What can machine learning do? Implications for citizen scientists

Anna Jia Gander, Department of Applied Information Technology, University of Gothenburg, Gothenburg, Sweden, anna.jia.gander@ait.gu.se

Alena Seredko, Department of Education, Communication and Learning, University of Gothenburg, Gothenburg, Sweden, alena.seredko@gu.se

Marisa Ponti, Department of Applied Information Technology, University of Gothenburg, Gothenburg, Sweden, marisa.ponti@ait.gu.se, ORCID ID: http://orcid.org/0000-0003-4708-4048

ABSTRACT

Citizen science projects set up in research fields such as astronomy, ecology and biodiversity, biology, and neuroimaging produce large datasets; thus they hold promise for applying artificial intelligence for the social and environmental spheres. Human-machine integration in citizen science can harness the contributions of many human observers and use machine learning (ML) to process their contributed data. Several citizen science projects have designed complex human-machine systems, taking advantage of the complementarity of the strengths of humans and machines, and aiming to optimize for efficiency and human engagement. Using document analysis of 12 citizen science projects deploying ML techniques to optimize classification tasks, we describe the distribution of work between citizens and researchers and between humans and algorithms, as well as configurations of human-in-the-loop. The results indicate that experts are involved in every aspect of the loop, from annotating or labeling data to giving them to algorithms to train and make decisions from such predictions. Experts also test and validate models to improve accuracy by scoring their outputs when algorithms are not able to make the right decisions. While experts are the humans mainly involved in the loop, citizens are also involved at various stages of the process. We present three main examples of citizens-in-the-loop: (a) when algorithms provide incorrect suggestions; (b) when algorithms do not know to perform classification, and (c) when algorithms are active learners. We contend that, unlike automated systems that tend to remove or reduce the need for humans, the examined projects are heteromated systems that do not function without the indispensable human mediation of engaged citizens.

Keywords: citizen science; classification; function allocation; heteromation; human-machine integration; machine learning.

1. INTRODUCTION

All over in the world, for several years now, many members of the general public have collaborated with professional scientists in research and data gathering in what is known as “citizen science”. Although involving amateurs in science is a several centuries-old practices, the term *citizen science* was coined in the 1990s by Irwin (1995) and has gained attention since from policy makers, academia, civil society, and the media. Members of the public – which we call here “citizen scientists,” or simply “citizens” to be meant not as citizens of nation states but as “members of a broadly construed community” (Eitzel et al. 2017, p. 6) – can participate in different types of citizen science and associated initiatives in several research fields. Citizen science projects are set up in astronomy and astrophysics, ecology and biodiversity, archeology, biology, and neuroimaging, among the others. For example, in ecology, citizens use sensors to contribute to data collection programs and monitor air or water quality, while in astronomy they classify galaxies. Citizen science projects often create large-scale observational datasets including citizen-generated images crowdsourced through smartphone apps, or galactic data collected by astronomers with telescopes. These data can benefit both science and society. For example, large-scale data can be used to complement official data sources to improve the Sustainable Development Goals reporting (Fritz et al. 2019). AI can help identify knowledge gaps, create awareness, and expand the dialogue on relevant issues.

AI has been used in citizen science for about 20 years (Ceccaroni et al. 2019). It is increasingly used to classify data and improve their quality (for example, by providing hints to volunteers based on automatic recognition of species from photographs) (Fritz et al. 2019). A subfield of AI, machine learning (ML), has been able to learn input/output relationships from data for years and thus solve problems such as classification or regression tasks with high accuracy. As artificial intelligence grows" smarter", people become increasingly concerned with being replaced in many domains of activities. A question on a hypothetical AI takeover in citizen science was also raised by the participants at the 3rd European Citizen Science 2020 Conference (https://www.ecsa-conference.eu/), during a discussion panel aimed to initiate a dialogue on how citizen scientists interact and collaborate with algorithms. As mentioned during a presentation given by Miller (2020), the current rapid progress in machine learning for image recognition and labeling, in particular the use of deep learning through convolutional neural networks, generative adversarial networks, and more, presents an obvious threat to human engagement in citizen science; if machines can confidently carry out the work required, then there can be no space for authentic engagement in the scientific process.
Several citizen science projects have designed complex human-machine systems, trying to take advantage of the complementarity of the strengths of humans and machines, and optimize for efficiency and human engagement (Trouille et al. 2019). Complementary abilities of both humans and machines need to be identified and leveraged to increase the accuracy and efficiency of the system (Kelling et al. 2013).

A crucial step in the design of human-machine systems is function allocation (a term used interchangeably with task allocation in this paper), that is, deciding which tasks or jobs should be allocated to humans and which ones to machines (de Winter and Dodou 2014). For decades, the standard method to decide which tasks are better performed by machines or 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. While the influence of these principles persists today in the human factors' literature (de Winter and Dodou 2014), HCI scholars have criticized Fitt's attributes for being static and insensitive to context dependency (Sheridan, 2000), and for neglecting the consequences of allocation decisions on the cooperative nature of complex work situations (Rognin, Salembier, and Zouinar 2000). The practical utility for the design of conventional methods used to allocate tasks has also been debated in HCI by Wright, Dearden, and Fields (2000), who suggested the use of naturalistic studies as well as abstract representations to understand "articulations of work" in ordinary work settings.

