



# Supporting Human and Machine Co-Learning in Citizen Science: Lessons From Gravity Spy

CARSTEN ØSTERLUND

KEVIN CROWSTON

COREY B. JACKSON

YUNAN WU

ALEXANDER O. SMITH

AGGELOS K. KATSAGGELOS

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

COLLECTION:  
THE FUTURE  
OF ARTIFICIAL  
INTELLIGENCE AND  
CITIZEN SCIENCE

RESEARCH PAPER

ubiquity press

## ABSTRACT

We explore the bi-directional relationship between human and machine learning in citizen science. Theoretically, the study draws on the zone of proximal development (ZPD) concept, which allows us to describe AI augmentation of human learning, human augmentation of machine learning, and how tasks can be designed to facilitate co-learning. The study takes a design-science approach to explore the design, deployment, and evaluations of the Gravity Spy citizen science project. The findings highlight the challenges and opportunities of co-learning, where both humans and machines contribute to each other's learning and capabilities. The study takes its point of departure in the literature on co-learning and develops a framework for designing projects where humans and machines mutually enhance each other's learning. The research contributes to the existing literature by developing a dynamic approach to human-AI augmentation, by emphasizing that the ZPD supports ongoing learning for volunteers and keeps machine learning aligned with evolving data. The approach offers potential benefits for project scalability, participant engagement, and automation considerations while acknowledging the importance of tutorials, community access, and expert involvement in supporting learning.

## CORRESPONDING AUTHOR: Carsten Østerlund

Syracuse University, School of  
Information Studies, US

[costerlu@syr.edu](mailto:costerlu@syr.edu)

## KEYWORDS:

learning; zone of proximal  
development; human-artificial  
intelligence augmentation

## TO CITE THIS ARTICLE:

Østerlund, C., Crowston, K,  
Jackson, CB, Wu, Y, Smith, AO  
and Katsaggelos, AK. 2024.  
Supporting Human and Machine  
Co-Learning in Citizen Science:  
Lessons From Gravity Spy. *Citizen  
Science: Theory and Practice*,  
9(1): 42, pp. 1–14. DOI: [https://  
doi.org/10.5334/cstp.738](https://doi.org/10.5334/cstp.738)

## INTRODUCTION

The growing capability of artificial intelligence (AI) technologies has sparked considerable rethinking of the interactions between humans and machines. The traditional narrative has emphasized unidirectional flows of knowledge, machines enhancing human capabilities through automation or decision support or humans providing labeled data to train machines. Few systems envision a bi-directional relationship in which machines not only augment and extend human capabilities, but humans work to augment and extend the capabilities of machines. Such a symbiotic relationship requires that both humans and machines are engaged in co-learning, a continuous learning process in which the two support one another.

The possibility for a mutually beneficial system is particularly compelling for citizen science (CS). An increasing number of CS projects are deploying AI technologies (Ceccaroni et al. 2019), for example, iNaturalist, eBird, Snapshot Safari, and the Koster Seafloor Observatory in biology; Muon Hunters, Galaxy Zoo, and Gravity Spy in astronomy; and Etch-a-Cell, Phylo, and Eyewire in medicine. However, these applications tend to focus on ways that AI extends the capabilities of volunteers or science teams. We see opportunities in CS for continuous refinement and expansion of machine learning (ML) by volunteers in the same setting where machines offer humans ways to amplify their learning. These efforts raise several new and old issues in the sciences, especially about the division of labour between humans and machines (Kullenberg and Kasperowski 2016), and how that relationship might affect scientific standards, inclusion, exploitation, and learning (Danielsen et al. 2005; Riesch and Potter 2014).

The challenge is how to facilitate human and machine learning in a way that the two do not simply counter one another. For instance, ML systems simply automating volunteers' tasks may remove opportunities for productive learning among volunteers. However, if ML gradually takes over low-level tasks, it might allow volunteers to focus on and learn more intricate tasks as the technology efficiently manages routine responsibilities (Norouzzadeh et al. 2018; Willi et al. 2019). While this scenario might facilitate human learning in a project, attention must also be paid to ML. For instance, many algorithms struggle with novel categories in the data that humans can detect. From a project design perspective, strategies are needed to facilitate and build synergies between human and machine learning.

The literature on co-learning (Huang et al. 2019; Kent et al. 2021; Schoonderwoerd et al. 2022; van den Bosch et al. 2019; Wilder et al. 2020) and augmentation (Brynjolfsson and McAfee 2014; Daugherty and Wilson 2018; Davenport and Kirby 2016), offer complementary approaches for facilitating human and machine learning. First, the emerging

literature on co-learning takes as its point of departure that neither humans nor AI are static (van den Bosch et al. 2019; van Zoelen et al. 2021); rather, they constantly develop and adapt, and in the process, influence one another. In other words, it is not desirable to focus on only the human or the AI side of a human-AI collaboration because both are adaptable. Second, the literature highlights that AI and humans have different strengths and shortcomings. Current AI tends to do well when it comes to handling narrow and well-defined tasks but falls short when a task requires abstract thinking, common sense, and the transfer of knowledge between areas. Rich tasks with an unknown variety of conditions tend to lead to low AI performance (van den Bosch et al. 2019). In contrast, humans, unlike AI, struggle to store and process large amounts of information but excel when it comes to solving abstract problems based on common sense, scarce information, and knowledge transfer (Blake and Butcher-Green 2009).

Several co-learning studies have taken inspiration from human team literature (Kent et al. 2021; van Zoelen et al. 2021; Wilder et al. 2020) in framing their approach to the co-evolution and the divergent strengths of human-AI collaborators. However, we have concerns that framing human-AI interactions as teaming raises unrealistic expectations, for example, about the possible level of mutual understanding or shared mental models. Instead we consider the interaction from the perspective of augmentation (Brynjolfsson and McAfee 2014; Daugherty and Wilson 2018; Davenport and Kirby 2016), suggesting that AI and humans can draw on their divergent strengths and mitigate the other's weaknesses. Here, augmentation can facilitate the collaborative accomplishment of a task (Agrawal et al. 2023; Raisch and Krakowski 2021), but it can also help us understand how AI and humans can support each other's learning and growth over time. Approaching augmentation from the point of view of co-learning adds a dynamic perspective to the augmentation and allows us to consider co-learning beyond the dyad as a unit of analysis. For instance, AI might augment the learning of an entire team of human and machine learners in some situations, but in others, the learning of individuals.

This leads us to the following broad research question: How do we secure human and machine co-learning in CS projects by allowing humans to augment machine learning and AI to augment human learning?

## THEORY

Learning theories focus on the development of internalized knowledge or skills that create a lasting change in behaviour (Engeström 2001). The question we address is how to structure activities to support such learning. We draw on

the notion of a zone of proximal development (ZPD) to shift the focus from mental and cognitive processes of learning to observable behaviours of people and AI working to achieve some objective. In this perspective, skills are not assessed as abstract demands; instead, what matters is the exhibition of skills and demands through the process of achieving an objective (Kaptelinin and Nardi 2009, p. 31). For instance, Kaptelinin and Nardi (2009) note that saying that someone is “good at math” can be misleading because performance can vary significantly depending on how the problem is posed (2009, p. 31).

Vygotsky defined the ZPD as:

“The distance between the actual level of development as determined by independent problem solving and the level of potential development as determined through problem-solving under adult guidance or in collaboration with more capable peers.” (Vygotsky and Cole 1978, p. 86).

