kerkow et al., 2020). Bowley et al. (2019) reported on comparing the results of ML training using citizen data and using expert classifications. Other authors such as Curry et al. (2018) and Jackson et al. (2015) reported on comparing the performance of different ML methods to predict species distribution using citizen data.

Furthermore, the results of ML classifications were compared with manual classifications done by field experts (Nguyen et al., 2018; Pearse et al., 2018; Wright et al., 2017). A similar approach was reported on by Kress et al. (2018) in relation to the Leafsnap, where a deep learning algorithm was used to define the contours of a leaf and visual recognition software was employed to find an ordered set of possible matches for it in the database. However, Leafsnap participants needed to confirm the classification suggestions made by the algorithm. Therefore, in this project, the validation of

Table 7 Tasks performed by AI computational technologies and skills needed across the activities—examples from the literature.

	CS project	Specific tasks performed	Nature of the tasks		Skills
			Task complexity	Task structure	
Data collection					
Data processing	Galaxy Zoo 1	Classify galaxy images using CNN (Jiménez et al., 2020)	High-level	Well-structured	Object recognition
	Gravity Spy	Clustering galaxy images using transfer learning (Coughlin et al., 2019)	High-level	Well-structured	Object recognition
	EteRNA	Design molecules using CNN (Koodli et al., 2019)	High-level	Well-structured	Prediction
	AirBeam	Mitigate errors and biases using automated ML (Adams et al., 2020)	High-level	Well-structured	Prediction
	FreshWater Watch	Predict water quality using a regression model (Thornhill et al., 2017)	High-level	Well-structured	Prediction
	Heritage Quest	Detect objects in remotely sensed data using CNN (Lambers et al., 2019).	High-level	Well-structured	Object recognition
	Coral Map	Predict coral bleaching using a regression model (Kumagai et al., 2018).	High-level	Well-structured	Prediction
	Serengeti Wildebeest Count	Count wildebeests on images using deep learning (Torney et al., 2019).	High-level	Well-structured	Object recognition
Data analysis	Galaxy Zoo	Evaluate consistency of volunteer annotations (Shamir et al., 2016).	High-level	Well-structured	Pattern recognition

accuracy referred to the results from both citizen scientists and ML models. Unique among these tasks were the validation procedures reported by Lambers et al. (2019), where experts and citizen scientists together validated the new potential archeological objects identified using ML by going into the field.

AI computational technologies: nature of the tasks performed, skills needed, and activities. In the reviewed papers, there is a variety of computational technologies using machine learning and/or neural network-based paradigms. Interested readers can find more details about the types of technologies and their reported use in the Supplemental Information 1 (Annotated review articles). These technologies used several common ML methods such as classification, regression, transfer learning, deep learning, and clustering. For the definitions of these methods, we refer the readers to Castañón (2019).

The skills developed by computational technologies can be grouped in two main categories: recognition and prediction. Recognition refers to classification and detection of objects in images, but also to clustering to classify data into specific groups. Classification and object detection are the most popular tasks performed by these technologies in a variety of projects in the fields of ecology and biodiversity, astronomy and astrophysics. Classification and object detection use various ML algorithms (e.g., Brut algorithm based on random forest, CNN), which are often based on a supervised paradigm and consist of two main steps: training a classifier with samples that are considered as "gold-standard" (expert classifications) or volunteer consensus data ("ground truth") or a combination of both, and testing the effectiveness of the classifier using other samples, ensuring that none of the samples of the test set are also used for training.

Prediction refers to making predictions of future outcomes by given data, or reducing the errors of a model. Prediction tasks include predicting environmental conditions (e.g., air quality, or variations in the data); addressing biases in the original data or in citizen scientist classification and detection results; improving performance by learning from citizen scientist contributions, modeling species geographical distribution, and learning from

player moves in a game. Table 7 provides some examples of the main tasks performed by technologies and the skills they need.

Data processing. Examples of classification and object detection include a convolutional neural network and a residual neural network trained on both citizen scientist and expert labels to classify galaxy images (Jiménez et al., 2020), or a deep learning algorithm trained on Serengeti Wildebeest Count project data used for counting wildlife in aerial survey images (Torney et al., 2019). It was argued that with the limited number of citizen scientists and increasingly large databases of images, computational technologies offer an approach to scale up data processing, overcome the analysis 'bottleneck' problem, and also relieve some burden from researchers and citizen scientists who would only have to classify enough images for training ML, rather than the whole dataset (Torney et al., 2019; Wright et al., 2017).

Clustering is another task performed by ML (Coughlin et al., 2019; Wright et al., 2019). Coughlin et al. (2019) reported on DIRECT, a transfer learning algorithm, which is a ML method consisting of reusing a model previously developed for a different task. The aim of DIRECT was facilitating the discovery of new glitches by citizen scientists in the Gravity Spy project. Owing to the sheer volume of available images, it is extremely difficult for volunteers to identify new classes by finding a sufficient number of similar objects that do not belong to any of the known classes. Thus, DIRECT clustered similar images together and offered this set to volunteers to make their judgment. Wright et al. (2019) reported on using Deep Embedded Clustering (DEC), a method that learns feature representations and cluster assignments, to produce an initial grouping of similar images. In the Supernova Hunters project, grouped images were shown to citizen scientists, who had to mark all of the objects belonging to one glitch class. Then, citizen scientists' labels were fed back to the DEC algorithm to make clustering purer. Compared to the standard image-by-image presentation, Wright et al. found that the DEC model helped reduce volunteer effort to label a new dataset to about 18% of the standard approach for gathering labels.