In this paper, we do not aim to contribute to the design of human-machine systems, nor do we discuss or propose criteria for task allocation in citizen science classification projects. The question we address here is simpler: *when designing a citizen science classification project involving humans and algorithms, which tasks or functions are allocated between citizens and researchers and between humans and algorithms?* Typically, the focus on allocating tasks between humans and algorithms is related to an increasing endeavor to automate parts of human contributions. The use of ML presents opportunities in citizen science to improve speed, accuracy, and efficiency to analyze massive datasets and improve scientific discovery. However, 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 (Leach, Parkinson, Lichten, and Marjanovic, 2020). Citizen science projects are not ordinary workplaces. Unpaid participants volunteer time and effort, therefore deriving personal meaning and value from performing a task is important to sustain engagement.

Our primary goal for this paper is to describe the distribution of work between citizens and researchers and between humans and algorithms, as reported in documents about classification projects. ML has been used in several projects to improve the classification of plants, animals, and galaxies, among others. The term "classification" refers to a single unit of output in a project, e.g., the coding of one image or the tagging of a video (Sauermann and Franzoni 2015). To respond to our question, we surveyed 12 citizen science classification projects to see what we could learn about function allocation between humans – specifically citizen scientists and experts (i.e., researchers) – and machines. In examining each project using documentary analysis, we described task allocation and configurations of the human-machine interplay.

Our secondary goal is to raise questions about human-machine integration in citizen science and set directions for future research. A better understanding of human-machine integration in citizen science could be also relevant for the field of collective intelligence, which is increasingly interested in combining human intelligence with AI (Mulgan 2018).

2. BACKGROUND

While task allocation to participants in citizen science projects has been studied by Wiggins and Crowston (2012), function allocation between experts, non-experts, and machines in projects using human-machine systems does not appear to have been investigated. Kelling et al. (2012) described the challenges to be addressed for human-machine systems to succeed in citizen science projects. They pointed out that tasks should be identified that humans can complete but machines cannot complete on their own, and should be sufficiently uncomplicated to motivate citizen scientists to participate. They also posited that the complementary abilities of both humans and machines should be identified to be leveraged to improve the accuracy and efficiency of the system.
Despite the growing attention, our empirical understanding of human-machine integration in citizen science appears to be limited. The existing literature has used case examples to illustrate projects in which a combination of humans and ML performs data-centered tasks (Willi et al. 2019; Sullivan et al. 2018). We can single out three major types of projects employing both human and machine efforts:

1. Projects that relate to classifying objects or observations, when the large size of a dataset makes expert classification unfeasible (e.g., Nguyen et al. 2018; Lukic et al. 2018).

2. Projects that benefit from citizen scientists' ability to collect data in the field covering large territories, while ML approaches are used to predict the distribution of species or probability of phenomenon occurrence (e.g., Jackson et al. 2015; Robinson et al. 2018).

3. Projects focused on clustering data to discover new classes (Coughlin et al. 2019; Wright et al. 2019). In contrast to the first two types of projects, where citizen science data is used to 'help' an algorithm, here an algorithm is used to 'help' citizen scientists. For example, Coughlin et al. (2019) employed a transfer learning algorithm to quantify similarities between Gravity Spy images, which allowed citizen scientists to search for glitches of similar morphology facilitating the identification of new classes. These types of projects are the fewest in number.

Collecting, coding/classifying, and interpreting data are some of the most common activities that participants carry out, depending on their level of engagement in the scientific research process (Shirk et al. 2012). Similarly, ML is used at various stages of the data–science life cycle through algorithms that perform tasks like classification, regression, clustering, and association, especially when dealing with massive datasets.

3. MATERIALS AND METHOD

We selected a convenience sample of classification projects. Note that the selected projects reflect those that were documented at the moment in time, rather than a truly representative sample of the population. We used document analysis (Bowen 2009) of the sampled projects. The selection of documents was based on two main criteria: there must have been an implementation – or a proof-of-concept – of the application and the texts must have been produced by personnel directly involved in the design and development of the project. We manually compiled and summarized data from a host of published sources, such as documents retrieved through websites, research articles, reports, and blogs. The used sources are referenced in Appendix 4. We collected all data in 2020.

We created a spreadsheet containing summarized data about the following 12 classification projects: Galaxy Zoo AI, Virus Spot, Multiple Sclerosis, Human Atlas, Plantsnap, MAIA (Machine Learning Assisted Image Annotation), iNaturalist, Milky Way, Twittersuicide, Mindcontrol, Observation.org, and Snapshot Serengeti. For each project, we described the tasks performed by citizen scientists, experts, and algorithms, respectively; the types of algorithms used; the sequence of tasks between humans and machines, and why the project combined humans and machines [data summaries are available in Appendix 1 and the anchors pointing to the reference sources used in the summaries are in Appendix 2]. For a proper understanding of the term *task*, we refer to

Hackman's (1969, p. 113) definition of the term as a job assigned to a person (or group) by an external agent or that can be self-generated. A task includes a set of instructions which specify which operations need to be performed by a person 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. The emphasis is placed on the reported behavior of the task performer. This conceptualization applies to both humans and machines performing tasks.

