CITIZEN SCIENCE:
THEORY AND PRACTICE

Image Recognition as a "Dialogic AI Partner" Within Biodiversity Citizen Science—an empirical investigation

COLLECTION: THE FUTURE OF ARTIFICIAL INTELLIGENCE AND CITIZEN SCIENCE

RESEARCH PAPER

ubiquity press

NIRWAN SHARMA
LAURA COLUCCI-GRAY
POPPY LAKEMAN-FRASER
ANNIE ROBINSON
JULIE NEWMAN
RENÉ VAN DER WAL
STEFAN RUEGER
ADVAITH SIDDHARTHAN

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

ABSTRACT

We investigate the potential of a new citizen science paradigm that facilitates collaborative learning between humans and artificial intelligence (AI). Recognising the potential of AI to support and empower rather than replace human participation, we explore the integration of image recognition as a 'dialogic AI partner' in citizen science (CS) projects, interacting with participants in real time. We study this in the context of a biodiversity monitoring project that relies on volunteers to identify biological species from images taken in the wild. Guided by the idea of Bakhtin's dialogism and Bayesian inference principles, we developed a web interface that integrated an image recognition model, fine-tuned for classifying 22 UK bumblebee species, into an interactive interface based on visual feature keys to enable real-time dialogue between humans and AI. We report a significant improvement in identification accuracy for both humans and AI when they engage in such dialogue and retain the ability to reach independent conclusions rather than achieve consensus. Given the inherent need for convergence in decision-making within scientific processes such as species identification tasks, we augmented the dialogic process with a Bayesian model that unifies potentially divergent human and AI perspectives post collaboration to achieve a more accurate consensus decision than that achieved by either AI or citizens. Our work provides new understandings around the design of a dialogic space for CS practice that effectively builds on the complementary strengths of human and AI visual recognition approaches.

CORRESPONDING AUTHOR:
Nirwan Sharma
The Open University, UK
nirwan.sharma@open.ac.uk

KEYWORDS:
biodiversity citizen science; dialogism; human-AI collaboration; species identification; machine learning; Bayesian reasoning; image recognition

TO CITE THIS ARTICLE:
Sharma, N, Colucci-Gray, L, Lakeman-Fraser, P, Robinson, A, Newman, J, Van der Wal, R, Rueger, S and Siddharthan, A. 2024. Image Recognition as a "Dialogic AI Partner" Within Biodiversity Citizen Science—an empirical investigation. *Citizen Science: Theory and Practice,* 9(1): 35, pp. 1–16. DOI: https://doi.org/10.5334/cstp.735

INTRODUCTION

The integration of artificial intelligence (AI) technologies into society has raised significant concerns about the potential replacement of human effort and consequent de-skilling of individuals. These concerns are similarly expressed within citizen science (CS) communities, where the increasing utilisation of AI for data collection and analysis poses the risk that valuable skills and knowledge traditionally developed by volunteers may diminish (Sambasivan and Veeraraghavan 2022). However, alongside these concerns, AI integration also offers opportunities for enhancing CS projects by not only processing and analysing large datasets, but also by introducing novel methods for public participation, thus enriching the volunteer experience and fostering deeper engagement. This can lead to a potentially transformative shift in public engagement with scientific research. We propose that a human-AI collaborative partnership may be central to this integration. Here, we report the findings of a study conducted in the context of biodiversity monitoring that included the use of a specific AI for image recognition. We assessed the particular role this played in supporting and empowering the collective efforts of citizen science volunteers engaged in collaborative species identification activities.

BACKGROUND

Biological recording has a long history of utilising efforts from the general public to record and identify biological species for use by science and society (Pocock et al. 2015; Silvertown 2009). This field faces unique challenges; For instance, the UK alone is home to over 70,000 species, many of which are not well known to the public beyond familiar groups like mammals or birds (Hopkins and Freckleton 2002). Furthermore, environmental change further necessitates continuous monitoring, as species distributions are ever-changing (Pimm et al. 2014). Central to effective monitoring is the ability to accurately identify species, a task that requires familiarity with species observation and recording, as well as problem solving and decision-making skills. Recent AI advancements are fuelling interest in human-AI collaboration for tasks like species identification (Memmert and Bittner 2022; Palmer et al. 2021; Truong and Van der Wal 2024) as automated image recognition plays a crucial role, both in lab environments and in CS projects involving photo submissions, highlighting the integration of technology in traditional practices (Martineau et al. 2017; Horn et al. 2018). However, in CS projects, AI predictions are presented to the participants as either suggested or likely identifications (Truong and Van der Wal 2024), with no meaningful collaboration with participants around the task, thereby limiting opportunities for learning within such projects. Moreover, effective use of image recognition for photos by CS participants is further challenged by technical issues like inconsistent image quality and biases affecting less common species. Addressing these issues through effective interaction design for human-AI collaboration is thus important to help improve data quality for ecological research and also enhance learning and engagement within CS initiatives.

The introduction of technology is often associated with concerns of deskilling and disempowerment (Rafner et al. 2022a). Hence, as we integrate more sophisticated technologies into CS, it is important to address new concerns and explore how AI can augment, rather than diminish, the skills of volunteers. For volunteers, the honing of identification skills is a key motivator of species recording (Ellis 2011; Sharma et al. 2019). Hence, designing technology that provides opportunities for learning for humans as well as machines is crucial.

To address such concerns, we offer a productive point of departure by drawing on emerging conceptions in AI literature, suggesting that AI technologies are best integrated as complementary to human input for decision-making tasks (Steyvers et al. 2022). Such an approach aims to improve efficiency but also increase volunteer engagement in citizen science projects by preserving opportunities for unexpected, serendipitous findings (Trouille et al. 2019). Hence, in practice, AI technologies can be designed as partners capable of collaborating with human participants to enhance their experience and learning. While human-AI collaboration through dialogue is largely well established in language-based applications, such as through conversational agents and large language models (LLMs) like ChatGPT, its application to visual tasks for biodiversity conservation remains largely unexplored. This includes human-AI dialogues in the context of visual tasks such as species identification from images. To address this, we draw upon a prior project in biodiversity CS to examine the potential for collaboration between human participants and AI technologies in CS with a framework derived from the linguistic tradition of dialogism. Our approach combines ideas of dialogism (Bakhtin 1984) and Bayesian inference as applied within CS (e.g., Siddharthan et al. 2016). By designing AI as a dialogic partner in visual tasks, we aim to create a richer, more interactive dialogue between human participants and AI systems, enhancing learning and improving species identification.