Figure 1 expresses this relationship, distinguishing between three categories of tasks: (A) tasks that can be done without assistance, (B) tasks that can be done only with assistance (the ZPD), and (C) tasks that cannot be done even with assistance. The theory posits that people learn best when they work on tasks in their ZPD. Repeating tasks they can already do (category A) will not expand their capabilities, nor will attempting and failing at tasks that are impossible for them (category C). However, accomplishing tasks with assistance increases their capacity as they gradually learn to do them on their own, shifting them to category A. Further, these new skills may be the foundation for attempting tasks that would earlier have been impossible, thus moving some tasks from category C to the ZPD.

The original ZPD concept posited that other, more knowledgeable people would provide assistance on tasks in the ZPD. Such assistance may be possible in some CS projects. For instance, eBird’s novice bird watchers may go birding with a more experienced birder, learning from them about new species or observational techniques. However, other projects do not afford human support. For instance, Zooniverse projects generally do not let learners directly

see how others work or ask for advice while doing a task. Yet, even in such projects, there are non-human sources of assistance: tutorial materials; feedback; or what Mugar et al. (2014) termed practice proxies, community discussions of a task viewable after it is completed that hint at how to perform it.

As an example, many CS projects rely on humans’ ability to recognize patterns. Some of the patterns may be readily apparent even to newcomers to a project. Galaxy Zoo asked about simple patterns and most Snapshot Serengeti volunteers likely came to the project able to distinguish lions and elephants. However, distinguishing similar species of antelopes or gazelles may require frequent reference to the training materials, and even then, uncertainties may remain. This situation indicates a newcomer’s ZPD: Identifying antelopes is possible but requires assistance. Moving to the far right of Figure 1, there will also be tasks that volunteers cannot do even with assistance, for example, identifying animals photographed from a distance. Yet, even antelopes at a distance may become interpretable with practice and assistance.

So far, we have been discussing human activities and human ZPD. Where does this leave us regarding machines’ learning capabilities and our search for synergies between human and machine learning? If learning is defined as expanding the range of tasks mastered, it applies whether the learner is a human or a system. By considering when machines may learn and humans can take on the assistant role, we can approach the ZPD from a machine-learning perspective. Figure 2 depicts a situation in which a machine can do some tasks without assistance (i.e., automation), other tasks only with assistance (e.g., when volunteers confirm or correct an ML classification), and many tasks the machine cannot do even with assistance. In short, we envision a ZPD supporting ML parallel to a ZPD supporting human learning.

These two ZPDs, one serving humans and the other machines, do not have to work independently. If we accept that humans and machines can both learn and assist each other, then synergies between the two should be possible. Figure 3 summarizes this perspective, where human- and machine-driven learning help move each other’s ZPD

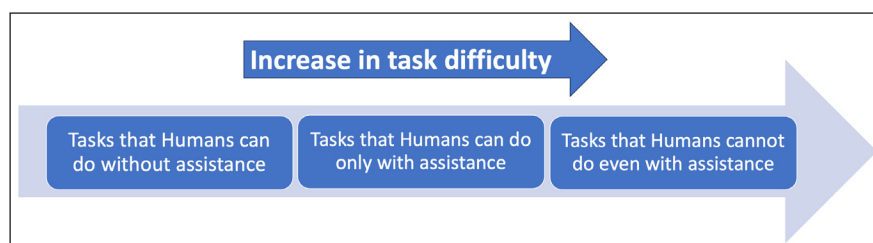
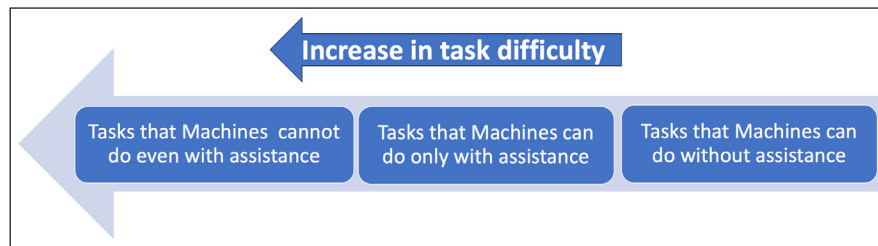
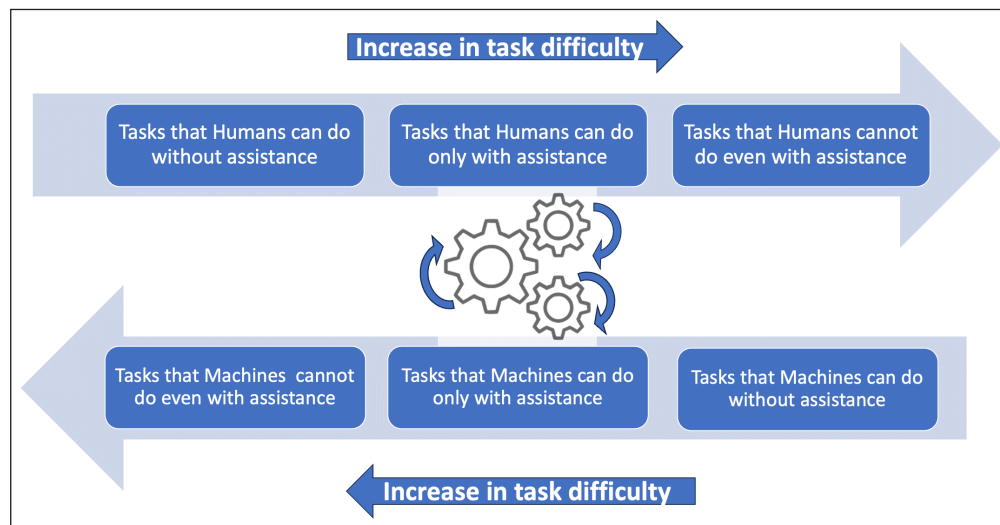


Figure 1 Individuals’ increase in capability, with the zone of proximal development (ZPD) in the centre of the figure.



**Figure 2** Machine zone of proximal development (ZPD).



**Figure 3** Co-augmentation of human and machine zones of proximal development (ZPD).

forward. The human and the machine ZPD augment each other by assisting the other in their ZPD, that is, helping them to do tasks they cannot do alone.

Given this theory, we can restate our research question as follows:

1. How can we use AI in citizen science projects to facilitate volunteer learning by keeping them in their ZPD?
2. How can human volunteers facilitate machine learning by keeping the machine in its ZPD?
3. How can we humans use AI to address tasks that neither can do independently?

## METHODS

To answer the research questions, we employ a design-science approach, which combines traditional research methods with system development to tackle research questions through design (Peffer et al. 2007). Design science is practiced in many domains, notably in human-computer interaction (HCI) and computer science, though rarely with this explicit label (Gregor and Jones 2007).

Its formal conceptualization originates from the field of information systems, aiming to overcome the perception of system design as atheoretical. This reconceptualization is valuable even in fields regularly embracing system design, as it shifts the focus from simply designing artifacts to using design to address larger research questions. In HCI, for instance, system evaluations (Brynjolfsson and McAfee 2014; Daugherty and Wilson 2018; Davenport and Kirby 2016; Raisch and Krakowski 2021) are often criticized for being usability-focused rather than scientifically oriented.