We coded the summarized data using qualitative content analysis (QCA) (Hsieh and Shannon 2005, p.1278) with NVivo 12 software (QSR 2020). We open coded the text describing the tasks performed by citizens, experts, and machines, and categorized the codes based on their conceptual similarity.

4. RESULTS

Table 1 illustrates the main characteristics of the sampled projects and exemplars of tasks.

Project	**Application category**	**Field of Study**	**Project Goal**	**Reason for human-machine integration**	**ML Model**	**Citizen tasks - Examples**	**Researcher tasks - Examples**	**Machine tasks - Examples**
Mindcontrol	Recognition	Neuroscience	Scale expertise in neuroimaging.	Train machines to emulate neuroimaging expert decision making.	Classification - Convolutional neural network	Amplify expert decisions; review machine annotations	Validate classifications; develop a gold-standard dataset	Learn the training dataset; classify
Twitter Suicide	Recognition	Sociology	Bridge big data and close reading and nuanced interpretations of manual qualitative coding and analysis.	Bridge the gap between close readings and nuanced interpretation of data and large-scale patterning.	Classification Convolutional neural network	Amplify expert decisions; classify	Aggregate labels to reach a consensus; identify classes of interest	Classify
MAIA-Machine Learning Assisted Image Annotation	Recognition	Oceanography	Speed up manual image annotation for environmental monitoring and exploration.	Efficiently identify good candidates for annotation.	Autoencoder networks (AEN) in combination with fully convolutional networks (FCN)	Review machine annotations; review machine training data	Boost training data; validate classifications	Help identify objects of interest; learn the training dataset
iNaturalist	Recognition	Environmental science	Map and share observations of biodiversity across the globe.	Maximize the utility of natural history collections	Classification	Annotate observations;	Provide a taxonomic nomenclature; curate data	Learn the training dataset; rank observations

				in a new era of data use.		follow authoritative taxonomies		
Observation.org	Recognition	Environmental science	Provide a free tool for all field observers around the world to record and share their plant and animal sightings.	Enhance the completeness and quality of the observation database.	Deep learning (neural network)	Annotate observations; classify	Develop a gold-standard dataset; validate classifications	Detect automatically
Snapshot Serengeti	Recognition	Environmental science	Provide a database of camera-trap-images.	Investigate multi-species dynamics in an ecosystem.	Plurality algorithm	Classify	Develop a citizen-labeled dataset; validate classifications	Produce a consensus dataset
Plantsnap	Recognition	Botany	Help people identify plants all over in the world.	Bridge the gap between nature and technology.	Customized classification algorithm	Annotate observations; peer review	Train machine classification	Detect automatically; classify
Milky Way	Recognition	Astronomy	Classify drawings on infrared images to identify targeted classes of astronomical objects.	Create a hybrid model of machine learning combined with crowdsourced training data.	Random Forest	Classify	Label data; curate data	Machine classify; finding missing items in the database

Human Protein Atlas	Combination of pattern recognition and classification of images	Life sciences	Map all the human proteins in cells, tissues, and organs using the integration of various omics technologies.	Improve large-scale image classification to reduce development costs for labs and show that game players can produce high-quality data.	Deep learning (neural network)	Classify	Label data; train citizen scientists	Classify
Galaxy Zoo AI	Recognition	Astronomy	Combine ML and big data to classify galaxies much faster than the most expert volunteers.	Combine human skills with AI speed to classify far more galaxies and do science.	Bayesian neural network	*Provide data for Galaxy Zoo*	Train machine classifications	Classify
Virus Spot	Prediction	Microbiology	Identify rotavirus infected cells in cryo-electron microscopy (cryo-EM) images.	Sort cryo-EM data in the future and streamline the data analysis pipeline.	Convolutional neural network - deep learning semantic segmentation	Classify; do collaborative analysis	Develop a citizen-labeled data set	Classify

Multiple Sclerosis	Prediction	Medicine	Predict the course of the disease based on randomly chosen clinical records of patients.	Attain prognostic ability above that of machine learning algorithms and groups of humans alone.	Random Forest	Make predictions; provide feedback and bugs reporting	Create hybrid predictions; train machine classifications	Make predictions

Table 1 Synopsis of project characteristics and tasks

We summarize the dataset according to the major categories and codes aggregated by the number of references (portions of coded text) across the 12 projects. In Figures 1, 2, and 3, we present the distribution of tasks performed by citizens, machines, and experts across projects. Given the small sample, the results do not have general representativeness. A description of the codes used to label citizen, expert, and machine tasks with examples from the data and detailed results from this work is in Appendix 3.