Lambers et al. (2019) also reported on using contributions from citizen scientists to improve the results of a multi-class archeological object detection based on CNN. In their project,

volunteers participated in field expeditions to validate arche-ological objects detected by the algorithm, and the results were used to tune the algorithm object detection results. In other fields, computational technologies such as random forest classification (Thornhill et al., 2017), stacked ensemble model (Lim et al., 2019), and generalized linear model (Kumagai et al., 2018) have been used on data collected by citizen scientists and environ-mental or urban data collected by scientists, to predict water quality, air quality, and coral bleaching respectively. A CNN has been used to model the distribution of species in biodiversity research, such as the White-tailed Ptarmigan distribution over Vancouver Island (Jackson et al., 2015) or the Asian bush mosquito distribution area (Kerkow et al., 2020). The datasets used for training were usually combined from different sources: observations collected and reported by citizen scientists and sometimes by experts as well, and environmental or climate data extracted by the researchers. In another project, Adams et al. (2020) used an automated ML process to adjust AirBeam sensor measurements, which showed errors during times of high humidity. They employed a temporal adjustment algorithm to mitigate biases or errors in the data collected and/or classified by citizen scientists.

In the area of games, the papers on EteRNA emphasize the development of algorithmic approaches that learn from player actions. Lee et al. (2014) developed and trained the EteRNABot algorithm, incorporating machine-learning regression with five selected and cross-validated design rules discovered by players and used to predict the design of RNA structures. An EternaBrain convolutional neural network (CNN) trained on expert moves was described by Koodli et al. (2019). Based on the test results, the algorithm achieved accuracy levels of 51% in base prediction and 34% in location prediction, indicating that top players' moves were sufficiently stereotyped to allow a neural network to predict moves with a level of accuracy much higher than chance (p. 2). *Data Analysis.* A concern addressed through the use of AI computational technologies is that citizen scientists' participation in data collection may not be uniformly distributed in space and can be skewed toward capturing observations rather than absences. For example, Derville et al. (2018) compared five species distribution models to see how they account for the sampling bias present in nonsystematic citizen science observations of humpback whales. Other papers also discussed employing transfer learning when only a small training dataset was available. An example is Willi et al. (2019), who developed a model based on the data from the Snapshot Serengeti citizen science project, and then applied it in another project where only smaller datasets were available to improve accuracy.

Another reason for applying AI computational technologies is that volunteers may misclassify data due to their varying expertize levels and proneness to human error. The issue is addressed in several ways. Tar et al. (2017) evaluated the false-positive contamination in the Moon Zoo project by utilizing predictive error modeling. Keshavan et al. (2019) compared the citizen science ratings to the golden standard created by experts. In the Galaxy Zoo project, Shamir et al. (2016) proposed using a pattern recognition algorithm to evaluate the consistency of annotations made by individual volunteers.

To measure the expertize of eBird volunteer observers, Kelling et al. (2012) used a probabilistic machine-learning approach. In their study, they used the occupancy-detection experience model to measure the probability of a given species being detected at a given site, and to distinguish expert observers from novice observers who are more likely to misidentify common bird species. Researchers used this approach to provide volunteers with feedback on their observation accuracy and to improve a training dataset for an ML algorithm. In addition to eBird, mentioned above, Gravity Spy is one of the few projects that used feedback and training to evaluate citizen contributions (Crowston et al., 2020; Jackson et al., 2020; Zevin et al., 2017). Volunteers were guided through several training levels with the ML system: first showing glitches belonging to two classes with a high-level of ML-determined probability, and later increasing the number of classes and offering images with lower ML confidence scores as they learned to classify them.

A project profile matrix based on distribution of work: discussion

The interdependence of AI computational technologies and domain experts. The results of the review guided our develop-ment of a matrix to classify CS projects on the basis of our adaptation of Franzoni and Sauermann's (2014) framework. Table 8 presents a summary of examples. Since ill-structured tasks were not found in the reported projects, this sub-dimension is not included in the matrix.

We have plotted some of these examples on two axes in Fig. 2. The horizontal axis represents the nature of the task and is broken into three levels of interdependence, while the vertical axis represents ths skills requirements, and is also broken into three.

While task complexity is largely defined by Franzoni and Sauermann (2014) from a social perspective, the degree of complexity of tasks changes in CS projects using computational technologies (i.e., experts perform tasks with a high-level of complexity, while citizens and technologies can be assigned tasks with a lower level of complexity). In our review we found that a large share of the CS projects involve citizens in performing tasks that tend to be of low complexity, well-structured, and requiring only skills that are common among the general population. Classifying species or marking potential archeological sites, for example, can involve a large number of individuals working in parallel, independently. At a medium-level of complexity is a moderately complex and relatively well-structured task, such as solving two-dimensional puzzles, which citizens perform in the EteRNA game. In this task, players work individually or in groups to explore different solutions collaboratively. Despite the fact that the game can be played without background scientific knowledge, success lies in visualizing and manipulating the design interface to create two-dimensional structures that can include complex patterns such as lattices, knots and switches (Lee et al., 2014). Moreover, players seem to adapt their puzzle strategies based on the results of the laboratory experiments (Lee et al., 2014). The case suggests some interdependence, as the results reached by individual players can be aggregated into a single outcome, which can be referred to as additive/pooled coordination (Nakatsu et al., 2014), and then reused/adapted by other players.

Unlike citizen scientists, experts primarily work on well-structured and medium- and high-level tasks that require expert skills in specific domains. For example, trained neuroimaging experts created cooperatively a gold-standard dataset to be used by citizens for "amplifying" expert decisions (Keshavan et al., 2019). In this task, expert workers are highly interdependent and are expected to consider what each is doing. Interdependence is demonstrated again by the involvement of trained life scientists who collaborate to compare the predictive accuracy of ML species distribution models. The tasks performed by experts and reported in the reviewed literature did not require common skills. Now let us consider the tasks that computational technologies perform. Unsurprisingly, we see that these tasks are on the right side of the diagram. The results of our study indicate that these technologies are capable of performing mostly well-structured, high-level tasks. Tasks appear interdependent in a sequential manner or reciprocal

Table 8 Project profile matrix—examples.