Design science research starts with problem identification and motivation, followed by definition of objectives for a solution. Through a process of design and development, it yields two outcomes: a functional system that addresses a specific practical design problem. The functionality of the system can be demonstrated and evaluated. In addition to the system, the project results in communication of meaningful scholarly contributions (Peffer et al. 2007). In contrast to typical social science research, design science involves additional components such as interactions with subject-matter experts, situational focus, and system building (Gregor and Jones 2007). Unlike typical systems research, design science emphasizes the explicit use of theory for design decisions, meaning that theory, design,

and evaluation are interrelated and iterative processes (Hevner et al. 2008). The term theory in design science research encompasses adopting existing theories, consulting experts, and reviewing project-specific design literature (Gregor and Jones 2007). New theories may come before, during, or after system development. The core of design science research is designing and developing a system informed by theoretical and practical concerns (Gregor and Jones 2007; Hevner et al. 2008; Peffers et al. 2007). The artifact's design may contribute to new theories, necessitating consideration for future evaluations. Finally, the evaluation stage in design science research goes beyond confirming success or failure. The design scientist must align artifact design with the research questions to be addressed through evaluation. It validates adopted theories or generates new ones, addressing broader research questions. Evaluation is thus not merely an endpoint; it often guides improvements for the next design iteration of the artifact, whether as a problem-solving system or a research tool.

We engaged in design science by designing, deploying, and evaluating a CS project called Gravity Spy (Zevin et al. 2024, 2017), hosted on the Zooniverse platform (Simpson et al. 2014). The motivation for and objectives of the project will be introduced in the following section. The theory for this project is ZPD, as described in the previous section. In the findings section, we describe how the concept of ZPD was incorporated into the project's design and how it influences the ongoing design of a follow-up project. We also present our evaluations of the project and reflect on the implications of the evaluations for the theory. We draw on many different sources of data collected throughout the project: interviews with the Laser Interferometer Gravitational-Wave Observatory (LIGO) and ML scientists (domain experts), interviews with volunteers, trace data documenting system use, participant observation, and our use of the system. Other publications provide more details about these data collection and analysis efforts (e.g., Crowston et al. 2023; Jackson et al. 2020a,b).

## SETTING

Gravity Spy supports LIGO (Aasi et al. 2015), which detects gravitational waves created by cosmic events such as black-hole mergers. Because of the extraordinary sensitivity of the detectors, they record orders of magnitude more noise events (called glitches) than genuine detections. Glitches can obstruct or confuse astronomical detections, so LIGO scientists seek to find and fix their causes to improve detector performance. The motivation for our project was to create a dataset to support the analysis of glitches. The specific objective was to increase the number of glitches

analyzed by enabling volunteers to contribute by helping them to learn about the glitches and the LIGO detector.

The task assigned to volunteers and the ML in Gravity Spy is to identify the classes of glitches. LIGO scientists have identified many classes of glitches with distinct appearances and presumed common known or unknown causes. For instance, a scattered light glitch is caused by laser light scattering off some component in the detector, creating a characteristic arch shaped glitch. Having a collection of glitches of the same class enables searches for their causes, for example, by looking for coincident activity in the detector.

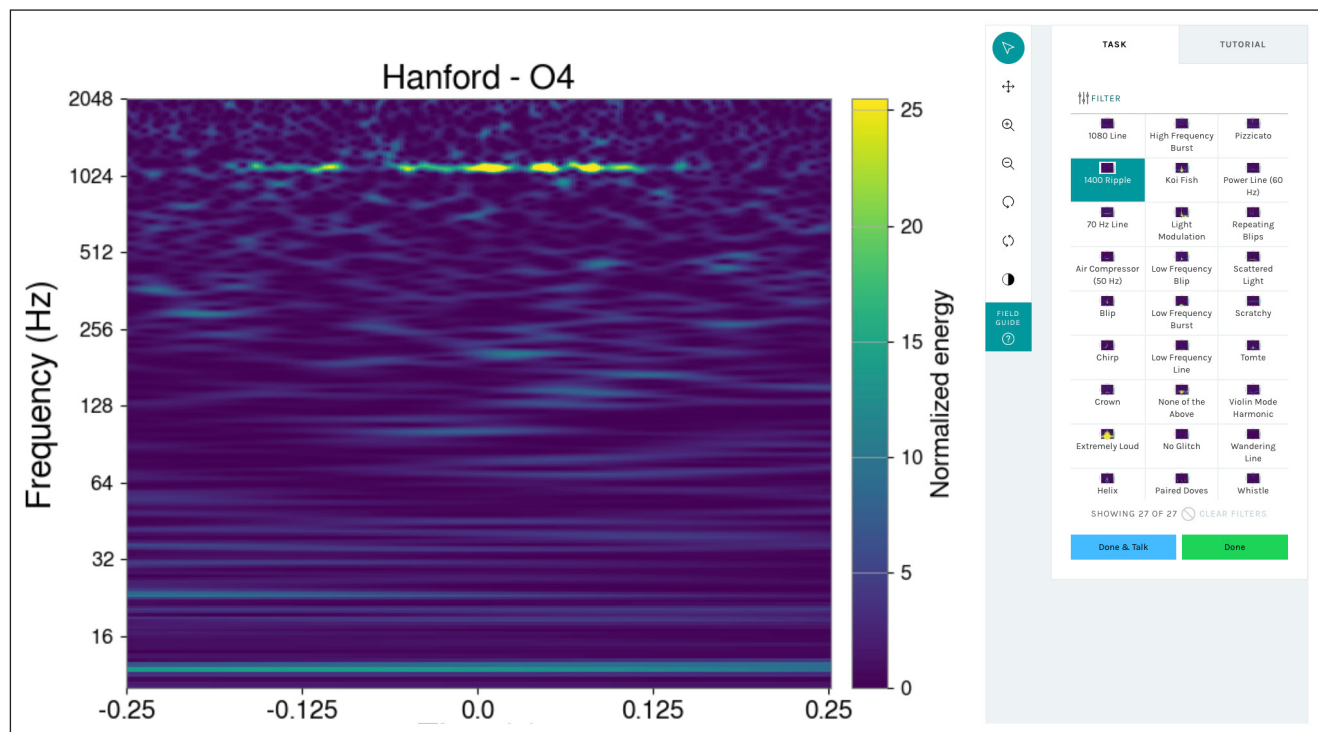
Glitches are presented to volunteers for classification as spectrograms, specifically, Omega scans (Chatterji et al. 2004), a visual representation with time on the horizontal axis, frequency on the vertical axis, and intensity of the signal represented by the colour from blue to yellow. The classification interface is shown in Figure 4, with a spectrogram on the left (of a so-called Tomte glitch in the figure) and the possible classes on the right. The classification interface was created using the Zooniverse project builder, in which a project can be created with a few mouse clicks and uploads of data and other materials (e.g., sample images).

Most glitches are classified by volunteers into one of the 26 known classes (or “None of the above”), creating a dataset of identified glitches to support exploration for their root causes. More advanced volunteers handle glitches that do not fit a known class by compiling sets of glitches of similar appearance that may be instances of a new class.

Gravity Spy also incorporates an ML algorithm. The specific task for the ML at this stage is to produce a score for the likelihood that a glitch belongs to each known class. The model is a convolutional neural network (CNN), a class of deep learning algorithms with exceptional performance for image recognition (Krizhevsky et al. 2012). (A discussion of the architecture for the CNN originally used can be found in Bahaadini et al. 2018a; Soni et al. 2021; Zevin et al. 2017). The current version of the model uses a new architecture with several advances, for example, fusion across multiple images and label smoothing (Wu et al. 2024).