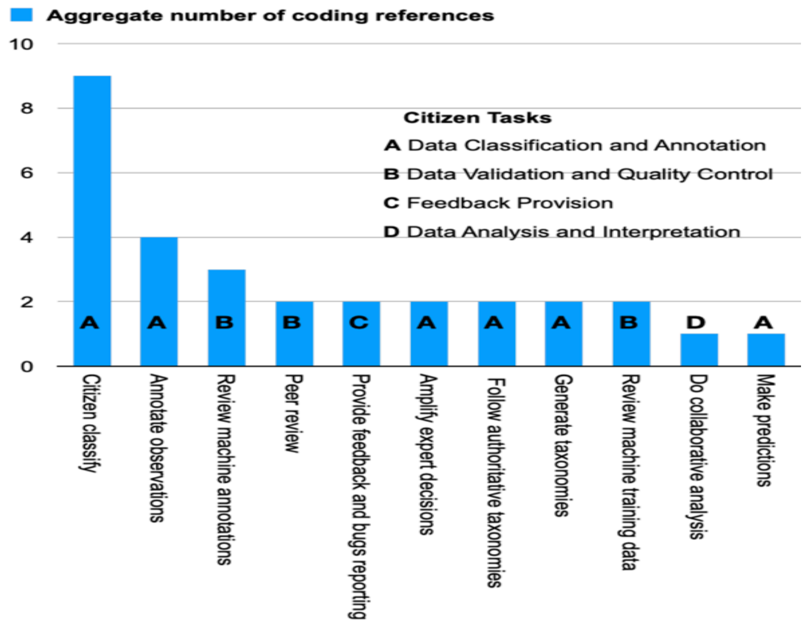

Fig 1 Citizen tasks, aggregated and sorted by number of references

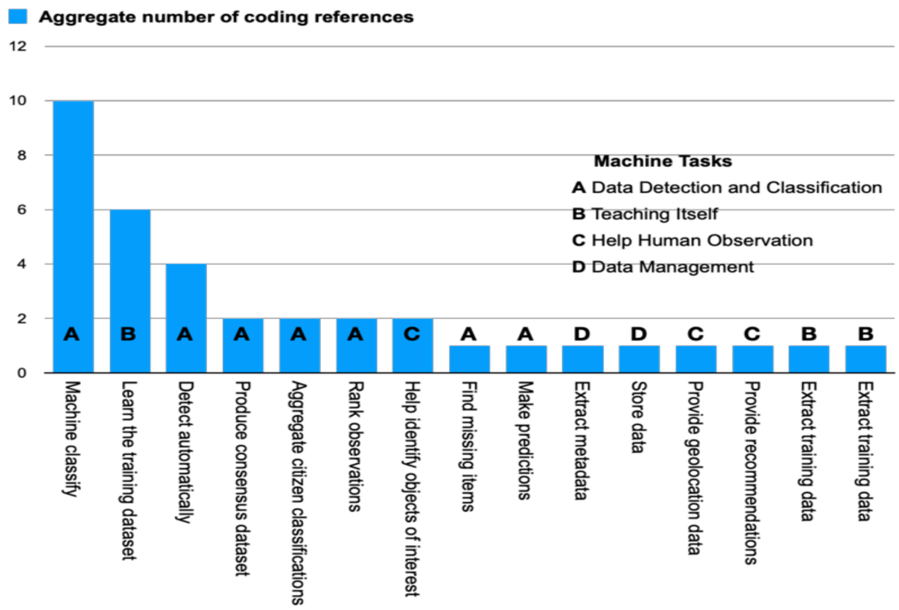

Fig 2 Machine tasks, aggregated and sorted by number of references

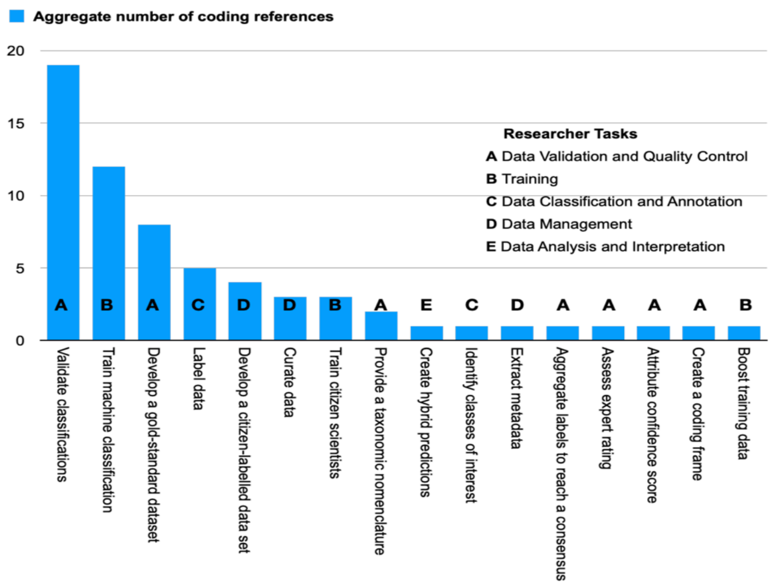

Fig 3 Expert tasks, aggregated and sorted by number of references

In Table 2, we present the three main tasks performed by each actor within the sampled projects. A complete description of tasks is in Appendix 3.