	Human skills			AI computational technologies skills	
	Common	**Specialized**	**Expert**	**Recognition**	**Prediction**
Nature of the task Low-level of complexity —Well-structured tasks	Data collection - Audio recording and taking photos (Citizens) Data Processing - Classification (Citizens)	Data collection - Train citizens to collect data (Experts) Data processing - Map 3D structures of retinal neurons (Citizens) Data Analysis - Solve two-dimensional puzzles	Data processing - Assist citizens in validating objects (Specialized/Experts) Data Processing - Volunteer data classification (Experts) Data analysis - Solve two-dimensional puzzles (Citizens)		
Medium-level of complexity —Well-structured tasks			Data collection - Obtain pre-classified data (Experts) - Assist citizens in validating objects (Experts)		
High-level of complexity —Well-structured tasks			Data processing - Gold-standard dataset creation (Experts) Data analysis - Validate results (Experts) - Compare different ML and statistical approaches (Experts)	Data Processing - Classification - Clustering objects - Count objects on images - Object detection Data Analysis - Evaluate consistency of citizen annotations	Data Processing - Mitigate errors and biases - Design of molecules - Predict environmental conditions

manner. Sequential interdependence takes place when the output of a task serves as the input into another (Haeussler and Sauermann, 2015). This seems to be the case when computational technologies mitigate errors in the data provided by citizens. Reciprocal interdependence refers to tasks that depend on each other and can require a mutual adjustment (Haeussler and Sauermann, 2015). For example, when performing tasks like clustering and classifying images, or predicting environmental conditions, these technologies need to build on human work in a reciprocal fashion, as they must be trained on specific datasets to develop a predictive model and then deploy it. Then experts need to check and validate the results produced by the algorithmic model and adjust such a model when necessary.

The skills of computational technologies make them a scalable complement to citizens and researchers, for example by structuring large amounts of unfiltered data into information, or estimating the probability of an occurrence of an event based on input data. However, to assume that machine learning and other computational technologies can replace humans entirely in citizen science is to downplay their current limited autonomy and "smartness", as they still require human intervention of experts and engaged citizens. The distribution of tasks resulting from the review indicates that experts work *on* and *with* computational technologies. For example, they work *on* them by training models, but once models are trained, they still require a human expert-in-the-loop to work *with* them to interpret their predictions and possibly refine them to acquire the most accurate results for unseen and unknown data (Budda et al., 2021).

Having examined the tasks performed by computational technologies and the rationale on which functions are allocated to them in CS projects, we can infer mechanisms that make certain tasks more suitable for existing computational methods. According to these mechanisms, Brynjolfsson and Mitchell (2017)

set criteria to identify tasks that are likely to be suitable for ML, based on the currently dominant paradigm, particularly supervised learning. Brynjolfsson and Mitchell's criteria include (a) learning a function that maps well-defined inputs to well-defined outputs, as in the classification of images, and the prediction of the likelihood of events; (b) the task provides a clear feedback, and goals and metrics for performance are clearly defined. When training data are labeled according to gold standards, for example, ML is particularly powerful to achieve its set goals; (c) ML excels when learning empirical associations in data, but is less successful when long chains of reasoning or complex planning require common sense or background knowledge that is unknown to the computer; (d) tolerance to errors of the learned system, as most ML algorithms derive their solutions statistically and probabilistically. As a result, it is seldom possible to train them to obtain total accuracy - even the best object recognition computer systems make errors, and (e) performing tasks where the inability of ML to explain why or how they made a certain decision is not critical. Brynjolfsson and Mitchell made the example of systems capable of diagnosing types of cancer as well as or better than expert doctors, but unable to explain why or how they came up with the diagnosis.

However, ML will be advanced, and other methods will be more suitable for different tasks. Cognitive work between humans and computational technologies will be shifting, challenging the ontological boundaries between them. Hence, we should be careful not to essentialize the qualities of humans and machines, both of which are constantly evolving, and whose lists of what each is "good at" (whether relative or absolute) are constantly changing. The processing power and the sophistication of algorithms have already increased at previously unimaginable levels, and, for example, some computer programs outperform humans in abstract games (Gibney, 2016), or at image recognition
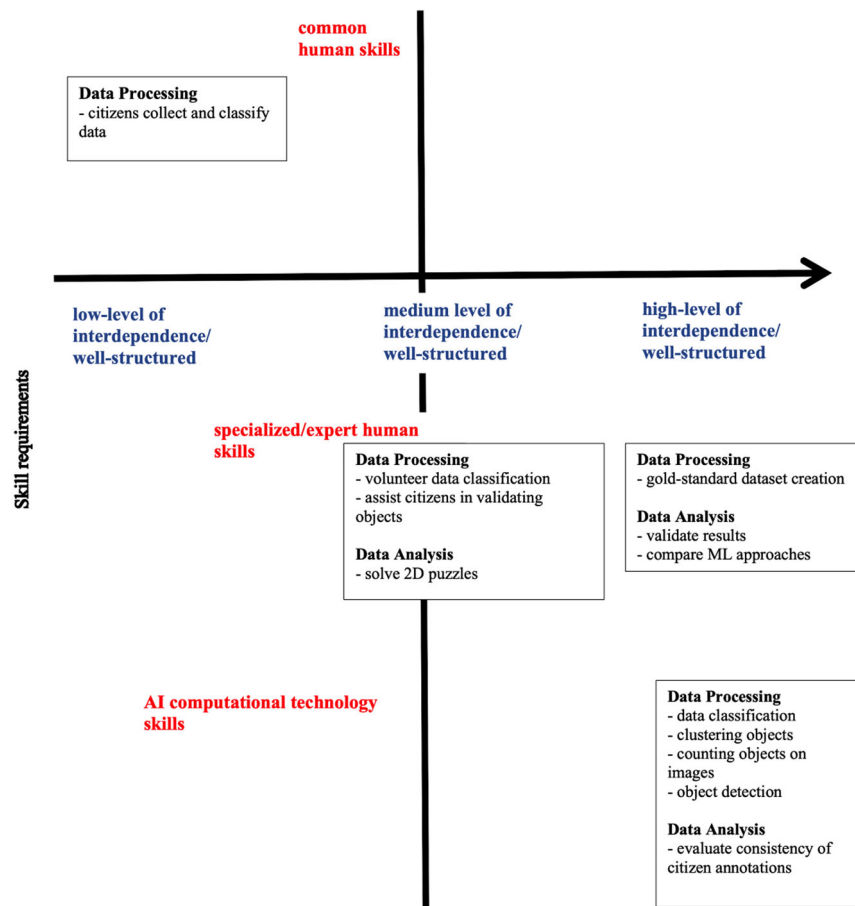
Fig. 2 Nature of the task outsourced to humans and AI computational technologies.