## FINDINGS FROM GRAVITY SPY

In Gravity Spy, human learners, in the form of volunteers, and the ML model serve as subjects striving to classify LIGO glitches. Learning to identify existing and new glitches is crucial to that process. Humans and machines each have distinctive strengths. In the context of Gravity Spy, machines excel at identifying patterns in large



**Figure 4** The Gravity Spy classification interface, showing a glitch to be classified on the left and the potential glitch classes on the right.

datasets. On the other hand, humans can discern and rationalize the underlying causes of those relationships, a task that necessitates complex reasoning, contextual knowledge, and intuition. Another example can be found in the visual separation of glitches to learn their distinctive morphological characteristics. ML algorithms can easily separate glitches through sophisticated pattern recognition, requiring enormous computational power. Such classifications would be a time-consuming and highly subjective task for volunteers. In that sense, humans and ML act as mediators for each other's activities. We will discuss each, starting with the volunteers' human learning mediated by the ML model.

## MACHINE CLASSIFICATION SUPPORTING HUMAN LEARNING

In many image-classification CS projects, newcomers face the daunting task of learning to distinguish among many options. For instance, in the popular Snapshot Serengeti project, volunteers must select from 56 possible species, many unfamiliar and distinguished only by subtle features. In contrast, in Gravity Spy, participants progress from learning a few obvious glitch types to classifying many glitch types with less obvious features, a design approach informed by ZPD as they advance through increasingly challenging workflows. The ML guides the human learner by determining the workflow to which a particular glitch is assigned. Glitches that the ML confidently classifies are

assigned to beginner workflows. High ML confidence for a glitch means they are likely (though not certain) to be examples of a class. Multiple beginner workflows contain an increasing number of glitch classes, as shown in the table in Supplemental File 1: Table 1.

Specifically, in the current version of workflow 1 (WF1), volunteers are shown glitches that the ML has classified with high confidence as belonging to just two common and easily distinguished glitch classes: Blip and Whistle. The classification interface offers just those two options, plus "None of the above" to capture instances where the ML is mistaken, meaning that even beginners can make scientifically useful contributions to the project by correcting ML errors. Note that the volunteers are never shown the ML classification, but they have access to tutorial materials describing the task and the classes of glitches. When the volunteers have mastered these glitches, as assessed by their correctness in classifying gold-standard data, that is, data classified by LIGO scientists, they are promoted to the next level. Volunteers are given feedback after classifying gold-standard data, another kind of assistance. In WF2, volunteers are shown additional glitches with high ML confidence, that is, Koi Fish, Power Line, and Violin Mode classes. As the volunteers move to increasing workflow levels, new classes are added as options until they see all of the glitches.

We conducted an online A/B field experiment on Gravity Spy to evaluate the scaffolded learning regimen

(Jackson et al. 2020a). The experiment involved 368 volunteers randomly assigned to a control or treatment condition to assess their accuracy on gold-standard data provided by the science team. The control group started in WF5, where they were shown glitches of all classes with a wide range of ML confidence without training, as this is the default approach adopted by similar projects. In contrast, the treatment group received the scaffolded training described above (i.e., they started in WF1). As anticipated, the average accuracy for volunteers in the treatment group (95%) was statistically significantly higher on a Mann-Whitney-Wilcoxon test compared to the control group (60%). They also contributed significantly more classifications on average to the project (228 versus 121 classifications) and returned to the project more frequently (2.5 versus 2 sessions).

In summary, the ML facilitates the human learner as their ZPD evolves. Accordingly, we can state a first design principle: CS project developers seeking to support volunteers' ongoing learning should identify and assign tasks likely to be in a volunteer's ZPD because such tasks will expand the volunteer's capabilities without overwhelming them. AI can be useful to evaluate a volunteer's status and to select appropriate tasks.

## HUMAN CLASSIFICATION SUPPORTING MACHINE LEARNING

The ML is also a learner. Its ZPD moves by increasing its accuracy in identifying known glitch classes and expanding the range of classes known. In both cases, the volunteers assist the ML. The model was initially trained on approximately 7,700 glitches classified by LIGO scientists into 19 initial classes. In the project's initial phase, it was not accurate enough, so input from volunteers was crucial to increase confidence in the classifications. The training set has now been supplemented with glitches classified by volunteers to include nearly 10,000 labelled glitches over 23 classes (Zevin et al. 2024), and it continues to expand. The increase in the training set along with model improvements has greatly improved the model's accuracy. Indeed, the ML is now sufficiently accurate that we are reconsidering whether volunteers need to be involved in classifying all glitches (i.e., the machine seems capable of some tasks without assistance). LIGO scientists also regularly use the Gravity Spy ML model when dealing with glitches rather than CS results. ML has advantages, for example, it can be applied to glitches on demand, whereas CS results take time to accumulate. More subtly, LIGO scientists prefer the apparent objectivity of the ML model while expressing doubts about the reliability of volunteer classifications, that is, they do not perceive the machine as operating in its ZPD.

Second, a significant limitation of the ML classifier is its inability to cope with novelty, being able to identify only the classes on which it was trained. This limitation is a particular concern for LIGO, as the detectors continually evolve: Some glitches are fixed, but new ones emerge. For instance, in the most recent LIGO detector run, volunteers noticed that the ML had started to misclassify a new class of glitch as Whistles, one of the known classes (Wu et al. 2024). After a closer examination, the science team realized that Whistle glitches seemed to have disappeared after the detectors were updated. However, new glitches had emerged that the ML was not trained on and which it misclassified. Meanwhile, the human volunteers had little trouble distinguishing the novel glitches and brought them to the science team's attention relatively quickly.

To cope with novel classes of glitches in lower-level workflows, volunteers have the option of "None of the above" to correct the ML. In higher workflows, the volunteer task shifts from classifying to searching for novel glitch classes to retrain the ML. Higher workflows include glitches with lower ML scores, and the highest-level workflow contains only glitches that the ML had trouble classifying or that were identified as "None of the above" in a lower workflow. Volunteers in these workflows develop collections of glitches with similar novel appearances that are possible instances of novel glitch classes.

Following the identification and curation of a potential new glitch class, volunteers can nominate the class for addition to the system, which expands ML's capability. They do so by creating a New Glitch Proposal, including a name, description, exemplar, and their collections of similar images. LIGO scientists evaluate the proposals for robustness and usefulness of the proposed glitch class for debugging the detector. If accepted, the new class is included in the list of glitch classes on which the ML is trained (using the provided examples initially) and made available to volunteers in the classification interface. Accordingly, we can state a second design principle: CS developers seeking to expand the capabilities of AI systems should incorporate feedback from volunteers on tasks that the volunteers can do but the AI cannot as yet do as those tasks will expand the AI's capabilities. Unfortunately, proposal review and system updates are time-consuming, manual processes. As such, neither is done regularly, meaning that the ML's capacity expands only rarely (and indeed, currently lags the Zooniverse site in the number of glitch classes included). We are currently working on automating the retraining process to lower this barrier.

Third, the inability of ML to deal with novelty could be addressed technically by employing unsupervised learning techniques to cluster glitches to identify classes beyond those already known. We have explored such techniques,

for instance, using the ML model to extract properties of glitches in a high-dimensional feature space and then clustering in that space to identify morphologically similar images. However, the resulting clusters still require inspection by a human for coherence and vetting by LIGO scientists before they can be considered for addition to the ML training set and the Zooniverse system. In other words, even in this mode, the machine learning needs human supervision to learn.

## CO-LEARNING

As discussed, human and machine ZPDs do not exist independently in Gravity Spy. The project design strives to build synergies between human and machine learning where the human activities augment the machine's ZPD while the machine activities augment the human ZPD (see [Figure 3](#)). We offer the development of new glitch classes as an example of such co-learning.