Task	**It describes…**	**Data examples**
Citizen		
Citizen classify	Citizen scientists classifying data. They can assign golden standard labels to objects, or assign their own labels.	*citizen participants identify patterns that machine cannot identify in bubble detection and contribute to a dataset called (Milky Way).*
Annotate observations	Citizen scientists adding context information associated with an observation, e.g., location and date/time of a photo.	*citizen participants provide evidence to their observation (e.g., time, place, descriptions) by using the tools provided by the database system (Observation.org).*
Review machine annotations	Instances in which citizen scientists have to approve, reject, or edit the automatically suggested matches.	*citizens eliminate the false positive detections, which mark regions of the images that show no OOI, the annotation candidates are manually reviewed in the last step (MAIA).*
Expert		
Validate classifications	Methods used to validate the quality of classifications. For example, it can include experts seeking to ensure that	*researchers evaluate MAIA on three marine image datasets and validate that MAIA is a promising and efficient method for marine image analysis (MAIA).*

	citizen scientist-trained machine model generates correct classifications.	
Train machine classifications	The use of labeled datasets for training machine learning to classify objects.	*researchers used the newly labelled DES data to do unsupervised recursive training to retrain their deep transfer learning model again, in order to boost the accuracy of classification (Galaxy Zoo AI).*
Develop a gold-standard dataset	A dataset that is accepted as the most accurate and reliable and needs to be used as a reference for checking and comparing machine labeling.	*the ratings of the top 4 expert raters (including the lead author) were used to create a gold-standard subset of the data. The gold-standard dataset set contains 100 images that were failed by experts, and 100 images that were passed by experts (Mindcontrol).*
Machine		
Machine classify	Machine classifying data by assigning labels to objects.	*the machine learning model was used for automated image classification of subcellular protein distribution patterns (e.g., recognizing patterns of the cell periphery such as plasma membrane, focal adhersions, cell junctions and making predictions of each cell) (Human Atlas).*
Learn the training dataset	Machines learning the "gold-standard" labeled dataset to predict labels when classifying data.	*in MAIA, the training is performed with the set of boosted training samples and the default configuration of the Mask R-CNN implementation. The training samples are used to train the machine supervised learning neural networks to produce a set of annotation candidates for a whole image dataset (MAIA).*
Detect objects automatically	An algorithm able to recognize objects, e.g., plants and animals.	*the system will provide a suggestion of what species it could be, and the suggestions come from the automatic image recognition model "Obsldentify" (Obsetvation.org).*

Table 2. The three main tasks performed by each actor

4.1 Examples of citizens-in-the-loop

Human-in-the-loop (HITL) is commonly described as the process in which humans help improve machine learning algorithms, e.g., by providing labels and features. HITL is said to leverage the benefits of human observation and classification skills, as well as machine computation abilities, to create better prediction models (Shih 2018). Our results indicate that experts are involved in every aspect of the loop, from annotating or labeling data to giving them to algorithms to train and make decisions from such predictions. Experts also test and validate models to improve accuracy by scoring their outputs, when algorithms are not able to make the right decisions. While experts are the humans mainly involved in the loop, citizens are involved as well at various stages of the process. We present three main examples of what we call *citizens-in-the-loop*, showing how citizens assist algorithms when they encounter difficulties. The examples show a type of interaction in which citizens and algorithms are interdependent and take turns to solve a task together, while the feedback loop allows a continuous improvement of the system (Berditchevskaia and Baeck 2020).

*When algorithms provide incorrect suggestions***.** In Observation.org, a free tool for field observers to record and share their plant and animal sightings, citizen scientists upload images of flora and fauna and if the recognition algorithm fails to provide correct identification of the species, then citizens can edit the wrong suggestion on the observation screen. Based on this, the system shows whether citizens have accepted or rejected observation data. Citizens contribute to creating a sort of gold-standard database used to train the machine learning model.
Also in Snapshot Serengeti machines can misclassify animals in the collected pictures. In such a case, citizen scientists in the role of spotters identify the animals and serve as teachers for the algorithms. Firstly, the algorithm classifies the picture. If the animal is detected with a certain probability, spotters come to the scene. AI offers a primary classification (animal recognition) to the spotter (also the trapper who uploaded records can pre-classify the image). A spotter validates/ invalidates the pre-classification and the image is not considered as validated until there is at least a 75% consensus (which can be adjusted in a certain project) among all the spotters involved. This is the input for the algorithms.

*When algorithms do not know to perform classification***.** In Milky Way, a system leveraging citizen science and machine learning to detect interstellar bubbles, citizens identify patterns that machines cannot identify in bubble detection and contribute to building a database. Researchers use the citizen identification output to train machine learning and build a model of an automatic classifier.