HUMAN-AI DIALOGUE FOR CITIZEN SCIENCE PRACTICE

Human-AI collaboration has been envisioned since the late 1950s through the work of Ashby on cybernetics and intelligence amplification (Ashby 1957), Licklider's concept of man-computer symbiosis (Licklider 1960) and Englebart's vision of augmenting human intellect (Engelbart 1962), all advocating systems that promote human-computer cooperation for societal benefit. This idea has further evolved into concepts like extended intelligence, hybrid intelligence, and collective intelligence (Dellermann et al. 2019; Peeters et al. 2021; Clark and Chalmers 1998). Although mainstream research has primarily focused on computer-supported human collaboration across disciplines (Dillenbourg 1999; Jeong and Hmelo-Silver 2016; Ludvigsen et al. 2021), advancements in machine learning have brought into focus the role AI tools can play in augmenting human decision-making in real-world settings (Reverberi et al. 2022; Tejeda et al. 2022).

As AI becomes more integrated in society, new models of knowledge co-production may be required to guide the design of interactive systems, fostering rich, multi-perspective interactions between AI and users. To this end, the framework of dialogism, first introduced by Mikhail Bakhtin, focuses on the emergence of meaning through a dynamic interaction of multiple voices or perspectives while maintaining independence within a conversation. In citizen science communities, knowledge creation can be considered as a collaborative dialogue among diverse participants from across a gradient ranging from expert to non-expert (Bakhtin 1984; Stahl et al. 2014; Trausan-Matu, Wegerif, and Major 2021). Adopting this concept, we aim to integrate AI as an independent voice in citizen science processes, utilising its analytical capabilities and unique data-driven insights to enrich the dialogue beyond what human participants can achieve themselves. This independent voice can be crafted through a carefully designed dialogic space to facilitate real-time, meaningful exchanges between humans and AI. As Wegerif and Major (2019) argue, while AI lacks human qualities, its 'ontological ambivalence'—when designed and presented as an independent voice or a dialogic partner for language-based applications—allows it to simulate engaging dialogues with humans (Wegerif and Major 2019). The intentional design of these interactions and underlying technology is crucial as it ensures that AI does not merely function as an automated tool but can act as a dynamic participant that enhances dialogic interactions. This approach demands careful design and specific affordances in interactive settings to accommodate computer-supported human-AI interactions, ultimately leading to an expansion of the dialogic space (Wegerif 2024).

In our study we facilitate these interactions between AI and human users through the development and assessment of an interface for identifying bumblebee species, a visual task that requires the critical skill of comparing and contrasting specimens with similar features. In this context of identifying species, exchange and evaluation of different viewpoints facilitates divergence and convergence of opinions, negotiation, self-reflection, and inference (Trausan-Matu, Wegerif, and Major 2021; Sharma et al. 2022). This method is especially effective in tasks like species identification, where distinguishing visual features between similar species is essential. By simulating a multivocal approach, participants can be guided to look closely and evaluate these features by taking into account different perspectives, thus enhancing their identification skills (Ellis 2011). Interacting with AI in this manner may prompt participants to attend to visual features that might have been initially overlooked, and encourage them to move beyond their inherent biases by considering alternative suggestions and details highlighted by AI models. This approach likely fosters richer collaboration and improved outcomes, enhancing both learning and data quality (Rezwana and Maher 2022).

Given the inherent need for convergence within scientific processes, we further focus on consensus-based decision-making based on the outcomes of human and AI participants, especially if the participants come to different decisions. We study decision-making processes where human and AI participants individually arrive at their decisions and, post collaboration, we apply Bayesian consensus modelling to integrate these decisions (Mugford et al. 2021; Siddharthan et al. 2016). Bayesian reasoning updates probabilities based on prior knowledge and new evidence, and it aligns well with dialogic principles as a pragmatic method for integrating diverse inputs in a structured manner. This statistical method is particularly relevant in scientific tasks that require continuous integration of new evidence, echoing the dialogic idea that knowledge evolves and that understanding is shaped by new evidence (Holquist 2002). By employing a Bayesian framework, we facilitate a methodological integration of diverse contributions for collective decision-making in CS. This approach fosters an environment where multiple perspectives are valued, each contributing uniquely to the consensus-building process.

Drawing on empirical findings from a CS project, we investigate how AI, integrated as a dialogic partner within citizen science processes, enhances learning and decision-making on visual identification tasks. We demonstrate that interactions between humans and AI integrated with a Bayesian consensus model not only improves individual and collective performance but also leads to more accurate outcomes than decision-making involving human participants and AI alone. Through this investigation, we contribute to the field of CS by advancing science practice and learning through a human-AI dialogue, and further quantify the effectiveness of such collaboration for CS practice.

METHODOLOGY

In this section, we describe our collaborative learning environment where an AI system for automated image recognition provides predictions for human users to support the task of species identification from images, and where—crucially—the user can input visual features to constrain AI predictions. We expect that this process of visual collaborative dialogue, which fosters cooperation and exchange of information, can lead to improved species identification accuracy by both AI and human users. Next, we describe the development of the AI system, the interfaces utilised for species identification tasks, the forms of dialogue that users and AI enter into, and who makes the final decision.

AI DEVELOPMENT

We first describe the development of the deep learning models for the automated species classification task.

Dataset

We used the BeeWatch dataset (Van der Wal et al. 2016) for fine-tuning the AI model described. The dataset consists of UK records of 22 bumblebee species submitted with images through an online platform by citizen scientists. Each record was verified with the help of bumblebee experts (Siddharthan et al. 2016). A bumblebee record could include multiple photographs of the species, as different angles help capture the relevant features needed for identification. The dataset consisted of a total of 21,688 images having 24 classes (22 UK bumblebee species and two further classes: not-a-bumblebee and not-identifiable-from-images).

Network architecture and implementation

We used the Inception V3 architecture for model training (Szegedy et al. 2016), which was state of the art when we deployed it in June 2020. To understand how the dynamic of collaboration is impacted by the quality of AI, we analysed results for human-AI collaboration separately for cases in which AI performed well and for cases where it performed poorly.