To assist the ML in dealing with new glitches, advanced volunteers focus on finding new glitch classes, noted above as a key volunteer activity. One of the challenges in this work is collecting a large enough sample of glitches to justify the need for a new class and on which to retrain the ML model. To augment this activity, we built Similarity Search, a tool using the unsupervised clustering approach described above to locate glitches similar to a given glitch. Details of the clustering algorithm and search approach can be found in ([Bahadini et al. 2018a,b](#); [Coughlin et al. 2019](#)). Users can evaluate the metadata of retrieved glitches, decide which images to include or exclude, and export the search results to a new Zooniverse collection. As we do not have ground truth for which glitches are related, our evaluation is based on volunteer feedback. The tool is felt to be effective in filtering out non-matching glitches, enhancing the purity of the set the volunteer will examine, thus saving time and effort. In short, we see co-learning, in which human learners assist the machine in learning new glitch classes while the machine assists the humans by easing the burden of sifting through a large dataset in search of those new classes. Accordingly, we can state a final design principle: CS project developers should seek opportunities for AI systems to support volunteers in performing tasks that neither can do alone, by decomposing the tasks volunteers perform and identifying steps that could benefit from support, for example, by processing larger volumes of data. Such analysis might also reveal the potential to expand the scope of the volunteers' work.

In summary, the Gravity Spy project illustrates how different learners can support each other: ML supports human volunteers learning to classify by keeping them in their ZPD, and the products of the human classification support improvements to the ML to enable it to be more accurate

and to do more. And further, the two can work together on tasks that neither can do entirely independently.

## DESIGNING FOR LEARNING AND CO-LEARNING IN CITIZEN SCIENCE

In this section, we describe how we are implementing the theoretical ideas about ZPD and design principles presented above to develop the next generation of the Gravity Spy project, Gravity Spy 2.0. The problem we seek to address is improving how volunteers describe new glitch classes to make the proposals more useful for LIGO scientists. In the current system, volunteers have limited knowledge about the underlying mechanisms within the detectors that generate glitches; nor do they have any way to explore those mechanisms. As a result, new glitch class identification is done phenomenologically, that is, by grouping glitches with similar appearance: Witness the fact that volunteer-identified glitch classes are named by shape, for example, Helix or Crown, in contrast to most LIGO-identified classes that are named by cause, for example, Whistle or Scattered Light. (The exceptions are the handful of glitches whose causes are unknown, for example, Tomte, Blip, and Koi Fish.)

This approach has been effective in identifying new glitch classes, but the essential next step of identifying the causes of glitches requires the attention of the overloaded LIGO science team. Being able to explore the underlying dynamics in the detectors causing glitches would make volunteers' searches for new glitch classes more helpful to the science team. Regarding volunteer involvement, the Gravity Spy 2.0 task is not just about handling the sheer volume of work, but also involves leveraging the abilities of motivated volunteers to mimic the scientific investigator practices of professional scientists, developing novel insights and creative ways of examining complex datasets. Volunteers may explore new data relations or make connections that the professionals might not (moving from correlation to causation), potentially leading to novel insights about the underlying causes of these glitches. Nevertheless, the science team will still need to be involved, for example, in identifying which glitch classes are worthy of investigation, something that the volunteers are not well placed to do ([Crowston et al. 2018](#)).

We first describe how the LIGO scientists address the task. To explore the cause of glitches (i.e., what is happening in the detector or the environment that causes particular glitches), LIGO scientists carry out studies using auxiliary-channel (AC) data. Along with gravitational waves, the LIGO detectors record more than 200,000 channels of data per detector from a diverse set of sensors that

continuously measure every aspect of the detectors and their environment, for example, equipment functioning, activation of components, seismic activity, and weather (Nguyen et al. 2021; Nuttall 2018). This dataset holds clues to the cause of glitches. Since different classes of glitches are created by different mechanisms, they are expected to be correlated with diverse ACs (Davis and Walker 2022). LIGO scientists investigating a set of glitches examine which auxiliary channels are recording noise simultaneously to look for insights into the causes of the glitches, for instance, noting that particular kinds of glitches increase in frequency in parallel with particular kinds of seismic noise.

### **MACHINE LEARNING SUPPORTING VOLUNTEER LEARNING**

Though the data volume poses some challenges, making AC data available to volunteers is straightforward. However, for volunteers to be able to make sense of them, a key challenge is their learning enough about the detector to understand potential connections. Similarly, ML must also learn about connections between glitches and activity in ACs. The problem is transforming the massive volume of data from disparate sources into useful information and enabling volunteers to learn how glitches relate to activity in different ACs.

Guided by the idea of the ZPD, we plan to facilitate volunteers' learning in part by staging their interactions with the data. In the first task, volunteers will examine individual glitches and seek to identify if there is a relation to activity in various individual ACs. This task is being performed in a Zooniverse project-builder project. We have identified a subset of several hundred ACs that are most informative. Following the first design principle, we plan to use ML to introduce glitches and also ACs gradually, so volunteers have time to learn the nature of that set of glitches or channels (the staging is arranged manually in the current prototype of the system). This staging will be supported by sorting glitches using the ML glitch classification models created for Gravity Spy and applying tools that assess correlations between glitches and AC activity (discussed below) to identify which ACs to present for a particular glitch. To learn about the process by which channels influence each other, volunteers need to have access to relevant information about the sensors, their spatial layout and the provenance of the datasets. This information is captured in a Wiki that supports the project, which is linked from the spectrograms presented.

### **VOLUNTEER LEARNING SUPPORTING MACHINE LEARNING**

As in Gravity Spy, ML will also be a learner, in particular, learning when a glitch and activity in an AC are related.

Currently, LIGO uses many algorithms (e.g., hveto [Smith et al. 2011], iDQ [Essick et al. 2020], Karoo GP [Cavaglia et al. 2019]) that identify statistically significant correlations between a loud event occurring in the main gravitational wave channel (a likely glitch) and an event in one of the ACs. As useful as these tools are for providing clues to the causes of glitches, statistical correlations represent an incomplete picture. Some channels experience loud events frequently, so the fact that they correlate with a glitch class might not be informative. Channels have complicated interdependencies (e.g., because they are in the same location or dependent on each other through feedback loops), so many channels can show a correlation with the same glitch. As a result, a channel may be a statistically significant witness for a class of glitch even though it is not actually close to the root cause. A further issue is that only some of the mechanisms connecting parts of the detector are well understood. Mechanisms can be complex and nonlinear, may involve complicated interactions (e.g., between environmental conditions and detector functioning), and some are yet to be discovered.

We are currently investigating ways to apply our ML models to this problem. Specifically, we are developing a system that, given a glitch, identifies which auxiliary channels are related (Wu et al. 2024). The Similarity Search tool can be applied to this problem to look for similarities between glitches in the main channel and in the auxiliary channels, but these can be related even without being identical. At the moment, we lack training data for a supervised learning approach, but the human volunteers are building the dataset, that is, the human volunteers will be guiding the machine learning, following design principle two.

We also see opportunities to leverage the work of LIGO scientists and volunteers to train ML models to establish the causal mechanisms of glitches. LIGO science teams monitor each detector and publish log books describing the internal state of the instrument and the physical environment around the detectors when issues arise. The logs reference auxiliary channels and other information about the detector's status. When presented to volunteers, they could conduct analyses that establish causal relationships between glitches and events at the detector site. As volunteers create textual data articulating the relationships, supervised ML models and NLP could learn from the descriptions and analyses provided by the volunteers, helping automate the identification of patterns and correlations. Additional design related to leveraging human knowledge are described in Crowston et al. (2023).