*When algorithms are active learners***.** In MAIA, a machine learning assisted image annotation for the analysis of marine environmental images, an algorithm poses queries to citizens, in the form of training data images. Citizens review these images and determine whether they contain objects of interest for classification or not. Then, they refine manually each image with a circle to mark the object of interest in the image, by modifying the circle position or size, so it closely fits the position and size of the object.

5. DISCUSSION

We highlight two aspects from the results. First, collecting, classifying, and validating data are some of the most common activities that humans carry out. In line with Kelling et al. (2012), the documented tasks appear to be sufficiently simple to motivate citizen scientists to participate, while machines cannot complete these tasks on their own. The results show that humans play a main role in classifying and annotating data to be used to train a model, but once a model is trained, it requires an expert-in-the-loop to interpret model predictions and potentially refine them to generate the most accurate results for unseen and unknown data. The sampled projects show the need to maintain human oversight over ML models, which can be explained as the need to repair mistakes made by machines, or because a combination of human and machin is most efficient (de Winter & Dodou 2014).

In most cases, citizen scientists still perform relatively simple tasks such as data collection, image coding, or object annotation. However, the roles of citizens can be reconfigured by algorithms as they come together with experts to constitute the project activities around data and objects (e.g., images). An example is the role of citizen scientists classifying images that serve as a training dataset for ML algorithms. Another role is that of citizens amplifying expert decisions. Instead of labeling thousands of training images, an expert can employ citizen scientists to help with this task, and machine learning can identify which citizen scientists provide expert-quality data.

Second, the sampled projects have used a relatively narrow set of ML applications aimed at helping humans become more efficient through new approaches to classification and prediction. This finding is consistent with the results of a Nesta's study of 20 project's bringing AI and collective intelligence together (Berditchevskaia and Baeck 2020). The authors argued that more attention should be paid to the relationship between AI and collective intelligence, in terms of the potential of AI to help groups to think and work together on identifying problems, finding solutions, and making decisions.
Furthermore, ML models are most useful when gold-standard or a "ground truth" datasets can be used as labels to properly train a classifier, because they represent the data that the model will encounter when it is used in the real world. Then, the ability of AI to identify patterns in huge amounts of data is useful for streamlining analysis.

For example, in Virus Spot researchers gather data from citizen participants and set a standard for how to use these data for an ML model. Then the algorithm learns how to sort and automate the segmentation of the cryo-EM data (images of viruses), and streamlines the data analysis from weeks to days.

Unlike automated systems that tend to remove or reduce the need for humans, the examined projects can be considered heteromated systems that do not function without the indispensable human mediation of engaged citizens (Ekbia and Nardi 2017). Ekbia and Nardi (2017) consider citizen science projects as examples of heteromation relying on what they called "inverse instrumentality" (p. 35), a concept they coined after examining complex technological systems like video games which insert humans strategically, to allow the systems to work in intended ways (Ekbia and Nardi 2012). Following their reasoning, citizen science projects are not heteromated in terms of experts' work, but through the participation of volunteers who act as indispensable mediators. As noted above, citizen science projects are not ordinary work settings employing paid staff. As Sauermann and Franzoni (2015) pointed out, citizen science projects provide important speed advantages as long as a large number of volunteers work in parallel, reducing the time required to perform a specific task. Insofar the success of these projects has relied on citizens willing to contribute time and effort. The sampled classification projects seek citizens to provide, classify, or annotate data. Their heteromated systems can push critical tasks to citizens, as the examples of citizens-in-the-loop show. These examples leverage the benefits of both the observation and classification skills of amateurs and non-professional scientists for data labeling, as well as the machine computation abilities, to create better prediction models (Shih, 2018). They also indicate the possibility of not just automating tasks but "leveraging" the knowledge of citizen scientists within a project, by capturing essential information and learning from the feedback they provide. In turn, this can result in motivating new citizen scientists to start contributing, and existing contributing citizens to generate a larger quantity of observations. In heteromated systems, efficiency is not opposed to engagement (Ekbia and Nardi 2017), as they cater to people's motivations to participate in projects that offer opportunities for meaningful tasks.

When allocating tasks, the argument about complementarity influenced by Fitt's (1951) list risks essentializing the attributes of humans and machines. We commonly think that humans are especially good at creativity, intuition, and abstraction, for example, while machines are good at speed, quantification, and efficiency. This complementarity is well described in the blog of Galaxy Zoo about using AI, wherein humans are acknowledged as being "much better classifier, able to make sense of the most difficult galaxies and even make discoveries," while AI is good at boring tasks (Galaxy Zoo Upgrade, 2019). As noted above, treating the attributes in Fitt's list as stable and natural does not take into account the dynamic redistribution of cognitive tasks or roles, involving interactions between humans and machines in a cooperative and situated way (e.g., Hutchins 1995). Cognitive work will be shifting between humans and machines (Ekbia and Nardi 2017), as the list of research tasks that machines can do is growing, although algorithms are still second to humans on recognizing patterns and they have longer learning curves.