Specifically, we used the Inception V3 model (Cui et al. 2018) trained on the iNaturalist dataset (Horn et al. 2018), and then used transfer learning to optimise for our task. We used TensorFlow Hub, which provides pre-built AI models (without the top classification layer) to adapt an existing model to our specific needs. This technique, known as transfer learning, allowed us to train this model on the BeeWatch dataset. The open-source Tensorflow was used to train the model (Abadi et al. 2016). A MacBook Pro 2.3 GHz Intel Core i9 processor with a dedicated Radeon Pro Vega 20 GPU was used. We trained the model using the Adam optimizer with a learning rate of 0.001 (Kingma and Ba 2014). A batch size of 200 images was used with image augmentation using random rotation of 40 degrees, horizontal flip, and vertical flip. The data was split into 80% training and 20% validation set at the record level, to avoid correlated images (e.g., different images of the same specimen) crossing training and test sets. A total of 17,371 images were used for training, and 4,317 (19.9%) were used for validation. We report a model accuracy of 51.9% (the percentage of images where the top prediction by the model is correct), and top-3 recall of 69.5% (the percentage of images where the correct answer is in the top 3 predictions). Inception V3 has performed well with accuracies of over 90% on bumblebee species in other datasets (Spiesman et al. 2021). However, it is important to highlight that these datasets are usually not balanced, dominated by a smaller number of species that are easier to distinguish and generally have higher quality images than submitted on BeeWatch, which engages members of the general public rather than naturalists. It is worth noting here that common species are arguably of less interest for biological monitoring, and monitoring rare species is challenging with AI because their majoritarian bias typically results in these being mislabelled as common species (Koch et al. 2022). Our method encourages proper consideration of rarer species and shows the value of collaborative approaches, which could benefit other challenging groups such as fungi.

INTERFACES DESIGN

We compared two different interfaces for species identification tasks, described below. For the control, users performed the identification tasks using an interactive identification key (Figure 1). For the human-AI collaboration interface, AI predictions were integrated into this identification key (Figure 2). The interfaces were co-designed with regular inputs from two bumblebee experts who tested them iteratively to improve the design and workflow.

Interactive key

This interface provides users with an image to identify (*left*), an interactive key (*middle*) and the full list of possible species (*right*), as depicted in Figure 1. Users can examine the visual features of the bumblebee in the image (e.g., sequence of stripes on the thorax) and select matching ones from the interactive key. The system then filters

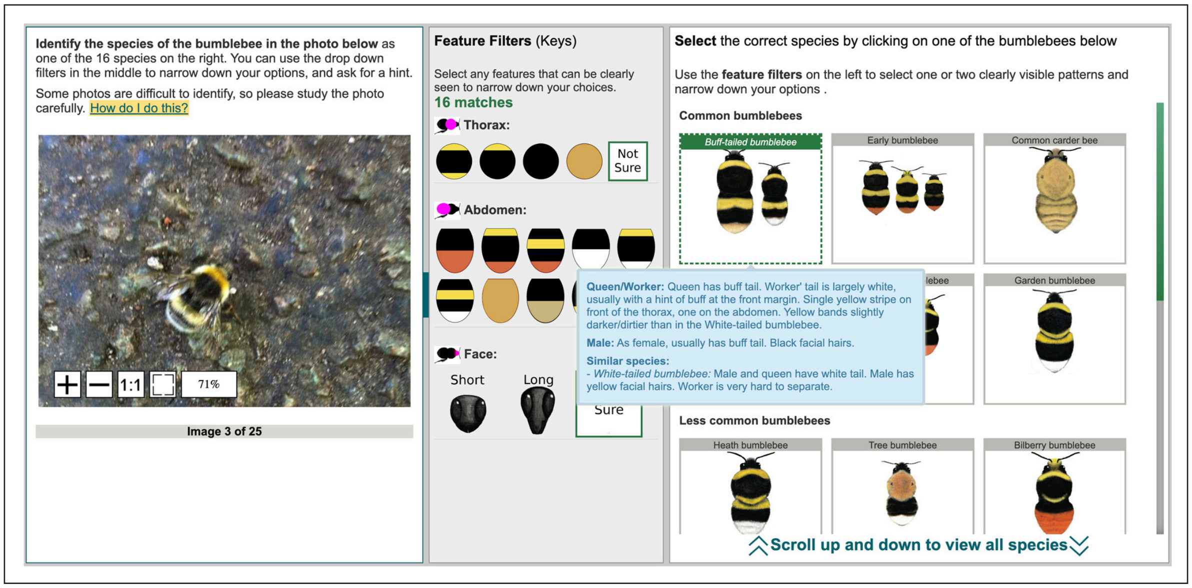

Figure 1 Interactive identification key. Bumblebee image (*left*), feature filters (*middle*), and bumblebee species (*right*). The feature filters can be used to shade out incompatible species. Detailed description of any species can be viewed by moving the mouse over the species.

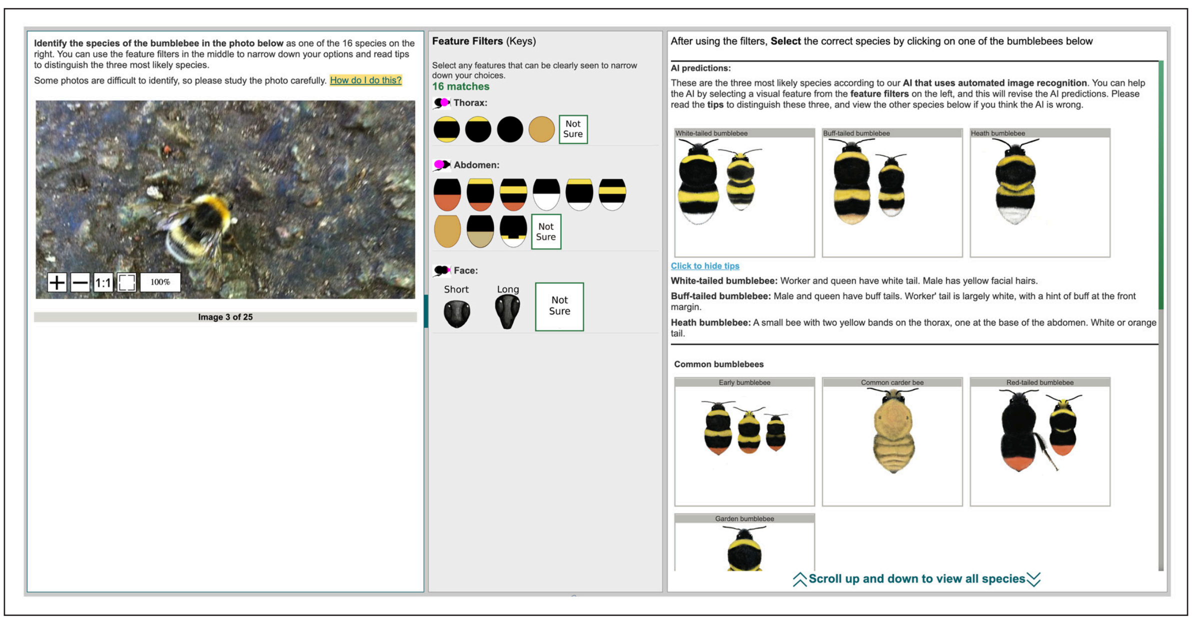

Figure 2 Interactive identification key with AI predictions. This figure shows the Bumblebee photo (*left*), feature filters (*middle*), bumblebee species (*right*), and AI predictions (*top*). The top 3 AI predictions are shown at the top of the species list with clickable tips to distinguish them. The feature filters can be used to change the top 3 AI predictions in addition to shading out the species choices not corresponding with the feature filters.