(Johnson, 2015). However, some scholars might argue that these are rather narrow domains, which cannot compare to the complexities of cognitive, emotional, and social human abilities (e.g., Dignum, 2019).

Scientists and AI computational technologies: will the role of citizens become unnecessary? A large share of the CS projects involve citizens in performing tasks in contributory projects that are "designed by scientists and for which members of the public primarily contribute data" (Shirk et al., 2012). The results of our review indicate a trend towards task polarization, with citizens performing well-structured and low-complexity tasks requiring primarily common skills, and experts performing well-structured and higher-level of complexity tasks requiring training and specialization. As technology races ahead, both types of task seem susceptible to computerization though, with both citizens and experts being reallocated to tasks that are not or less susceptible to computerization, i.e., tasks requiring creative and social intelligence. To this regard, the differentiation between task and skill made by Autor et al. (2003) is useful: task denotes a unit of activity performed at work and it produces output, while the concept of skill refers to the human capabilities required to perform a task. The Routine-Biased Technological Change (RBTC) (Arntz et al., 2016) approach builds on this differentiation and analyzes tasks according to the routine and nonroutine axis. Following this approach, a job's substitutability is determined by the number of routine tasks it requires, as opposed to the level of skills it needs (Arntz et al., 2016). Routine tasks can be performed both manually or cognitively, while nonroutine tasks, also known as abstract tasks, involve problem-solving, intuition, and creativity. Routine tasks

that follow a well-defined practice can be more easily codified and performed automatically by algorithms. In the reviewed CS projects, routine tasks, such as collecting data, counting or demarcating objects in images, seem to be prevalent—although not exclusively—in citizens' contributions. Even classification of objects following authoritative taxonomies can be considered a routine task that can be codified and performed by algorithms. Almost any task in CS projects reliant on pattern recognition is susceptible to automation, as long as adequate data are collected for training algorithms (Frey and Osborne, 2013).

Citizens are more likely to be involved in nonroutine tasks when playing games. Exemplary is Foldit where players have the opportunity to use and write and adapt recipes to manage the increasing complexity of the game (Cooper et al., 2011). Recipes are computer programs that allow players to interact automatically with a protein and repeat simple routines consistently, or perform a series of complex routines, which keep running in the background for ever. Although recipes embed a number of simple, time-consuming and repetitive manual actions, they have not yet replaced the skills that citizens learn over time, through training and playing the game intensively, and that are needed to perform nonroutine tasks in the game (Ponti et al., 2018). When it comes to designing RNA sequences that fold into particular shapes, computational technologies have proved to be second to humans, even after being endowed with insights from the human folders (Lee et al., 2014).

Regarding the use of skills, automation literature authors (e.g., Brynjolffson and Mcafee, 2016; Eurofound 2018; Goos et al., 2019) commonly assert that technological developments have upgraded the skill requirements for occupations by complementing the skills

of highly skilled, highly educated professionals. However, these technologies have lowered the demand for lower-skilled, less educated workers, as the tasks they perform are more susceptible to replacement by technologies. In the context of CS, though, it is not clear whether the use of algorithms could result in the disappearance of "low skill" roles and, by any means, it is unclear what we refer to when we talk about low-skilled or unskilled work in this specific area. Even tasks that are considered low-skill may not mean they can be done easily by computational technologies.

Conclusion and future research

A growing number of CS projects use human efforts and computational technologies. Yet the distribution of tasks among experts, citizen scientists, and this type of technologies does not seem to have been considered. Even though task allocation has long been a central topic in the literature, we are unaware of previous studies examining this topic in the context of citizen science.

We summarized the results of an integrative review to illustrate the current state of the distribution of tasks between humans and computational technologies in CS. We used an adapted version of the framework developed by Franzoni and Sauermann (2014) to analyze the results and highlight the differences in the nature of the task and the skills contributed by humans and computational technologies to perform those tasks. We hope that this framework may provide a useful tool for analyzing task allocation not just in citizen science but also in other contexts where knowledge production involves a crowd and computational technologies. The presented framework may also be useful not only for "mapping" projects, but also for inferring underlying causal mechanisms, such as, for example, that computational technologies seem to be better at certain tasks. An important next step would be to learn why certain CS projects cannot be entirely performed algorithmically and still require human contribution, while others are already suited for full automation (Rafner et al., 2021).

Although we conducted this review to include all relevant papers, there can be papers we did not include as we left out preprints, reports, and other types of non-peer-reviewed literature. Future research could consider using a larger number of databases, publication types and publication languages, in order to widen the scope of the review. Furthermore, we have two limitations related to our sampling strategy. First, we did not search catalogs of CS projects (e.g., Scistarter). Our current approach to search on publication databases has only retrieved articles that have been written about projects. This strategy may have introduced certain biases—for example, we may not have captured projects run by non-scientists (since they may not care to publish a paper about their project). We can assume that most research papers report on successful rather than unsuccessful projects, meaning we have been exposed to mostly successful divisions of labor instead of divisions that did not work. Second, we acknowledge that using Baert's (2019) game list has resulted in selection bias. As to our knowledge, there is not a complete sampling frame listing all the existing citizen science games from which to draw a representative sample. Therefore, we opted for a convenience sample, where the sample is taken from the game list, in addition to the articles selected for this review. Specifically, our convenience sample is a purposive sample, as we relied on our judgment when choosing the games to include in the sample (Fricker, 2012).