6. CONCLUSION

This survey represents the perspective of the authors. Given the small sample of 12 applications, our analysis of task allocation needs to be considered only as an illustration. The approach we used suffers from the limitation of employing only documentary evidence, without combining it with other methods to reduce bias and compensate for the absence or the incompleteness of documents. Nevertheless, the employed method tried to minimize the subjectivity of the analysis, which is an initial step toward empirically examining function allocation. However, we hope it stimulates further research on this topic in citizen science projects using AI. For example, future studies could examine how the cooperative and situated practices in which humans and algorithms co-evolve and whether they change the content and meaning of tasks allocated early in the design of projects. Another topic could examine whether human-machine integration results in a skill-biased technological change, wherein machines take over low-skill tasks, for example. As obvious boundaries and distinctions between humans and machines are subject to blurring, we might face unexpected obstacles, possibilities, and questions worth investigating.

Studying whether a task can be performed by a machine or a human is important to incorporate not only useful but also meaningful human participation into a project. Deriving personal meaning and value from participating is important to citizen scientists who typically volunteer time and effort driven by intrinsic or social motivations and not for financial compensation (Sauermann and Franzoni 2015). For this reason, optimizing the overall experience of participants, and not just speed and efficiency, is critical to realize the potential for the contribution of each volunteer (Trouille et al. 2019). When we automatize certain tasks, we need to balance the goals of a project with the meaning that citizen scientists can derive from their participation. Citizen science projects need to cater to diverse needs and expectations. 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 difficult tasks

As the boundaries between humans and machines blur, human roles in citizen science projects are likely to be reconfigured by ML. We argue that the value of citizen science applications using ML should lie in creating pathways for citizens that support their engagement and agency while balancing the quality and accuracy of classifications. In this respect, it is valuable to think about machine-in-the-loop approaches that aim to help citizen scientists, besides the citizens-in-the-loop helping machines.

7. REFERENCES

Blickhanb S, Trouille L, Lintott CJ (2018) Transforming research (and public engagement) through citizen science. *International Astronomical Union. Proceedings of the International Astronomical Union* 14(A30): 18-523. (Section 4). https://doi.org/10.1017/S174392131900526X

Berditchevskaia A, Baeck P (2020) *AI and Collective Intelligence: Case Studies. 20 Projects Bringing Artificial Intelligence and Collective Intelligence together*. Accessed from https://www.nesta.org.uk/feature/ai-and-collective-intelligence-case-studies/ on October 14, 2020.

Berditchevskaia A, Baeck P (2019) *The Futures of Minds and Machines: How Artificial Intelligence Can Enhance Collective Intelligence.* Nesta Report. Accessed from https://media.nesta.org.uk/documents/FINAL_The_future_of_minds_and_machines.pdf on October 10, 2020.

Ceccaroni L, Bibby J, Roger, E, et al (2019) Opportunities and risks for citizen science in the age of artificial intelligence. *Citizen Science: Theory and Practice.* 4(1):29. DOI: 10.5334/cstp.241

Coughlin S, Bahaadini S, Rohani N, et al (2019) Classifying the unknown: Discovering novel gravitational-wave detector glitches using similarity learning. *Physical Review D* 99(8): 082002.

de Winter JCF, Dodou D (2014) Why the Fitts list has persisted throughout the history of function allocation. *Cognition, Technology and Work 16*: 1–11. DOI 10.1007/s10111-011-0188-1

Ekbia HR, Nardi BA (2017) *Heteromation, and other stories of computing and capitalism*. MIT Press, Cambridge, MA.

Ekbia HR, Nardi BA (2012) Inverse instrumentality: how technologies objectify patients and players. In Leonardi P, Nardi BA, Kallinikos J (eds.), *Materiality and Organizing: Social Interaction in a Technological World*. Oxford University Press, Oxford, pp. 157–176.

Eitzel MV, Cappadonna JL, Santos-Lang, C, et al (2017) Citizen science terminology matters: exploring key terms. *Citizen Science: Theory and Practice* 2(1): 1. DOI: https://doi.org/10.5334/ cstp.96

Fitts PM (1951) *Human Engineering for an Effective Air-Navigation and Traffic-Control System*. Division of National Research Council, Oxford, England.

Galaxy Zoo Upgrade (2019) Accessed from https://blog.galaxyzoo.org/2019/05/21/galaxy-zoo-upgrade-better-galaxies-better-science/ on 30 September 2020.

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

Hsieh H-F, Shannon SE (2005) Three approaches to qualitative content analysis. *Qualitative Health Research* 15(9): 1277-1288. 10.1177/1049732305276687

Hutchins E (1995) *Cognition in the wild.* MIT Press, Cambridge, MA.

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. *Biological Conservation* 181: 162-172. http://dx.doi.org/10.1016/j.biocon.2014.11.010

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

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

Leach B, Parkinson S, Lichten CA, et al (2020) *Emerging Developments in Citizen Science: Reflecting on Areas of Innovation*. RAND Corporation, Santa Monica, CA.

