Mutual Learning in Human-AI Interaction

CARSTEN ØSTERLUND and KEVIN CROWSTON, School of Information Studies, Syracuse University, USA
COREY B. JACKSON, Information School, University of Wisconsin-Madison, USA
MARLENE TAKOU-AYAOH, College of Engineering and Computer Science, Syracuse University, USA
YUNAN WU and AGGELOS K. KATSAGGELOS, Department of Electrical and Computer Engineering, Northwestern University, USA

We explore the bi-directional relationship between human and machine learning in citizen science. Theoretically, the study draws on the Zone of Proximal Development concept, which allows us to describe the augmentation of human learning by AI, human augmentation of machine learning and how tasks can be designed to facilitate co-augmentation. Methodologically, the study utilizes 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-augmentation, where both humans and machines contribute to each other's learning and capabilities. The research contributes to the existing literature by emphasizing the role of ZPD in citizen science projects, showcasing how the concept supports ongoing learning for volunteers and keeps machine learning aligned with evolving data.

CCS Concepts: • **Theory of computation** → *Models of learning*; • **Human-centered computing** → *Empirical studies in collaborative and social computing*; • **Applied computing** → *Computer-managed instruction.*

Additional Key Words and Phrases: learning, zone of proximal development, human-artificial intelligence augmentation

ACM Reference Format:
Carsten Østerlund, Kevin Crowston, Corey B. Jackson, Marlene Takou-Ayaoh, Yunan Wu, and Aggelos K. Katsaggelos. 2018. Mutual Learning in Human-AI Interaction. In *Proceedings of Trust and Reliance in Evolving Human-AI Workflows Workshop at CHI 2024 (TREW '24).* ACM, New York, NY, USA, 6 pages. https://doi.org/XXXXXXX.XXXXXXX

1 INTRODUCTION

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

The possibility for a mutually beneficial system is particularly compelling for citizen science (CS), that is, scientific projects that involve members of the general public as contributors. An increasing number of CS projects deploy AI technologies [5], e.g., 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.

The challenge is how to facilitate human and machine learning so that the two do not simply counter one another. For instance, ML systems simply automating volunteers' tasks might 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 [15, 18]. While this scenario might facilitate human learning in a project, attention must also be paid to machine learning. For instance, many algorithms struggle with novel categories in the data that humans can detect, meaning that they would improve if they could learn from humans. From a project design perspective, strategies are needed to facilitate and build synergies between human and machine learning. The notion of augmentation [16] can be helpful where augmentation approaches AI as a collaborative tool, emphasizing close cooperation between humans and AI. Scholarly investigations suggest that augmentation is advantageous for complex and ambiguous processes such as learning [4, 8, 9].

2 THEORY

Learning theories focus on the development of internalized knowledge or skills that create a lasting behaviour change [10]. 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 to observable behaviours of people and AI working to achieve some objective. Skills or tasks are not assessed as abstract demands; instead, what matters is the exhibition of skills and demands through the process of achieving an objective [13, p. 31]. For instance, Kaptelinin & Nardi [13] note that saying that someone is "good at math" can be misleading since performance can vary significantly depending on how the problem is posed [13, p. 31].

We can thus distinguish 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, tasks that are done with assistance increase 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 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, e.g., tutorial materials, feedback, or what Mugar et al. [14] 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 for pattern recognition. Some of the patterns may be readily apparent even to newcomers to a project. Galaxy Zoo asked about simple shapes and most Snapshot Serengeti volunteers can 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. There will also be tasks that volunteers cannot do even with assistance, e.g., identifying animals at 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 does not matter if 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. We can have the situation in which a machine can do some tasks without assistance (i.e., automation), other tasks the machine can do only with assistance, and many tasks the machine cannot do even with assistance. In short, we envision a ZPD supporting machine learning 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. The human and the machine ZPD augment each other by assisting the other in their ZPD, i.e., helping them to do tasks they can not do alone.