Overall, the projects reported in the literature put an emphasis on process optimization. The use of computational technologies presents opportunities in citizen science to improve speed, accuracy, and efficiency to analyze massive datasets and improve scientific discovery. However, as mentioned earlier in this paper, concerns have been raised over the potential risks of disengaging citizen scientists by reducing the range of their possible contributions or making them either too simple or too complex (Trouille et al., 2019; Leach et al., 2020). If citizens come to think that the only thing they are good at can be replaced by machine learning, they may feel left out and useless (Ponti et al., 2021).

Citizen science projects are not ordinary workplaces. Unpaid citizens volunteer time and effort, therefore, deriving personal meaning and value from performing a task is important to sustain engagement. Arguably, if the organizers of CS projects focus primarily on efficiency and productivity goals, they may replace citizens as much as possible, and only make tasks as meaningful as needed to keep volunteers engaged. In contrast, if organizers also want to achieve "democratization" goals, they will use computational technologies more for the benefit of human engagement and may even use citizens if those technologies could do a more efficient job. Currently, the difference may not matter much because computational technologies are not yet capable to replace humans entirely. However, this difference can become critical once AI is more powerful—then organizers will have to decide if they intend to maximize efficiency by replacing citizens, or try to achieve/maximize engagement by keeping citizens in the loop and use computational technologies to make tasks more interesting and meaningful for people. Future research could look more closely on whether the use of computation technologies benefits a variety of citizens or only certain groups (e.g., those with more technical expertize), and whether certain cultural domains are more facilitated.

We are aware that one size does not fit all. A boring task for one person can be a joy for another, while some volunteers may prefer to engage their brains and choose more complex tasks. Nevertheless, taking into account meaningful roles that citizen scientists can play alongside experts and computational technologies, for example in the form of additional data validation and other types of human-based analysis that strengthen analytical rigor or the diversity of analytical lenses, remains an unavoidable design issue of task assignment. It may be useful to complement the focus on efficiency and speed with greater attention to other goals such as volunteer learning and development, an issue that becomes particularly salient when we think about division of labor in the context of democratization of science, diversity, and inclusion, which are long-standing challenges in citizen science (Cooper et al., 2021).

Data availability

References

Adams MD, Massey F, Chastko K, Cupini C (2020) Spatial modelling of particulate matter air pollution sensor measurements collected by community scientists while cycling, land use regression with spatial cross-validation, and applications of machine learning for data correction. Atmos Environ 230:117479. https://doi.org/10.1016/j.atmosenv.2020.117479

Anton V, Germishuys J, Bergström P, Lindegarth M, Obst M (2021) An open-source, citizen science and machine learning approach to analyse subsea movies. Biodiver Data J 9:e60548. https://doi.org/10.3897/BDJ.9.e60548

Arntz M, Gregory T, Zierahn U (2016) The risk of automation for jobs in OECD countries: a comparative analysis. OECD Social, Employment and Migration Working Papers No. 189

Autor DH, Levy F, Murnane RJ (2003) The skill content of recent technological change: an empirical exploration. Q J Econ 118(4):1279–1333

Baert C (2019) Citizen science games list. Available via https://citizensciencegames.com/games/. Accessed 18 Apr 2020

Bahaadini S, Noroozi V, Rohani N, Coughlin S, Zevin M, Smith JR et al. (2018) Machine learning for Gravity Spy: glitch classification and dataset. Inf Sci (Ny) 444:172–86. https://doi.org/10.1016/j.ins.2018.02.068

Beaumont CN, Goodman AA, Kendrew S, Williams JP, Simpson R (2014) The Milky Way Project: leveraging citizen science and machine learning to detect interstellar bubbles Astrophys J Suppl Ser 214(1):3, http://www.tinyurl.com/yymgqpye. Accessed 5 Feb2021

Blickhan S, Trouille L, Lintott CJ (2018) Transforming research (and public engagement) through citizen science. Proc Int Astron Union 14(A30):518–23. https://doi.org/10.1017/S174392131900526X. (Section 4)

Botella C, Joly A, Bonnet P, Monestiez P, Munoz F (2018) Species distribution modeling based on the automated identification of citizen observations. Appl Plant Sci 6(2):e1029. https://doi.org/10.1002/aps3.1029

Bowley C, Mattingly M, Barnas A, Ellis-Felege S, Desell T (2019) An analysis of altitude, citizen science and a convolutional neural network feedback loop on object detection in unmanned aerial systems. J Comput Sci 34:102–16. https://doi.org/10.1016/j.JOCS.2019.04.010

Brynjolffson E, Mcafee A (2016) The second machine age: work, progress, and prosperity in a time of brilliant technologies. W.W. Norton, London

Brynjolfsson E, Mitchell T (2017) What can machine learning do? Workforce implications. Science 358(6370):1530–1534. https://doi.org/10.1126/science.aap8062

Budda S, Robinson EC, Kainz B (2021) Survey on active learning and human-in-the-loop deep learning for medical image analysis. Med Image Anal 71:102062. https://doi.org/10.1016/j.media.2021.102062

Capinha C (2019) Predicting the timing of ecological phenomena using dates of species occurrence records: a methodological approach and test case with mushrooms. Int J Biometeorol 63(8):1015–24. https://doi.org/10.1007/s00484-019-01714-0

Castañón J (2019) 10 machine learning methods that every data scientist should know. In: Towards Data Science. Available via Medium. https://towardsdatascience.com/10-machine-learning-methods-that-every-data-scientist-should-know-3cc96e0eeee9. Accessed 7 Jul 2021