Fritz S, See L, Carlson T, et al. (2019). Citizen science and the United Nations Sustainable Development Goals. *Nat Sustain* 2*:* 922–930. https://doi.org/10.1038/s41893-019-0390-3

Lintott CJ, Reed J (2013) Human computation in citizen science. In Michelucci P (ed.), *Handbook of Human Computation*. Springer, New York, pp. 153-162.

Lukic V, Brüggen M, Banfield JK, et al (2018) Radio Galaxy Zoo: Compact and extended radio source classification with deep learning. *Monthly Notices of the Royal Astronomical Society* 476(1): 246-260. 10.1093/mnras/sty163

Miller G, Lintott CJ, Eisner N, et al (2020) *Can't we all just get along? Hybrid human-machine approaches to citizen science in the Zooniverse.* Presentation during the interactive session on Citizen science, AI and algorithms, 3rd European Citizen Science Conference, 10 September 2020, Trieste, Italy.

Mulgan G (2018) Artificial intelligence and collective intelligence: the emergence of a new field. *AI & Society*, 33:631–632. DOI: https://doi.org/10.1007/s00146-018-0861-5.

Nguyen T, Pankratius V, Eckman L, at al (2018) Computer-aided discovery of debris disk candidates: A case study using the Wide-Field Infrared Survey Explorer (WISE) catalog. *Astronomy and computing* 23: 72-82. 10.1016/j.ascom.2018.02.004

Ostermann FO, Garcia-Chapeton GA, Kraak M-J, et al (2017) Mining user-generated geographic content: an interactive, crowdsourced approach to validation and supervision. In Bregt A, Sarjakoski, T, van Lammeren R, Rip F (eds), *Societal Geo-Innovation: Short Papers, Posters and Poster Abstracts of the 20th AGILE Conference on Geographic Information Science*. Wageningen University Research, Wageningen, The Netherlands. Accessed from https://agile-online.org/images/conference_2017/Proceedings2017/shortpapers/74_ShortPaper_in_PDF.pdf.

QSR International (2020) *NVivo Qualitative Data Analysis Software* [Software].

Resnik DB, Elliott KC, Miller AK (2015). A framework for addressing ethical issues in citizen science. *Environ Sci Pol*. 2015;54:475–81.

Robinson OJ, Ruiz-Gutierrez V, Fink D (2018) Correcting for bias in distribution modelling for rare species using citizen science data. *Diversity and Distributions* 24(4): 460-472.

Rognin L, Salembier P, Zouinar M (2000) Cooperation, reliability of socio-technical systems and allocation of functions. *International Journal of Human-Computer Studies* 52(2): 357-379. https://doi.org/10.1006/ijhc.1999.0293

Sauermann H, Franzoni C (2015) Crowd science user contribution patterns and their implications. *PNAS* 112(3): 679-684. https://doi.org/10.1073/pnas.1408907112.

Sheridan TB (2000) Function allocation: Algorithms, alchemy or apostasy? *International Journal of Human-Computer Studies* 52(2): 203-216. DOI: 10.1006/ijhc.1999.0285

Shih PC (2018) Beyond human-in-the-loop: Empowering end users with transparent machine learning. In Zhou J, Chen F (eds) *Human and Machine Learning. Visible, Explainable, Trustworthy and Transparent*. Springer, Cham, pp. 37-54. https://doi.org/10.1007/978-3-319-90403-0_3

Shirk JL, Ballard HL, Wilderman CC, et al (2012). Public participation in scientific research: a framework for deliberate design. *Ecology and Society* 17(2): 29. http://dx.doi.org/10.5751/ES-04705-170229

Sullivan DP, Winsnes CF, Åkesson L, et al (2018) Deep learning is combined with massive-scale citizen science to improve large-scale image classification. *Nature biotechnology* 36(9): 820–828. https://doi.org/10.1038/nbt.4225

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

Wiggins A, Crowston K (2012) Goals and tasks: Two typologies of citizen science projects. *45th Hawaii International Conference on System Sciences.* IEEE. 10.1109/HICSS.2012.295.

Willi M, Pitman RT, Cardoso AW, et al (2019) Identifying animal species in camera trap images using deep learning and citizen science. *Methods in Ecology and Evolution*, 10(1): 80-91. https://doi.org/10.1111/2041-210X.13099

Wright P, Dearden A, Fields B (2000) Function allocation: a perspective from studies of work. *International Journal of Human-Computer Studies* 52(2): 335-355. https://doi.org/10.1006/ijhc.1999.0292

Wright DE, Fortson L, Lintott, CJ, et al (2019) Help me to help you: machine augmented citizen science. *ACM Transactions on Social Computing* 2(3): 1-20. https://doi.org/10.1145/3362741

Zevin M, Coughlin S, Bahaadini S, et al (2017) Gravity Spy: integrating advanced LIGO detector characterization, machine learning, and citizen science. *Classical and quantum gravity* 34(No 6): 064003. https://doi.org/10.1088/1361-6382/aa5cea