out (by shading) bumblebee species that do not have those features. Users can modify their choice of filters and mouse-over a bumblebee species to see its detailed visual description and focused tips on distinguishing it from similar species. At the end, they can submit any species from the list, even those greyed out that do not match the selected filters. Similar types of keys are commonly made available to citizen scientists for nature recording purposes, for example, the interactive keys on iSpot, Discover Life, and Butterfly Conservation identification tools for butterflies and moths, and the Royal Society for the Protection of Birds (RSPB) bird identification tool. The interaction using this interface is largely univocal and user driven. It restricts the system's responses to the user's input via a responsive and logical interface, with no external prompts or guidance. This design inherently limits the interface to unidirectional communication, focusing solely on the user's decisions and maintaining a single-voiced interaction throughout the identification process (Wegerif 2008).

Human-AI collaboration interface

We designed the human-AI collaborative interface as a dialogic space to foster real-time dialogue between humans and AI. The interface, depicted in Figure 2, shows the bumblebee images and an interactive key augmented with dynamic AI-generated species predictions. In this interface, the user and AI enter an interactive visual dialogue. AI provides visual predictions for the three most likely species, displayed at the top-right, accompanied by detailed explanations on distinguishing features, as illustrated in Figure 2. These predictions serve as visual prompts that can initiate divergent or convergent viewpoints with respect to the user's beliefs. The user can help AI revise these by inputting visual features using the identification key. Each time the user selects a feature, the interface filters out AI predictions if they do not match the selected features and picks the 3 highest-probability remaining species consistent with the user's feature selections. The features selected can be either similar or different from the ones they can observe from the AI predictions, and hence supports divergent and convergent interactions initiated from the user. However, the updated AI predictions converge towards only the user's feature selection. At the end, the user can choose to accept one of the AI predictions or submit a different species entirely, by scrolling down to select any species, as before. The interaction this interface supports is simultaneous between the participants, with both the AI and the user expressing opinions grounded in prior beliefs. The interface also provides an opportunity for divergence (from the user's perspective) and convergence of opinions through negotiation and self-reflection, and the dialogue modality permits both the user and AI to modify the other's decision. The goal of the interaction is to support a dialogue, and both the user and AI can come to different outcomes post collaboration. This interface thus supports polyphony more completely than the interactive key.

After completing an identification task using either of the interfaces, users were shown a popup (Figure 3) informing them if they were correct, and automatically generated feedback was provided to help improve their identification skills (Van der Wal et al. 2016). With reference to dialogism, this feedback process introduces multi-vocality, whereby the previously hidden voice of the expert is introduced to confirm the correct identification and support learning on the task through feedback.

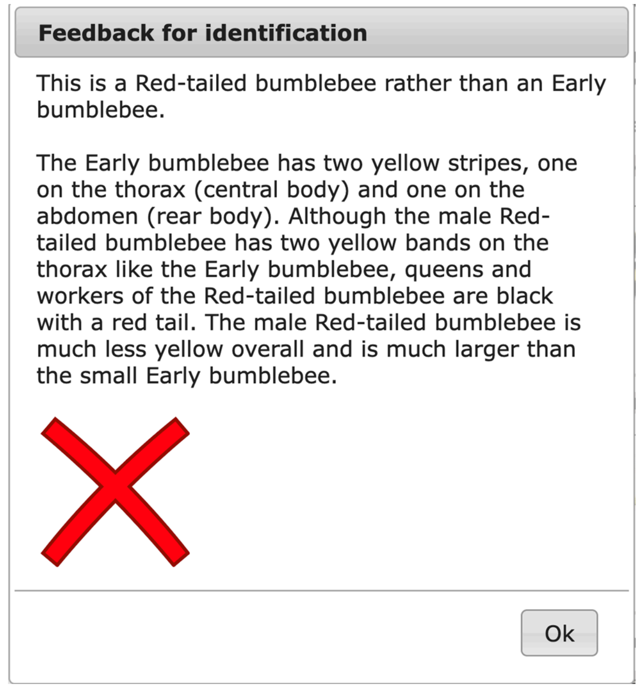

Figure 3 Natural language generated (NLG) texts providing feedback on an incorrect user identification.

DECISION-MAKING

The interface for human-AI collaboration requires the user to submit the final decision on the species. A key issue to consider is whether this leads to the most accurate species identification, for example, when the user and AI arrive at different identifications. We report results for three methods for assigning the final say: (i) Always letting the user have the final say, as already imposed by the interface, (ii) always letting AI have the final say, but taking into account the user's views, and (iii) combining the evidence from the user and AI identifications in a Bayesian framework that also models the biases in the dataset to determine the most likely identification.

User has final say: This is the standard setting on the platform, whereby the user submits the identification after considering the filtering through the keys and the suggestions from the AI.

AI has final say: We designed a method to allow the AI the final decision, after taking into consideration the identification by the user. We implemented rules to identify the top prediction of the AI that is visually compatible with the user's identification as follows:

1. The 22 bumblebee species were manually clustered into 5 distinct sets that shared visual features in the identification key. This approach grouped together visually similar species, and we would expect the correct identification to be within the cluster containing the user's identification.
2. The species-level identification by a user was mapped to one of the five sets.

3. The top 3 species-level AI predictions were also each mapped to one of the 5 sets.
4. The user-mapped set was matched to the top 3 AI-mapped sets in order and if any of them matched; then that AI predicted species-level identification was used as the final identification. If none of the top 3 predicted sets matched, the user's species-level identification was used as the final identification.

Bayesian method to integrate identifications: In this method, we select the species with the highest odds of being the correct identification, given the evidence from identifications by the human user and the AI. Consider Bayes' rule in odds notation:

$$O(H|E_1,\ldots,E_n) = O(H) \times \Lambda(H|E_1) \times \ldots \times \Lambda(H|E_n) \quad (1)$$

$$\text{where } \Lambda(H|E_i) = \frac{P(E_i|H)}{P(E_i|\neg H)} \quad (2)$$

These are the conditional odds O for a hypothesis H, given independent evidence E_1 to E_n. The hypothesis H in this context is a possible species identity. Each evidence E_i comes from an identification by a human or AI. The odds depend on $O(H)$, the prior odds of the hypothesis H (as not all species are equally abundant in our data set, a priori some are more likely than others before we have seen any identifications from human or AI), and Λ terms, each of which updates the existing odds for H based on the incoming evidence E_i. Intuitively, the conditional odds for a hypothesis H increase when the numerator of the term in (2), the likelihood of seeing this evidence E_i for the hypothesis H, is high and the denominator, the likelihood of seeing this evidence E_i for alternative hypotheses, is low.