### **CO-LEARNING**

We also have identified ways humans and ML can cooperate on tasks that neither can do alone. As volunteers advance

in the project, we expect their and the ML's ZPDs will also have advanced, allowing them to handle more advanced activities. The volunteer task will shift from examining individual glitches to examining collections of glitches of a particular class, building on the filtered dataset of glitches and auxiliary channels created in the beginner stage. We strive to co-augment human and machine learning for each of these activities. Volunteers and ML together will create collections of glitches that might represent a novel glitch class with hypothesized common causality, not just similar appearance, and search for relations between those glitches and groups of ACs over time using a Similarity Search tool that we will build that includes both glitches and AC data, following design principle three.

Specifically, we will extend the existing similarity search tool to consider not only the morphological similarities of the glitches but also similarities in the pattern of AC data, that is, glitches that look similar and are related to a similar set of ACs. We are also exploring extending the Gravity Spy Similarity Search tool by fusing the main channel and related ACs and learning across both data sets. An interesting element of performing this fusion in the project's later stages is that the volunteers' work will have established the degree of similarity of each channel to the main channel. As a separate impact, the fused model may also further improve the performance of the ML glitch classifier for Gravity Spy. A question is whether the ACs convey complementary information concerning the classification task and can improve classification performance compared with the classifications obtained using only the spectrogram of the main channel.

Human and machine co-learning efforts call for a close dialogue between the wishes and needs of the citizen scientists, the science team, and our machine learning experts. We find an iterative process is most productive when we gradually move between design mockups shared with volunteers to preliminary ML tool models tested by science team members and advanced volunteers.

## DISCUSSION

Designing CS projects that secure human and machine co-learning by allowing humans to augment ML and AI to augment human learning offers several contributions to the existing literature and points to new research opportunities.

First, using ML to facilitate volunteer learning by keeping volunteers in their ZPD in and of itself does not add new insights to the literature on learning and AI. The concept of ZPD has been around since the 1930s and many scholars of learning have drawn on its principles since (Engeström

2001). Most of this literature emphasizes the role of other people and communities in keeping people in their ZPD as they gradually master harder and harder tasks. Artifacts, rules, and the division of labor have also been discussed as central to facilitating ZPD (Kaptelinin and Nardi 2009). The emerging literature on AI augmentation (Raisch and Krakowski 2021) likewise highlights the role AI can play in augmenting complex and ambiguous processes such as learning. However, the literature on augmentation tends to offer a static perspective on how AI can help people cope with their current job and not how AI can help people become better at their job. Adding the dynamic perspective of ZPD to the idea of augmentation offers a change in perspective useful to CS projects seeking to support volunteers' ongoing learning (i.e., applying design principle one).

Second, in the design of Gravity Spy, we have found the ZPD concept productive when trying to keep ML up-to-date with changes in the underlying science data from the LIGO detectors. We are not only designing an AI system for the project but expect that it will keep improving. Assuming augmentation can go both ways opens new design opportunities for AI-supported CS projects (applying design principle two). Yet, one needs to keep in mind that machines do not learn as humans do, driven by an innate desire to explore. Rather, learning requires human effort to add new data to training sets and to retrain models or expand them to handle new and novel data. There is also a possibility that learning might not be desirable. For instance, a new model might not perform as well as an existing one on important tasks. Likewise, some applications of machine learning could lead to unlearning among its human counterparts. If we designed the ML to completely take over the classification of certain glitch classes, for instance, we could expect to see a reduction in the volunteers' abilities to master these classes.

Third, addressing tasks that neither human nor AI can do independently stand out as the ultimate goal for this project, as synthesized in Figure 3 and design principle three. To enable co-learning requires careful analysis of the CS task progressions possible and the desired and possible human and machine learning processes. As suggested by the co-learning literature (van den Bosch et al. 2019), AI and humans have different strengths and weaknesses, with AI excelling in storing and processing information, while humans do not struggle as much as AI when it comes to abstract thinking, common sense, and knowledge transfer. Whether this assumption is true for all cases of AI and human co-learning situations is an open question and one that needs further investigation. One could imagine new forms of AI with improved capabilities when it comes to, for example, knowledge transfer and taking its context into account.

Further, neither human nor the AI's ZPD stands still. Co-learning is an evolving process in which the strengths and weaknesses of each entity change over time (van Zoelen et al. 2021). As their ZPDs change, so does the need for augmentation. Gravity Spy has just started to explore these opportunities and the identified design principles. It will require some rethinking of system development processes and approaches to make co-learning broadly available to CS projects. One might need to operate with multiple units of analysis when demarcating a co-learning system and the appropriate ZPDs. In the Gravity Spy case, for instance, all volunteers go through the initial classification stages and with it the scaffolded workflow, but many drop off before they reach the higher-level workflows and more advanced ZPDs. In other words, while all volunteers benefit from the AI support in the beginner stage, it is not all volunteers that assist the AI by finding new glitch classes in later workflows. Thus, a one-to-one approach to co-learning in which one human and one AI system support each other's progress might not be productive. Instead, one needs to take both individual and collective progress into account in the design process.

Finally, co-learning adds degrees of freedom to system design. AI systems need not completely automate tasks if we can assume that humans can fill in the gaps and facilitate improvements over time.

## CONCLUSION

We argue that as the opportunities for AI-support gradually become more accessible to project designers, it is beneficial to consider a bi-directional relationship between AI and volunteers. If one accepts this challenge, it is important to find ways that best secure human and machine learning by allowing humans to augment ML and AI to augment human learning. ZPD is a helpful perspective as it allows us to consider not only how machines can support volunteers' ongoing learning and volunteers can facilitate ML, but also how we can design tasks where each supports the other as their capabilities improve and the underlying tasks change.

The approach suggests scaffolding projects, so newcomers do not get overwhelmed or more experienced participants stuck in a rut. For instance, we hypothesize that the human-machine co-learning process found in Gravity Spy could benefit projects like Snapshot Serengeti in being more immediately accessible to newcomers. The approach also allows us to consider automation of some tasks. While automation is probably not an immediate goal for most CS projects, it is important to monitor ML performance. If it can do certain tasks without human support, one should

consider the value of having volunteers switch to other tasks to benefit themselves, the science team, and the ML.

## LIMITATIONS AND FUTURE RESEARCH

Herein, we have emphasized the role of volunteer and machine learning. Yet, more goes into the supporting learning on Gravity Spy. Tutorials, community access through discussion boards, science team involvement, and ML experts are important in facilitating both human and machine learning. These remain central for CS project success and must be considered regarding co-learning processes.

Detecting when volunteers are in their ZPD is also not always straightforward. This difficulty is particularly true on crowdsourced project platforms such as Zooniverse, where you cannot look over participants' shoulders to see if they are struggling. In Gravity Spy 2.0, we are experimenting with new ways of detecting when volunteers and machines enter their ZPD and determining what types of support would help. Being able to track people and machine performances becomes central. To date, we have mined trace data to detect what activities lead to project progress or stagnation. Such activity patterns can point to situations where participants tend to get stuck or practices that help people learn. Regarding machines' ZPD, discussion boards are helpful places to detect the shortcomings of AI tools. However, more research in these areas is required.

## DATA ACCESSIBILITY STATEMENT

Classified glitches from the Gravity Spy project are available (Glanzer et al. 2021). Our IRB protocols do not permit sharing human subject data.

## SUPPLEMENTAL FILE

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

- **Supplemental File 1: Table 1.** Table of glitches available to classify in each workflow in Gravity Spy. DOI: <https://doi.org/10.5334/cstp.738.s1>

## ETHICS AND CONSENT

All human subjects research was reviewed and determined to be exempt by the Syracuse University Institutional Review Board (IRB) (protocols 17–183 and 22–215).

## ACKNOWLEDGEMENTS

The authors acknowledge their collaborators on the Gravity Spy project, the LIGO Scientific Collaboration, which provides data for the system, and the many volunteers, without whose contributions the project would not be possible.