3 METHOD

To explore these ideas, we have designed, deployed, and evaluated a CS project called Gravity Spy [20, 21], hosted on the Zooniverse platform [17]. Gravity Spy supports the Laser Interferometer Gravitational-Wave Observatory [LIGO, 1], 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 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 known or unknown causes. 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. We draw on many different sources of data collected throughout the project: interviews with LIGO and machine learning 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., 7, 11, 12].

4 FINDINGS

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. In doing so, 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.

4.1 Machine classification supporting human learning

In many image-classification citizen-science 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 ML confidently classified are assigned to beginner workflows. Multiple beginner workflows contain an increasing number of glitch classes.

Specifically, in workflow 1, volunteers are currently shown glitches that the ML has classified with high confidence as belonging to just two common and easily distinguished glitch classes: Blip and Whistle. High ML confidence for a glitch means they are likely (though not certain) to be examples of a class. The classification interface offers just those two options, plus "None of the above" to capture instances where the ML is mistaken. 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 work flow 2, volunteers are shown additional glitches with high ML confidence, i.e., 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.

4.2 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 or to override them if enough volunteers disagree with the ML classification. The training set has now been supplemented with glitches classified by volunteers to include nearly 10,000 labelled glitches over 23 classes [21] and 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 should be involved in classifying all glitches (i.e., the machine seems capable of some tasks without assistance).

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. In other settings, these issues might be due to out-of-distribution data, but for LIGO, they are expected, 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 [19]. After 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 as Whistles. Meanwhile, the human volunteers had little trouble distinguishing the novel glitches and brought them to the science team's attention relatively quickly (that is, even the novice workers in work flow 1 were able to correct the ML's misinterpretation). 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.

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.

4.3 Co-augmentation

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, meaning that the knowledge is not completely complementary. We will take the development of new glitch classes as an example of such co-augmentation. 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 [2, 3, 6]. 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 a co-augmentation where human learners (of all levels of expertise) assist the machine in learning new glitch classes while the machine assists the humans (mostly experts) by easing the burden of sifting through a large dataset in search of those new classes.

In summary, the Gravity Spy project illustrates how different learners can support each other: machine learning 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.

ACKNOWLEDGMENTS

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

REFERENCES

[1] J. Aasi and LIGO Scientific Collaboration. 2015. Advanced LIGO. *Classical and Quantum Gravity* 32 (2015), 074001. https://doi.org/10.1088/0264-9381/32/7/074001

[2] S. Bahaadini, V. Noroozi, N. Rohani, S. Coughlin, M. Zevin, and A.K. Katsaggelos. 2018. DIRECT: Deep Discriminative Embedding for Clustering of LIGO Data. In *2018 25th IEEE International Conference on Image Processing (ICIP)*. 748–752. https://doi.org/10.1109/ICIP.2018.8451708

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

[4] E. Brynjolfsson and A. McAfee. 2014. *The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Technologies.* WW Norton & Company.

[5] L. Ceccaroni, J. Bibby, E. Roger, P. Flemons, K. Michael, L. Fagan, and J.L. Oliver. 2019. Opportunities and Risks for Citizen Science in the Age of Artificial Intelligence. *Citizen Science: Theory and Practice* 4 (2019), 29. https://doi.org/10.5334/cstp.241

[6] S. Coughlin, S. Bahaadini, N. Rohani, M. Zevin, O. Patane, M. Harandi, C. Jackson, V. Noroozi, S. Allen, J. Areeda, M. Coughlin, P. Ruiz, C.P.L. Berry, K. Crowston, A.K. Katsaggelos, A. Lundgren, C. Østerlund, J.R. Smith, L. Trouille, and V. Kalogera. 2019. Classifying the unknown: Discovering novel gravitational-wave detector glitches using similarity learning. *Phys. Rev. D* 99 (2019), 082002. https://doi.org/10.1103/PhysRevD.99.082002

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

[8] P.R. Daugherty and H.J. Wilson. 2018. *Human+ Machine: Reimagining Work in the Age of AI.* Harvard Business Press.