We estimate the prior odds and the Λ terms from a confusion matrix of species identifications by users and the AI versus the correct identifications, as determined by taxonomic experts within BeeWatch. A Laplace smoothing function is used to add a count of one to each cell of the matrix, to account for previously unseen evidence, and to prevent the denominator in (2) being zero (cf. Siddharthan et al. 2016 for details of method). In this article, we consider two sources of evidence, E_1 and E_2, the identifications provided by a human user and the AI, and select the species H with the highest odds $O(H|E_1, E_2)$.

STUDY DESIGN

We collected data "in the wild" through an online CS platform (https://plantingforpollinators.org/). Participants did not need to provide any information about themselves or their skill levels to participate. Data was collected in each case through usage of a training tool for bumblebee identification that allowed participants to identify a sequence of bumblebee images using one of the interfaces. The training tool for bumblebee identification was initially designed in 2014, using the interactive key interface (Figure 1). Participants could practise their identification skills on initially one dataset, with a second set added in 2019. The interactive key interface was replaced in 2020 with the two-way dialogue interface for human-AI collaboration (Figure 2), though there was a highly visible option for users to turn AI off or back on at any point. For the human-AI collaboration interface, participants were not informed about AI's accuracy during the study to mirror CS practice, in which AI accuracy of a classification may be unknown to the user.

Image datasets

The training tool offered two image datasets for species identification (Table 1). The Easy dataset was created in 2014 with the tool's first version and consisted of a selection of 49 curated images by bumblebee experts.

	EASY	DIFFICULT
DATE CREATED	**05/2014**	**03/2019**
No. of images	49	35
No. of species (max. 22)	12	21
No. of users	849	105
No. of users who used the human-AI collaboration interface	48	62
No. of identifications	13949	1649
No. of identifications with two-way dialogue (human-AI collaboration)	765	931
No. of identifications with AI turned off by user	155	161
No. (and %) of images with AI prediction not in Top 3	8 (16%)	10 (29%)

Table 1 Characteristics of the two datasets resulting from online use of the identification tools.

These images were submitted by citizen scientists during a two-week period within the bumblebee season, predominantly including common species and relatively fewer species overall. This dataset was tailored for training novices on species which they may encounter in their everyday settings. The dataset was available for a longer period and a considerably higher number of users attempted this dataset using the interactive key. The Difficult dataset comprised a careful selection of 35 images to include almost all the bumblebee species (21 out of 22 possible). It was created to provide comprehensive species training and hence also included rare and very rare species. As the name suggests the Difficult dataset was of higher difficulty than the Easy dataset and may be targeted for training proficient users such as naturalists.

Data analysis

By comparing the interfaces, we assessed whether the collaboration of users with AI resulted in increased accuracy. For these comparisons, we analysed data at the level of individual identifications; i.e., where each data point represents a single instance of a user interacting with the interface to identify a species. The dependent variable was user accuracy, modelled as a binary outcome by comparing the user-submitted identification with the expert identification (correct = 1, incorrect = 0). We utilised a generalized linear mixed model with the "glmer" function from the lme4 package (Bates et al. 2015) using R statistical software (v.4.0.4; R Core Team 2021), with Interface type and Dataset as fixed effects and User and Image as random intercepts to account for variability among different users and unique characteristics of each image. We also compared human performance in the study with the accuracy of the AI on that set of photos, and the accuracy of the AI when assisted by the user as described previously.

To understand how the quality of AI predictions affected accuracy, we also separately analysed images in which the correct answer was in the top 3 AI predictions shown to the user and images in which the correct answer was not in the top 3 AI predictions. There were 8 images out of 49 in the Easy dataset and 10 images out of 35 in the Difficult dataset, where the top three AI predictions did not include the correct answer (Table 1).

To assess how the interface and AI prediction quality affected engagement, we analysed the time taken for identification. We focused on 2,892 records from 2019 onwards, when the Difficult dataset was introduced to the platform. After excluding 10 records where users had been inactive for between 5 minutes and 18 hours, indicating they had left the interface to do something else, 2,882 records were considered. A linear regression model was fitted to assess whether dataset type and type of AI assistance (no AI assistance, with AI and correct answer in top 3, with AI and correct answer not in top 3) influenced the time users spent on their tasks.

RESULTS

EFFECT OF INTERFACE AND DATASET TYPE ON ACCURACY (2 INTERFACES, 2 DATASETS)

We found a significant improvement in accuracy when the human-AI collaborative interface was used to identify species (β = 0.486, SE = 0.1191, z = 4.082, $p < 0.001$), showing that the collaboration with AI improved human performance. Additionally, the choice of dataset significantly impacted accuracy outcomes, with the Easy dataset leading to higher accuracy (β = 0.9289, SE = 0.3103, z = 2.994, $p < 0.01$). The mean human accuracy without AI assistance was 64.4% for the Easy dataset and 46.7% for the Difficult dataset (Figure 4). Similarly, automated image recognition was successful in 58.5% of photos for the Easy dataset and 45.4% of photos for the Difficult. The human-AI collaborative interface significantly outperformed the interactive key, which relies on interactions initiated by the user. Overall performance comparison across the two datasets, that is, Easy (49 images, 849 participants) and Difficult (35 images, 105 participants), revealed that the two-way dialogue improved accuracy of both participants (+3.8% Easy; +12.5% Difficult, relative to using the interface without AI) and the AI model (+18.9% Easy; +5% Difficult, compared with the original AI prediction).

It seems reasonable to assume that the collaboration with AI helps users more when the AI predictions are accurate, and users need just choose one of the answers from the top three predictions. For each dataset, we investigated images separately when the correct answer was in the top 3 predictions and when the answer was not in the top 3 predictions.

For the images where the correct answer was among the top 3 AI predictions, the average accuracy for these images improved from 65% (interactive key) to 69% (human-AI collaborative interface) for the Easy dataset and substantially from 50% (interactive key) to 65% (human-AI collaborative interface) for the Difficult dataset (see Table 2).