## FUNDING INFORMATION

Gravity Spy was funded by grants from the US National Science Foundation, INSPiRE 15–47880 and HCC 21–06865.

## COMPETING INTERESTS

The authors have no competing interests to declare.

## AUTHOR CONTRIBUTIONS

CØ and KC: Conceptualization. CØ, KC, CJ, AS: Investigation, Writing – Original Draft. YW: Software, Formal Analysis. CØ, KC, CJ, AK: Supervision, Funding acquisition.

## AUTHOR AFFILIATIONS

**Carsten Østerlund**  [orcid.org/0000-0003-0612-1551](https://orcid.org/0000-0003-0612-1551)

Syracuse University, School of Information Studies, US

**Kevin Crowston**  [orcid.org/0000-0003-1996-3600](https://orcid.org/0000-0003-1996-3600)

Syracuse University, US

**Corey B. Jackson**  [orcid.org/0000-0003-0828-4506](https://orcid.org/0000-0003-0828-4506)

University of Wisconsin-Madison, School of Computer, Data, and Information Sciences, US

**Yunan Wu**  [orcid.org/0000-0001-6980-9746](https://orcid.org/0000-0001-6980-9746)

Northwestern University, Department of Electrical and Computer Engineering, US

**Alexander O. Smith**  [orcid.org/0000-0002-3719-2232](https://orcid.org/0000-0002-3719-2232)

Syracuse University, School of Information Studies, US

**Aggelos K. Katsaggelos**  [orcid.org/0000-0003-4554-0070](https://orcid.org/0000-0003-4554-0070)

Northwestern University, Department of Electrical and Computer Engineering, US

## REFERENCES

**Aasi, J., Abbott, B.P., Abbott, R., Abbott, T., Abernathy, M.R., Ackley, K., Adams, C., et al.** (LIGO scientific collaboration) (2015) Advanced LIGO. *Classical and Quantum Gravity* 32, 074001. DOI: <https://doi.org/10.1088/0264-9381/32/7/074001>

**Agrawal, A., Gans, J. S., Goldfarb, A.** (2023) Do we want less automation? *Science* 381(6654), pp. 155–158. DOI: <https://doi.org/10.1126/science.adh9429>

**Bahaadini, S., Noroozi, V., Rohani, N., Coughlin, S., Zevin, M., Katsaggelos, A.K.** (2018a) DIRECT: Deep discriminative embedding for clustering of LIGO data, in: *2018 25th IEEE International Conference on Image Processing (ICIP)*. pp. 748–752. DOI: <https://doi.org/10.1109/ICIP.2018.8451708>

**Bahaadini, S., Noroozi, V., Rohani, N., Coughlin, S., Zevin, M., Smith, J.R., Kalogera, V., Katsaggelos, A.** (2018b) Machine learning for Gravity Spy: Glitch classification and dataset. *Information Sciences* 444, pp. 172–186. DOI: <https://doi.org/10.1016/j.ins.2018.02.068>

**Blake, M.B., Butcher-Green, J.D.** (2009) Agent-customized training for human learning performance enhancement. *Computers & Education* 53, pp. 966–976. DOI: <https://doi.org/10.1016/j.compedu.2009.05.014>

**Brynjolfsson, E., McAfee, A.** (2014) *The second machine age: Work, progress, and prosperity in a time of brilliant technologies*. WW Norton & Company.

**Cavaglia, M., Staats, K., Gill, T.** (2019) Finding the origin of noise transients in LIGO data with machine learning. *Communications in Computational Physics CiCP* 25. DOI: <https://doi.org/10.4208/cicp.OA-2018-0092>

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

**Chatterji, S., Blackburn, L., Martin, G., Katsavounidis, E.** (2004) Multiresolution techniques for the detection of gravitational-wave bursts. *Classical and Quantum Gravity*. 21, S1809–S1818. DOI: <https://doi.org/10.1088/0264-9381/21/20/024>

**Coughlin, S., Bahaadini, S., Rohani, N., Zevin, M., Patane, O., Harandi, M., Jackson, C., et al.** (2019) Classifying the unknown: Discovering novel gravitational-wave detector glitches using similarity learning. *Physical Review D*. 99, 082002. DOI: <https://doi.org/10.1103/PhysRevD.99.082002>

**Crowston, K., Jackson, C.B., Corieri, I., Østerlund, C.** (2023) Design principles for background knowledge to enhance learning in citizen science. Barcelona, Spain and virtual. DOI: [https://doi.org/10.1007/978-3-031-28032-0\\_43](https://doi.org/10.1007/978-3-031-28032-0_43)

**Crowston, K., Mitchell, E., Østerlund, C.** (2018) Coordinating advanced crowd work: Extending citizen science. *Hawaii International Conference on System Sciences 2018 (HICSS-51)*. DOI: <https://doi.org/10.24251/HICSS.2018.212>

**Danielsen, F., Burgess, N.D., Balmford, A.** (2005) Monitoring matters: Examining the potential of locally-based approaches. *Biodiversity & Conservation* 14, pp. 2507–2542. DOI: <https://doi.org/10.1007/s10531-005-8375-0>