Cooper CB, Hawn CL, Larson LR, Parrish JK, Bowser G et al. (2021) Inclusion in citizen science: the conundrum of rebranding. Science 372(6549):1386–1388. https://doi.org/10.1126/science.abi6487

Cooper S, Khatib F, Makedon I, Lu H, Barbero J, Baker D et al (2011) Analysis of social gameplay macros in the Foldit Cookbook. In: FDG'11, Proceedings of the 6th International Conference on Foundations of Digital Games, ACM, New York. pp. 9–14

Coughlin S, Bahaadini S, Rohani N, Zevin M, Patane O, Harandi M et al (2019) Classifying the unknown: discovering novel gravitational-wave detector glitches using similarity learning. Phys Rev D [Internet] 99(8). https://doi.org/10.1103/PhysRevD.99.082002

Crowston K, Osterlund C, Lee TK, Jackson C, Harandi M, Allen S et al. (2020) Knowledge tracing to model learning in online citizen science projects. IEEE Trans Learn Technol 13(1):123–134. https://doi.org/10.1109/TLT.2019.2936480

Curry CM, Ross JD, Contina AJ, Bridge ES (2018) Varying dataset resolution alters predictive accuracy of spatially explicit ensemble models for avian species distribution. Ecol Evol 8(24):12867–78. https://doi.org/10.1002/ece3.4725

de Winter JCF, Dodou D (2014) Why the Fitts list has persisted throughout the history of function allocation. Cogn Technol Work 16(1):1–11. https://doi.org/10.1007/s10111-011-0188-1

Dearden A, Harrison M, Wright P (2000) Allocation of function: scenarios, context and the economics of effort. Int J Hum Comput Stud 52(2):289–318

Derville S, Torres LG, Iovan C, Garrigue C (2018) Finding the right fit: comparative cetacean distribution models using multiple data sources and statistical approaches. Divers Distrib 24(11):1657–73. https://doi.org/10.1111/ddi.12782

Dignum V (2019) Responsible artificial intelligence. How to develop and use AI in a responsible way. Springer Nature, Cham Switzerland

Duo X, Offner SSR (2017) Assessing the performance of a machine learning algorithm in identifying bubbles in dust emission. Astrophys J 851(2):149. https://doi.org/10.3847/1538-4357/aa9a42

Eurofound (2018) Automation, digitisation and platforms: implications for work and employment. Publications Office of the European Union, Luxembourg

Everaars J, Strohbach MW, Gruber B, Dormann CF (2011) Microsite conditions dominate habitat selection of the red mason bee (Osmia bicornis, Hymenoptera: Megachilidae) in an urban environment: a case study from Leipzig, Germany. Landsc Urban Plan 103(1):15–23

Fitts PM (1951) Human engineering for an effective air-navigation and traffic-control system. Division of National Research Council, Oxford, England

Franzoni C, Sauermann H (2014) Crowd Science: the organization of scientific research in open collaborative projects. Res Pol 43(1):1–20. https://doi.org/10.1016/j.respol.2013.07.005

Frey CB, Osborne M (2013) The future of employment: how susceptible are jobs to computerisation? [Online]. University of Oxford. https://www.oxfordmartin.ox.ac.uk/downloads/academic/future-of-employment.pdf. Accessed Feb 18 2020

Fricker RD (2012) Sampling methods for web and e-mail surveys. In: Fielding N, Lee RM, Blank G (eds) The SAGE Handbook of Online Research Methods. SAGE Publications, London, pp. 195–216

Gibney E (2016) Google AI algorithm masters ancient game of Go. Nature 529:445–446. (28 Jan 2016)

Goos M, Arntz M, Zierahn U, Gregory T, Carretero Gomez S, Gonzalez Vazquez I, Jonkers K (2019) The impact of technological innovation on the future of work. JRC Working Papers on Labour, Education and Technology 2019–03, European Commission, Joint Research Centre

Hackman JR (1969) Toward understanding the role of tasks in behavioral research. Acta Psychol 31:97–128. https://doi.org/10.1016/0001-6918(69)90073-0

Haeussler C, Sauermann H (2015) The anatomy of teams: division of labour in collaborative knowledge production. Academy of Management Annual Meeting Proceedings. https://doi.org/10.5465/ambpp.2015.11383abstract

Hardison DR, Holland WC, Currier RD, Kirkpatrick B, Stumpf R, Fanara T et al. (2019) HABscope: A tool for use by citizen scientists to facilitate early warning of respiratory irritation caused by toxic blooms of Karenia brevis. PLoS ONE 14(6):e0218489. https://doi.org/10.1371/journal.pone.0218489

Hollnagel E, Bye A (2000) Principles for modelling function allocation. Int J Hum Comput. Stud 52(2):253–265

Jackson MM, Gergel SE, Martin K(2015) Citizen science and field survey observations provide comparable results for mapping Vancouver Island White-tailed Ptarmigan (Lagopus leucura saxatilis) distributions Biol Conserv 181:162–172. https://doi.org/10.1016/j.biocon.2014.11.010

Jackson C, Østerlund C, Crowston K, Harandi M, Allen S, Bahaadini S, Coughlin S, Kalogera V, Katsaggelos A, Larson S, Rohani N, Smith J, Trouille L, Zevin M (2020) Teaching citizen scientists to categorize glitches using machine learning guided training. Comput Human Behav 105:106198

Janssen CP, Donker SF, Brumby DP, Kun AL (2019) History and future of human-automation interaction. Int J Hum Comput Stud 131:99–107

Jiménez M, Torres MT, John R, Triguero I (2020) Galaxy image classification based on citizen science data: a comparative study. IEEE Access 8:47232–47246. https://doi.org/10.1109/ACCESS.2020.2978804

Johnson RC (2015) Microsoft, Google beat humans at image recognition. EENews Europe. Available at: https://www.eenewseurope.com/news/microsoft-google-beat-humans-image-recognition