When the correct answer was not in the original AI top 3 predictions, the collaboration still improved accuracy for the Easy dataset. For this dataset, the humans (like the AI) found these images substantially harder irrespective

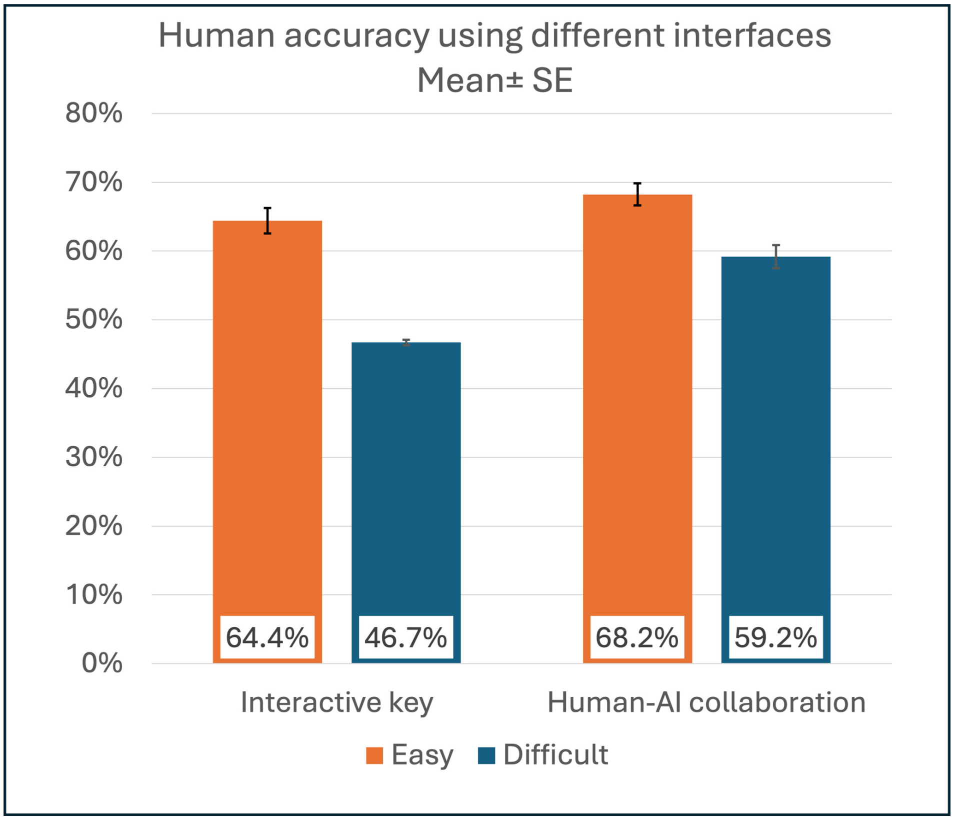

Figure 4 Performance across different image datasets for longer-term use by users. In comparison, the accuracy of automated image recognition over these images was 58.5% for the Easy dataset and 45.4% for the Difficult dataset.

	CORRECT ANSWER IN ORIGINAL AI TOP 3		CORRECT ANSWER NOT IN ORIGINAL AI TOP 3	
	INTERACTIVE KEY	HUMAN-AI COLLABORATION (HUMAN DECISION)	INTERACTIVE KEY	HUMAN-AI COLLABORATION (HUMAN DECISION)
Easy	65%	69%	35%	41%
Difficult	50%	65%	59%	59%

Table 2 Accuracy of the user for images when the correct answer is in the original top 3 predictions of the AI, and when it is not.

of interface (accuracy dropping from 65–59% to 35–41%). For the Difficult dataset, human accuracy was the same irrespective of interface for images when the AI predictions were originally wrong, and interestingly, these accuracies were relatively high even though the AI found these images hard and failed to predict the correct answer in its top 3.

Analysing the time taken to complete the identification revealed that dataset type significantly affected task time ($p < 0.001$), with the Easy dataset (24.6% faster) taking less time than the Difficult. AI assistance did not significantly affect task time, with the users taking 31.25 seconds on average using the interactive key and likewise with the human-AI collaboration interface when the correct answer was not in the top three predictions. When the correct answer was in the top 3 predictions, the average time to completion was 28.63 seconds.

EFFECT OF FINAL DECISION (HUMAN, ARTIFICIAL INTELLIGENCE, BAYESIAN CONSENSUS)

The dialogic nature of the human-AI collaboration through our interface offers the possibility for the user and AI to arrive at different decisions. However, for practical reasons, the expected outcome of an identification task is usually a final decision on one of several possible species. As described previously, this final decision can be determined in three possible ways: the user's decision after utilising AI's input (which is implemented within the interface as a submit button), AI's decision after utilising the user's input, or a Bayesian method, which combines these identifications from AI and the user to arrive at a decision. Figure 5 shows that over the combined dataset, the Bayesian method (results reported over the full dataset using 10-fold cross-

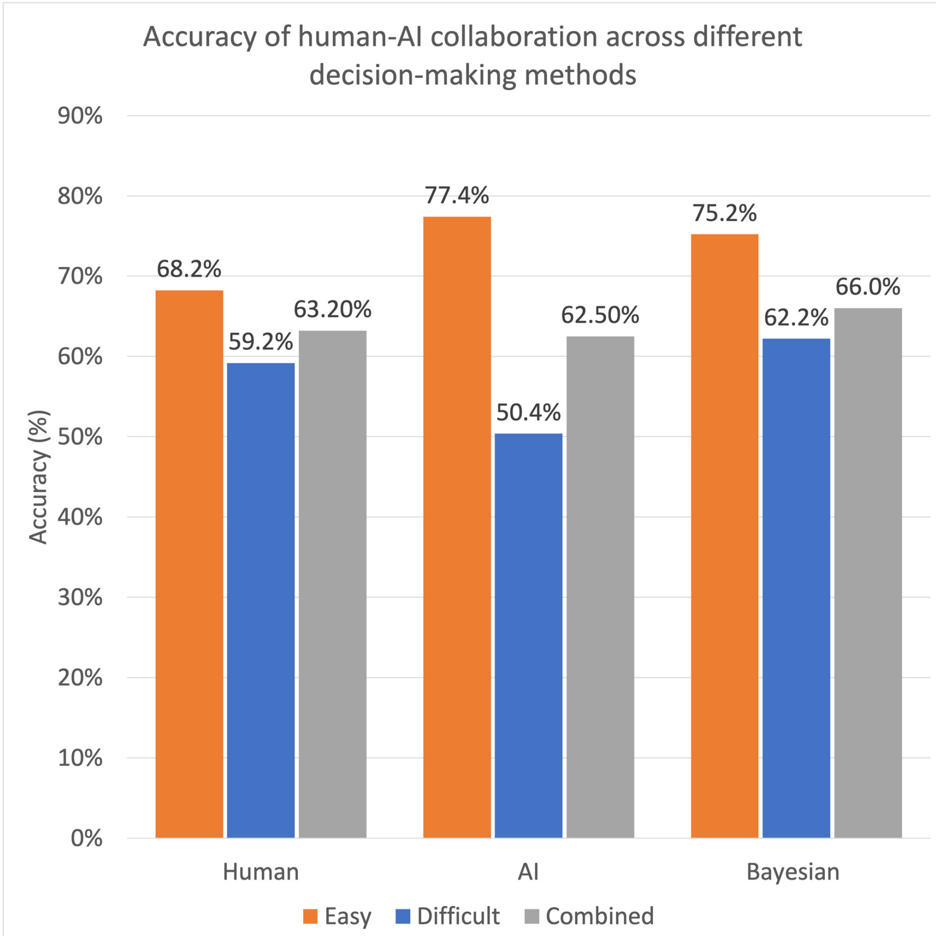

Figure 5 Final decision accuracy comparison after human-AI collaboration. The graph shows the average accuracies for each dataset individually and for a combined dataset, and for situations in which the final decision is taken by a human (taking AI suggestions into consideration), by AI (taking human decisions into consideration), and by the Bayesian method, deriving consensus from these two human and AI identifications.