[9] D. Davis and M. Walker. 2022. Detector Characterization and Mitigation of Noise in Ground-Based Gravitational-Wave Interferometers. *Galaxies* 10 (2022). https://doi.org/10.3390/galaxies10010012

[10] Y. Engeström. 2001. Expansive Learning at Work: Toward an activity theoretical reconceptualization. *Journal of Education and Work* 14 (2001), 133–156. https://doi.org/10.1080/13639080020028747

[11] C.B. Jackson, C. Østerlund, K. Crowston, M. Harandi, S. Allen, S. Bahaadini, S. Coughlin, V. Kalogera, A. Katsaggelos, S. Larson, N. Rohani, J. Smith, L. Trouille, and M. Zevin. 2020. Teaching citizen scientists to categorize glitches using machine-learning-guided training. *Computers in Human Behavior* 105 (2020). https://doi.org/10.1016/j.chb.2019.106198

[12] C.B. Jackson, C. Østerlund, M. Harandi, K. Crowston, and L. Trouille. 2020. Shifting forms of Engagement: Volunteer Learning in Online Citizen Science. *Proceedings of the ACM on Human-Computer Interaction* 36 (2020). https://doi.org/10.1145/3392841

[13] V. Kaptelinin and B.A. Nardi. 2009. *Acting with Technology: Activity Theory and Interaction Design.* The MIT Press, Cambridge, Mass. London.

[14] G. Mugar, C. Østerlund, K.D. Hassman, K. Crowston, and C.B. Jackson. 2014. Planet hunters and seafloor explorers: legitimate peripheral participation through practice proxies in online citizen science. (2014), 109–119.

[15] M.S. Norouzzadeh, A. Nguyen, M. Kosmala, A. Swanson, M.S. Palmer, C. Packer, and J. Clune. 2018. Automatically identifying, counting, and describing wild animals in camera-trap images with deep learning. *Proc. Natl. Acad. Sci. U.S.A.* 115 (2018). https://doi.org/10.1073/pnas.1719367115

[16] S. Raisch and S. Krakowski. 2021. Artificial Intelligence and Management: The Automation–Augmentation Paradox. *AMR* 46 (2021), 192–210. https://doi.org/10.5465/amr.2018.0072

[17] R. Simpson, K.R. Page, and D. De Roure. 2014. Zooniverse: Observing the world's largest citizen science platform. In *Proceedings of the 23rd International Conference on World Wide Web.* 1049–1054.

[18] M. Willi, R.T. Pitman, A.W. Cardoso, C. Locke, A. Swanson, A. Boyer, M. Veldthuis, and L. Fortson. 2019. Identifying Animal Species in Camera Trap Images Using Deep Learning and Citizen Science. *Methods Ecol Evol* 10 (2019), 80–91. https://doi.org/10.1111/2041-210X.13099

[19] Y. Wu, M. Zevin, C.P. Berry, K. Crowston, C. Østerlund, Z. Doctor, S. Banagiri, C.B. Jackson, V. Kalogera, and A.K. Katsaggelos. 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* (2024).

[20] M. Zevin, S. Coughlin, S. Bahaadini, E. Besler, N. Rohani, S. Allen, M. Cabero, K. Crowston, A.K. Katsaggelos, S.L. Larson, T.K. Lee, C. Lintott, T. Littenberg, A. Lundgren, C. Østerlund, J. Smith, L. Trouille, and V. Kalogera. 2017. Gravity Spy: integrating advanced LIGO detector characterization, machine learning, and citizen science. *Classical and Quantum Gravity* 34 (2017), 064003. https://doi.org/10.1088/1361-6382/aa5174

[21] M. Zevin, C. B. Jackson, Z. Doctor, Y. Wu, C. Østerlund, L. C. Johnson, C. P. L. Berry, K. Crowston, S. B. Coughlin, V. Kalogera, S. Banagiri, D. Davis, J. Glanzer, R. Hao, A. K. Katsaggelos, O. Patane, J. Sanchez, J. Smith, S. Soni, L. Trouille, M. Walker, I. Aerith, W. Domainko, V.-G. Baranowski, G. Niklasch, and B. Téglás. 2024. Gravity Spy: lessons learned and a path forward. *The European Physical Journal Plus* 139, 1 (Jan. 2024), 100. https://doi.org/10.1140/epjp/s13360-023-04795-4

Received 26 February 2024