- Daugherty, P.R., Wilson, H.J.** (2018) *Human + machine: Reimagining work in the age of AI*. Harvard Business Press.
- Davenport, T.H., Kirby, J.** (2016) *Only humans need apply: Winners and losers in the age of smart machines*. Harper Business New York.
- Davis, D., Walker, M.** (2022) Detector characterization and mitigation of noise in ground-based gravitational-wave interferometers. *Galaxies* 10. DOI: <https://doi.org/10.3390/galaxies10010012>
- Engeström, Y.** (2001) Expansive learning at work: Toward an activity theoretical reconceptualization. *Journal of Education and Work* 14, pp. 133–156. DOI: <https://doi.org/10.1080/13639080020028747>
- Essick, R., Godwin, P., Hanna, C., Blackburn, L., Katsavounidis, E.** (2020) iDQ: Statistical inference of non-gaussian noise with auxiliary degrees of freedom in gravitational-wave detectors. *Machine Learning: Science and Technology* 2, 1. DOI: <https://doi.org/10.1088/2632-2153/abab5f>
- Glanzer, J., Banagari, S., Coughlin, S., Zevin, M., Bahaadini, S., Rohani, N., Allen, S., et al.** (2021) Gravity Spy machine learning classifications of LIGO glitches from observing runs O1, O2, O3a, and O3b. *Zenodo*. DOI: <https://doi.org/10.5281/zenodo.5649212>
- Gregor, S., Jones, D.** (2007) The anatomy of a design theory. *Association for Information Systems*.
- Hevner, A.R., March, S.T., Park, J., Ram, S.** (2008) Design science in information systems research. *Management Information Systems Quarterly* 28, 6. DOI: <https://doi.org/10.2307/25148625>
- Huang, Y.-C., Cheng, Y.-T., Chen, L.-L., Hsu, J.Y.** (2019) Human-AI Co-Learning for Data-Driven AI. DOI: <https://doi.org/10.48550/arXiv.1910.12544>
- Jackson, C.B., Østerlund, C., Crowston, K., Harandi, M., Allen, S., Bahaadini, S., Coughlin, S., et al.** (2020a) Teaching citizen scientists to categorize glitches using machine-learning-guided training. *Computers in Human Behavior* 105. DOI: <https://doi.org/10.1016/j.chb.2019.106198>
- Jackson, C.B., Østerlund, C., Harandi, M., Crowston, K., Trouille, L.** (2020b) Shifting forms of engagement: Volunteer learning in online citizen science. *Proceedings of the ACM on Human-Computer Interaction* 36. DOI: <https://doi.org/10.1145/3392841>
- Kaptelinin, V., Nardi, B.A.** (2009) *Acting with technology: Activity theory and interaction design*. The MIT Press, Cambridge, Mass. London.
- Kent, C., Chaudhry, M.A., Cukurova, M., Bashir, I., Pickard, H., Jenkins, C., du Boulay, B., Moeini, A., et al.** (2021) Machine learning models and their development process as learning affordances for humans, in: Roll, I., McNamara, D., Sosnovsky, S., Luckin, Rose, Dimitrova, V. (Eds.), *Artificial Intelligence in Education*. Springer International Publishing, Cham, pp. 228–240. DOI: [https://doi.org/10.1007/978-3-030-78292-4\\_19](https://doi.org/10.1007/978-3-030-78292-4_19)
- Krizhevsky, A., Sutskever, I., Hinton, G.E.** (2012) Imagenet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems* 25.
- Kullenberg, C., Kasperowski, D.** (2016) What is citizen science?—A scientometric meta-analysis. *PLoS one* 11, e0147152. DOI: <https://doi.org/10.1371/journal.pone.0147152>
- Mugar, G., Østerlund, C., Hassman, K.D., Crowston, K., Jackson, C.B.** (2014) Planet hunters and seafloor explorers: legitimate peripheral participation through practice proxies in online citizen science, in: *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing*. ACM, pp. 109–119. DOI: <https://doi.org/10.1145/2531602.2531721>
- Nguyen, H.N., Gonzalez, J., Guo, J., Nguyen, N.V.T., Dang, T.** (2021) VisMCA: A visual analytics system for misclassification correction and analysis. *VAST Challenge 2020*, Mini-Challenge 2 Award: Honorable Mention for Detailed Analysis of Patterns of Misclassification.
- Norouzzadeh, M.S., Nguyen, A., Kosmala, M., Swanson, A., Palmer, M.S., Packer, C., Clune, J.** (2018) Automatically identifying, counting, and describing wild animals in camera-trap images with deep learning. *Proceedings of the National Academy of Sciences of the United States of America*, 115. DOI: <https://doi.org/10.1073/pnas.1719367115>
- Nuttall, L.K.** (2018) Characterizing transient noise in the LIGO detectors. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 376, 20170286. DOI: <https://doi.org/10.1098/rsta.2017.0286>
- Peppers, K., Tuunanen, T., Rothenberger, M.A., Chatterjee, S.** (2007) A design science research methodology for information systems Research. *Journal of Management Information Systems* 24, pp. 45–77. DOI: <https://doi.org/10.2753/MIS0742-1222240302>
- Raisch, S., Krakowski, S.** (2021) Artificial intelligence and management: The automation–augmentation paradox. *Academy of Management Review* 46, pp. 192–210. DOI: <https://doi.org/10.5465/amr.2018.0072>
- Riesch, H., Potter, C.** (2014) Citizen science as seen by scientists: Methodological, epistemological and ethical dimensions. *Public Understanding of Science* 23, pp. 107–120. DOI: <https://doi.org/10.1177/0963662513497324>
- Schoonderwoerd, T.A.J., Zoelen, E.M. van, Bosch, K. van den, Neerincx, M.A.** (2022) Design patterns for human-AI co-learning: A wizard-of-Oz evaluation in an urban-search-and-rescue task. *International Journal of Human-Computer Studies* 164, 102831. DOI: <https://doi.org/10.1016/j.ijhcs.2022.102831>
- Simpson, R., Page, K. R., De Roure, D.** (2014, April) Zooniverse: observing the world’s largest citizen science platform. In *Proceedings of the 23rd international conference on world wide web* (pp. 1049–1054). DOI: <https://doi.org/10.1145/2567948.2579215>

- Smith, J.R., Abbott, T., Hirose, E., Leroy, N., Macleod, D., McIver, J., Saulson, P., Shawhan, P.** (2011) A Hierarchical Method for Vetoing Noise Transients in Gravitational-Wave Detectors. *Classical and Quantum Gravity* 28, 235005. DOI: <https://doi.org/10.1088/0264-9381/28/23/235005>
- Soni, S., Berry, C.P.L., Coughlin, S.B., Harandi, M., Jackson, C.B., Crowston, K., Østerlund, C., et al.** (2021) Discovering features in gravitational-wave data through detector characterization, citizen science and machine learning. *Classical and Quantum Gravity* 38, 195016. DOI: <https://doi.org/10.1088/1361-6382/ac1ccb>
- van den Bosch, K., Schoonderwoerd, T., Blankendaal, R., Neerincx, M.** (2019) Six challenges for human-AI co-learning, in: Sottolare, R.A., Schwarz, J. (Eds.), *Adaptive instructional systems*. Springer International Publishing, Cham, pp. 572–589. DOI: [https://doi.org/10.1007/978-3-030-22341-0\\_45](https://doi.org/10.1007/978-3-030-22341-0_45)
- van Zoelen, E.M., van den Bosch, K., Neerincx, M.** (2021) Becoming team members: Identifying interaction patterns of mutual adaptation for human-robot co-learning. *Front. Robot. AI* 8. DOI: <https://doi.org/10.3389/frobt.2021.692811>
- Vygotsky, L.S., Cole, M.** (1978) *Mind in Society: Development of Higher Psychological Processes*. Harvard university press.
- Wilder, B., Horvitz, E., Kamar, E.** (2020) Learning to complement humans. DOI: <https://doi.org/10.24963/ijcai.2020/212>
- Willi, M., Pitman, R.T., Cardoso, A.W., Locke, C., Swanson, A., Boyer, A., Veldthuis, M., et al.** (2019) Identifying animal species in camera trap images using deep learning and citizen science. *Methods in Ecology and Evolution* 10, pp. 80–91. DOI: <https://doi.org/10.1111/2041-210X.13099>
- Wu, Y., Zevin, M., Berry, C.P., Crowston, K., Østerlund, C., Doctor, Z., Banagiri, S., et al.** (2024) Advancing glitch classification in Gravity Spy: Multi-view fusion with attention-based machine learning for advanced LIGO's fourth observing run. arXiv preprint [arXiv:2401.12913](https://arxiv.org/abs/2401.12913).
- Zevin, M., Coughlin, S., Bahaadini, S., Besler, E., Rohani, N., Allen, S., Cabero, M., et al.** (2017) Gravity Spy: integrating advanced LIGO detector characterization, machine learning, and citizen science. *Classical and Quantum Gravity* 34, 064003. DOI: <https://doi.org/10.1088/1361-6382/aa5cea>
- Zevin, M., Jackson, C.B., Doctor, Z., Wu, Y., Østerlund, C., Johnson, L.C., Berry, C.P.L., et al.** (2024) Gravity Spy: Lessons learned and a path forward. *The European Physical Journal Plus* 139, 100. DOI: <https://doi.org/10.1140/epjp/s13360-023-04795-4>

#### TO CITE THIS ARTICLE:

Østerlund, C., Crowston, K., Jackson, C.B., Wu, Y., Smith, A.O. and Katsaggelos, A.K. 2024. Supporting Human and Machine Co-Learning in Citizen Science: Lessons From Gravity Spy. *Citizen Science: Theory and Practice*, 9(1): 42, pp. 1–14. DOI: <https://doi.org/10.5334/cstp.738>

**Submitted:** 16 February 2024    **Accepted:** 22 August 2024    **Published:** 09 December 2024

#### COPYRIGHT:

© 2024 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>.

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