validation) provides the highest accuracy of 66%. The results indicate that the Bayesian method is indeed useful for building consensus as opposed to delegating the final decision to either a human or AI. This consensus building is thus a continuation of the dialogic process. The Bayesian method has the advantage that we can filter out images for which we have particularly low confidence in identification. For example, if we only accept consensus identifications for which the odds are greater than 1 (indicating the case in which one species is more likely than all the other species together), then we are able to identify 91.8% of the images in the combined dataset with an accuracy of 70.3%.

DISCUSSION

CS processes have emerged as a practice to engage the wider public in scientific activities, both as a means to enhance public understanding but also to foster public participation in issues—such as biological conservation—that are deemed relevant to wider society (Irwin 1995). This reframing of scientific research towards open knowledge-production processes has been crucial in developing participatory models of CS, which recognises various levels of participation, from experts to non-experts, within groups and communities with varying levels of knowledge, and participants possessing a range of skills and familiarity with research (Conrad and Hilchey 2011; Haklay 2013; Shirk et al. 2012). Drawing on a Bakhtinian lens, we have further expanded the idea of participation by proposing a view of CS as a dialogic space inviting contributions from people with different levels of expertise, which may be expressed through a multi-modality of registers and forms. This space typically ranges from monologic interactions, exemplified by crowdsourcing and contributory projects (Haklay 2013), to dialogic interactions as they occur in collaborative projects (Shirk et al. 2012) and community-based monitoring groups (Conrad and Hilchey 2011). Although less common in CS practice, interactions can be further classified as carnivalesque, where traditional boundaries between stakeholders diminish, as seen in Extreme Citizen Science (Haklay 2013), collegial contributions (Shirk et al. 2012), and community led and funded community-based monitoring groups (Conrad and Hilchey 2011).

In this frame, technology has already played and continues to play a central role in facilitating interactions between scientific experts and the broader public, illustrated in large-scale projects such as eBird and Zooniverse for species monitoring and data classification tasks, and on platforms like iSpot and iNaturalist for creating online biological recording communities. However, as AI becomes integral to various facets of society, there emerges a need to further explore its integration within such communities. Here, we suggest that CS, with its diverse applications and participatory models, provides an ideal setting for such explorations. We interpret the interactions between humans and AI in CS as multi-voiced, or polyphonic, showcasing a novel framing of dialogism for CS practice.

Building on this concept, this paper introduces a novel application of dialogism to a visual modality, rather than a linguistic one, expanding the interactive capabilities of AI in enhancing human decision-making processes in CS. We exemplify this by incorporating AI into the species identification process, designing it as a dialogic partner. This AI system prompts users to consider the top three predictions, enabling them to refine their decisions and/or alter the AI predictions based on visual features observed. This interaction occurs through an interface that serves as a dialogic space to maximise the complementary capabilities of humans and AI, that is, human expertise in visual features analysis and AI pattern recognition at the pixel level. We found that such an interactional setting, which supports a dialogue through complementarity, improved identification outcomes for both human participants and AI models across datasets of varying difficulty. The results highlighted that the collaboration benefited both humans and AI, but in different, complementary ways. For the easier dataset with more common species reflecting real-world settings, collaboration allowed AI accuracy to improve considerably whereas human accuracy saw a modest increase. In contrast, for the difficult dataset, AI gains were much smaller but human accuracy increased substantially. Interestingly, users benefitted from the dialogue even when AI did not initially offer any correct identifications. Since AI was introduced on the platform, we found that not many participants turned off the AI predictions and that most identifications were submitted in collaboration with AI (83.2% in the Easy dataset and 85.3% in the Difficult). Additionally, the engagement data showed no significant difference in the time taken across interfaces and regardless of whether the correct answer was among the top three predictions. These results indicate that while the user engagement with the identification process remained similar, the performance improved through collaboration with AI. It appears that when AI is correct, users engage with the interface to validate AI predictions. Conversely, when AI predictions are incorrect, users collaborate to arrive at a decision, which leads to better learning processes in both scenarios. These results highlight a critical role played by AI as a learning partner, helping users validate identifications in some cases and supporting critical reflection in other cases.

However, our implementation also highlighted the challenges of maintaining true multivocality in a CS setting, which often involves maintaining a balance between data quality (independent classifications) and citizen learning. Although the interface allowed for an exchange of perspectives, it guided users towards converging on a single, ideally correct identification. This suggests a tension between the dialogic principle of maintaining multiple voiced interactions and the practical necessity of converging on a decision, an expected scientific outcome (Riesch and Potter 2014). We developed these consensus-building processes using Bayesian modelling to integrate conclusions from humans and AI participants into a combined decision, a technique increasingly utilised across human-AI collaboration research in other domains (Steyvers et al. 2022; Reverberi et al. 2022). This method outperformed the decisions made by either AI or humans alone, demonstrating its effectiveness in unifying diverse perspectives. The Bayesian method brings a novel form of vocality as it expands the dialogic space to include methods and voices beyond the originally intended participants, and generates expertise by highlighting consensus, continuing the dialogic process. Additionally, the Bayesian method provides a framework to include any additional sources of evidence available, such as through additional human participants or different AI predictions. All these are useful outcomes for CS practice. However, a limitation of the Bayesian method is its inherent design to converge on a simplified conclusion, which contrasts with the multivocal nature of dialogism. Despite the convergence, preserving multiple opinions within the dialogic space remains valuable in biological classification processes as it can maintain traceability, transparency, and accountability, while also yielding practical outcomes and new research directions for projects. For instance, on the BeeWatch project platform, images with diverging volunteer classifications (often poor-quality images, rare species, or other pollinators such as solitary bees) have been automatically referred to experts for classification when they fail to meet a consensus threshold. Additionally, this dataset with diversity of volunteer opinions has been useful for exploring new consensus-building models

(Siddharthan et al. 2016) and collaboration techniques between volunteers (Sharma et al. 2022). Finally, keeping dissenting opinions can also be useful for science when revisiting historical data or considering new evidence, for example, when searching for new or non-native species.