Kelling S, Gerbracht J, Fink D, Lagoze C, Wong W-K, Yu J et al. (2012) A human-computer learning network to improve biodiversity conservation and research. AI Mag 34(1):10. https://doi.org/10.1609/aimag.v34i1.2431

Kerkow A, Wieland R, Früh L, Hölker F, Jeschke JM, Werner D et al. (2020) Can data from native mosquitoes support determining invasive species habitats? Modelling the climatic niche of Aedes japonicus japonicus (Diptera, Culicidae) in Germany. Parasitol Res 119(1):31–42. https://doi.org/10.1007/s00436-019-06513-5

Keshavan A, Yeatman JD, Rokem A (2019) Combining citizen science and deep learning to amplify expertise in neuroimaging. Front Neuroinform [Internet] 13. https://doi.org/10.3389/fninf.2019.00029

Kim JS, Greene MJ, Zlateski A, Lee K, Richardson M, Turaga SC et al. (2014) Space–time wiring specificity supports direction selectivity in the retina. Nature 509(7500):331–336. https://doi.org/10.1038/nature13240

Koodli RV, Keep B, Coppess KR, Portela F, Das R, Eterna participants (2019) EternaBrain: automated RNA design through move sets and strategies from an Internet-scale RNA videogame. PLoS Comput Biol 15(6):e1007059. https://doi.org/10.1371/journal.pcbi.1007059

Kress WJ, Garcia-Robledo C, Soares JVB, Jacobs D, Wilson K, Lopez IC et al. (2018) Citizen science and climate change: Mapping the range expansions of native and exotic plants with the mobile app leafsnap. Bioscience 68(5):348–358. https://doi.org/10.1093/biosci/biy019

Kumagai NH, Yamano H, Committee Sango-Map-Project (2018) High-resolution modeling of thermal thresholds and environmental influences on coral bleaching for local and regional reef management. PeerJ 6:e4382. https://doi.org/10.7717/peerj.4382

Kuminski E, George J, Wallin J, Shamir L (2014) Combining human and machine learning for morphological analysis of galaxy images. Publ Astron Soc Pac 126(944):959–67. https://doi.org/10.1086/678977

Lambers K, Verschoof-van der Vaart W, Bourgeois Q (2019) Integrating remote sensing, machine learning, and citizen science in Dutch archaeological prospection. Remote Sens 11(7):794. https://doi.org/10.3390/rs11070794

Latour B (1994) On technical mediation: philosophy, sociology, genealogy. Common Knowl 94(4):29–64

Leach B, Parkinson S, Lichten CA et al. (2020) Emerging developments in citizen science: reflecting on areas of innovation. RAND Corporation, Santa Monica, CA

Lee J, Kladwang W, Lee M, Cantu D, Azizyan M, Kim H et al. (2014) RNA design rules from a massive open laboratory. Proc Natl Acad Sci USA 111(6):2122–7. https://doi.org/10.1073/pnas.1313039111

Lim CC, Kim H, Vilcassim MJR, Thurston GD, Gordon T, Chen L-C et al. (2019) Mapping urban air quality using mobile sampling with low-cost sensors and machine learning in Seoul, South Korea. Environ Int 131:105022. https://doi.org/10.1016/j.envint.2019.105022

Lintott C, Reed J (2013) Human computation in citizen science. In: Michelucci P (ed) Handbook of human computation. Springer, New York, NY, p 153–162. https://doi.org/10.1007/978-1-4614-8806-4

Mac Aodha O, Gibb R, Barlow KE, Browning E, Firman M, Freeman R et al. (2018) Bat detective—Deep learning tools for bat acoustic signal detection. PLoS Comput Biol 14(3):e1005995. https://doi.org/10.1371/journal.pcbi.1005995. Bottom of Form

McClure EC, Sievers M, Brown CJ, Buelow CA, Ditria EM, Hayes MA et al. (2020) Artificial intelligence meets citizen science to supercharge ecological monitoring. Patterns (NY) Oct 9 1(7):100109. https://doi.org/10.1016/j.patter.2020.100109

Nakatsu RT, Grossman EB, Iacovou CL (2014) A taxonomy of crowdsourcing based on task complexity. J Inf Sci 40(6):823–834. https://doi.org/10.1177/0165551514550140

Nguyen T, Pankratius V, Eckman L, Seager S (2018) Computer-aided discovery of debris disk candidates: a case study using the Wide-Field Infrared Survey. Explorer (WISE) catalog. Astron Comput 23:72–82. https://doi.org/10.1016/j.ascom.2018.02.004

Panel for the Future of Science and Technology (STOA) (2021) Digital automation and the future of work. European Parliamentary Research Service 656:311. https://doi.org/10.2861/826116

Pearse WD, Morales-Castilla I, James LS, Farrell M, Boivin F, Davies TJ (2018) Global macroevolution and macroecology of passerine song. Evolution 72(4):944–60. https://doi.org/10.1111/evo.13450

Ponti M, Stankovic I, Barendregt W, Kestemont B, Bain L (2018) Chefs know more than just recipes: professional vision in a citizen science game. Hum Comput 5(1):1–12. 10.15346/hc.v5i1

Ponti M, Kloetzer L, Ostermann FO, Miller G, Schade S (2021) Can't we all just get along? Citizen scientists interacting with algorithms. Hum Comput 8(2):5–14. https://doi.org/10.15346/hc.v8i2.128

Rafner J, Gajdacz M, Kragh G, Hjorth A, Gander A, Palfi B et al. (2021) Revisiting citizen science through the lens of hybrid intelligence. arXiv:2104.14961 [cs.HC]. Available at https://arxiv.org/pdf/2104.14961.pdf

Shamir L, Diamond D, Wallin J (2016) Leveraging pattern recognition consistency estimation for crowdsourcing data analysis. IEEE Trans Hum Mach Syst 46(3):474–80. https://doi.org/10.1109/THMS.2015.2463082

Sheridan TB (2000) Function allocation: algorithm, alchemy or apostasy? Int J Hum Comput Stud 52(2):203–16. https://doi.org/10.1006/ijhc.1999.0285

Shirk JL, Ballard HL, Wilderman CC, Phillips T, Wiggins A, Jordan R, McCallie E, Minarchek M, Lewenstein BV, Krasny ME (2012) Public participation in scientific research: a framework for deliberate design. Ecol Soc 17(2):29. https://doi.org/10.5751/ES-04705-170229

Sullivan DP, Winsnes CF, Åkesson L, Hjelmare M, Wiking M, Schutten R et al. (2018) Deep learning is combined with massive-scale citizen science to improve large-scale image classification. Nat Biotechnol 36(9):820–8. https://doi.org/10.1038/nbt.4225

Tar PD, Bugiolacchi R, Thacker NA, Gilmour JD, MoonZoo Team (2017) Estimating false positive contamination in crater annotations from citizen science data. Earth Moon Planets 119(2–3):47–63. https://doi.org/10.1007/s11038-016-9499-9

Tausch A, Kluge A (2020) The best task allocation process is to decide on one's own: effects of the allocation agent in human–robot interaction on perceived work characteristics and satisfaction. Cogn Tech Work. https://doi.org/10.1007/s10111-020-00656-7

Terry JCD, Roy HE, August TA (2020) Thinking like a naturalist: Enhancing computer vision of citizen science images by harnessing contextual data. Methods Ecol Evol 11(2):303–15. https://doi.org/10.1111/2041-210X.13335

Theodorou A, Dignum V (2020) Towards ethical and socio-legal governance in AI. Nat Mach Intell 2:10–12. https://doi.org/10.1038/s42256-019-0136-y

Thornhill I, Ho JG, Zhang Y, Li H, Ho KC, Miguel-Chinchilla L et al. (2017) Prioritising local action for water quality improvement using citizen science; a study across three major metropolitan areas of China. Sci Total Environ 584–585:1268–1281. https://doi.org/10.1016/j.scitotenv.2017.01.200

Torney CJ, Lloyd-Jones DJ, Chevallier M, Moyer DC, Maliti HT, Mwita M et al. (2019) A comparison of deep learning and citizen science techniques for counting wildlife in aerial survey images. Methods Ecol Evol 10(6):779–87. https://doi.org/10.1111/2041-210x.13165

Trouille L, Lintott CJ, Fortson LF (2019) Citizen science frontiers: efficiency, engagement, and serendipitous discovery with human-machine systems. Proc Natl Acad Sci USA 116(6):1902–1909. https://doi.org/10.1073/pnas.1807190116

Van Horn G, Oisin MA, Yang S, Cui Y, Sun C, Shepard A et al (2018) The iNaturalist species classification and detection dataset. In: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018. https://doi.org/10.1109/CVPR.2018.00914

Vohland K, Land-zandstra A, Ceccaroni L, Lemmens R, Perelló J, Ponti M, et al., (eds.) (2021) The science of citizen science. Springer, Cham, CH. https://doi.org/10.1007/978-3-030-58278-4

Wardlaw J, Sprinks J, Houghton R, Muller J-P, Sidiropoulos P, Bamford S, Marsh S (2018) Comparing experts and novices in Martian surface feature change detection and identification. Int J Appl Earth Obs Geoinf 64:354–364. https://doi.org/10.1016/j.jag.2017.05.014

Wiggins A, Crowston K (2012) Goals and tasks: two typologies of citizen science projects. In: Proceedings of the 45th Hawaii International Conference on System Sciences (HICSS), IEEE. https://doi.org/10.1109/HICSS.2012.295

Willett KW, Lintott CJ, Bamford SP, Masters KL, Simmons BD, Casteels KRV et al. (2013) Galaxy Zoo 2: detailed morphological classifications for 304 122 galaxies from the Sloan Digital Sky Survey. Mon Not R Astron Soc 435(4):2835–60. https://doi.org/10.1093/mnras/stt1458

Willi M, Pitman RT, Cardoso AW, Locke C, Swanson A, Boyer A et al. (2019) Identifying animal species in camera trap images using deep learning and citizen science. Methods Ecol Evol 10(1):80–91. https://doi.org/10.1111/2041-210X.13099

Winter M, Bourbeau J, Bravo S, Campos F, Meehan M, Peacock J et al. (2019) Particle identification in camera image sensors using computer vision. Astropart Phys 104:42–53. https://doi.org/10.1016/j.astropartphys.2018.08.009

Wright DE, Fortson L, Lintott C, Laraia M, Walmsley M (2019) Help me to help you: machine augmented citizen science. ACM Trans Soc Comput 2(3):1–20. https://doi.org/10.1145/3362741

Wright DE, Lintott CJ, Smartt SJ, Smith KW, Fortson L, Trouille L et al. (2017) A transient search using combined human and machine classifications. Mon Not R Astron Soc 472(2):1315–1323. https://doi.org/10.1093/mnras/stx1812

Zevin M, Coughlin S, Bahaadini S, Besler E, Rohani N, Allen S et al. (2017) Gravity Spy: integrating advanced LIGO detector characterization, machine learning, and citizen science. Class Quantum Gravity [Internet] 34(6). https://doi.org/10.1088/1361-6382/aa5cea

Zilli D, Parson O, Merrett GV, Rogers A (2014) A hidden Markov model-based acoustic cicada detector for crowdsourced smartphone biodiversity monitoring. J Artif Intell Res 51:805–827. https://doi.org/10.1613/jair.4434

Acknowledgements

Funding

Competing interests

The authors declare no competing interests.

Ethical approval

Not applicable

Informed consent

Not applicable

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1057/s41599-022-01049-z.

Correspondence and requests for materials should be addressed to Marisa Ponti.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.