The interfaces explored here illustrate how AI can be designed to simulate active participation in CS processes to move away from mostly univocal to multivocal interactions even in visual modalities, leading to improved performance and learning. There are additional roles that AI can play in CS processes involving individual participants, wider community members, multimodal AI systems, and domain experts. We contextualise these roles from the perspective of Legitimate Peripheral Participation (LPP) (Lave and Wanger 1991), in which a novice engages with these processes through legitimate participation such as training or observation, and through continued interaction and learning, gradually moves towards more central roles within the community. During initial engagement and training, AI models and systems can support individual learning by providing tailored feedback on practice classifications, catered specifically to novice participants. AI could take on a "learning partner" role, offering guidance and explanations during the identification process. It could also simulate an expert voice by delivering insights on volunteer classifications through multimodal feedback like highlighting key visual features. Additionally, AI could serve as a facilitator, mediating communication between novices and more experienced volunteers or scientific experts, keeping track of evolving skills. As volunteers develop their skills through practice, there is opportunity for increased engagement in more complex collaborative tasks like analysing real scientific data. In this phase, AI's participation can significantly enhance the data processing workflow. AI could be integrated as an additional participant, making classifications while also facilitating consensus-building by combining different human and AI perspectives using techniques like Bayesian modelling. AI systems as facilitators could streamline communication with experts when conflicts need resolution. With sustained involvement, novices can transition to taking on central expert roles themselves within the community, diminishing traditional hierarchies. Here, AI could continue collaborating through advanced analyses while also strategically identifying areas where human experts or experienced volunteers could provide valuable contributions based on their prior work. Facilitated by AI, these participants could engage in more open-ended, peer-driven dialogues and real-time collaboration, fostering the vibrant exchange of ideas aligned with dialogic principles.

While citizen science projects often explore human-AI collaboration focused on interactive machine learning to improve model performance (Rafner et al. 2022b; Willi et al. 2019), our study underscores the importance of facilitating true two-way interaction between humans and AI for learning tasks like species identification. This dialogic approach not only enhances identification accuracy and consensus-building by integrating human intuition with AI pattern recognition, but also helps address ethical concerns around de-skilling participants. Involving participants in identification tasks, even if they can be easily performed by AI models, helps develop connectedness with nature and provokes interest in socio-ecological issues such as biodiversity loss and climate change, and can lead to attitudinal change and positive action (Sharma et al. 2019; Deguines et al. 2020).

Central to our work is a transformative shift in CS practice, highlighting a dialogic space for human-AI collaboration to enhance scientific outcomes and learning experiences. We achieve this by transcending traditional dichotomies, neither proposing AI as a tool or solution, nor viewing participants as experts or lay. By fostering a dialogic interaction, we aim to create a more inclusive and engaging citizen science participation that has the potential to better address complex environmental challenges.

CONCLUSIONS

We have unfolded how dialogic collaboration can play out between humans and AI around a biological species identification task. We found that a two-way dialogue between AI and the human user improved accuracy over either acting alone. With respect to Bakhtin's dialogism, our developed collaborative space supported a polyphony of distinct voices and prior beliefs, allowing for different species identifications by AI and human participants. We further examined which participant should make the final identification, considering its practical and ethical implications with regards to how new knowledge is validated and by whom. We found that the best results are achieved by deriving post-collaboration consensus from the decisions of the AI and the human using a Bayesian framework, thus highlighting the potential of this approach for designing effective human-AI collaborative spaces guided by dialogism and Bayesian inference. Our application of dialogism to visual interfaces rather than linguistic ones is novel, and our evidence suggests that when used in combination with Bayesian ideas, they offer a framework for human-AI collaboration research in decision-making and scientific inquiry. Their focus on prior beliefs and dialogic consensus-building offers a mixed-method approach to encourage the formation of new communities of practices for participatory learning in CS and conservation practices. While our findings are derived from a specific image identification task and our interface thus implemented visual methods, we suggest that our approach can be applied to wider CS applications to broaden participation and engagement through the integration of explainable AI and interactive, language-based generative AI models. Through our research, we therefore advocate and call for further research that incorporates a multimodal dialogic space design in order to cultivate participatory and inclusive citizen science.

ETHICS AND CONSENT

These studies were approved by The Open University's Human Research Ethics Committee (Application Number HREC/3485/).

ACKNOWLEDGEMENTS

We thank the thousands of volunteers who contributed data via the online platform, whose efforts were indispensable to this research. We also extend our sincere thanks to the bumblebee experts for their critical role in verifying species identifications. We are also grateful to Kevin Crowston and the anonymous referees for their insightful feedback and constructive comments, which have significantly improved the quality of this publication. The research has been supported through two grants from the Engineering and Physical Sciences Research Council [EP/S027513/1 and EP/V042351/1].

COMPETING INTERESTS

The authors have no competing interests to declare.

AUTHOR AFFILIATIONS

Nirwan Sharma orcid.org/0000-0002-6576-3848
The Open University, UK

Laura Colucci-Gray orcid.org/0000-0003-0390-7364
University of Edinburgh, UK

Poppy Lakeman-Fraser orcid.org/0000-0002-1221-0094
Imperial College London, UK

Annie Robinson
University of Aberdeen, UK

Julie Newman orcid.org/0009-0008-4111-7500
St Alban's C of E Primary School, UK

René Van der Wal orcid.org/0000-0002-9175-0266
Swedish University of Agricultural Sciences (SLU), SE

Stefan Rueger orcid.org/0000-0002-6013-9018
The Open University, UK

Advaith Siddharthan orcid.org/0000-0003-0796-8826
The Open University, UK

REFERENCES

Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., et al. (2016). TensorFlow: A system for large-scale machine learning, In: *12th USENIX Symposium on Operating Systems Design and Implementation* (OSDI 16). pp. 265–283.

Ashby, W.R. (1957) *An Introduction to Cybernetics*. London: Chapman & Hall Ltd. DOI: https://doi.org/10.1063/1.3060436

Bakhtin, M. (1984) *Problems of Dostoevsky's Poetics*. C. Emerson, Trans. C. Emerson Eds. University of Minnesota Press: Minneapolis, MN. DOI: https://doi.org/10.5749/j.ctt22727z1

Bates, D., Mächler, M., Bolker, B., and **Walker, S.** (2015) Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), pp. 1–48. DOI: https://doi.org/10.18637/jss.v067.i01

Clark, A., Chalmers, D. (1998) The extended mind. *Analysis* 58, pp. 7–19. DOI: https://doi.org/10.1093/analys/58.1.7

Conrad, C.C. and **Hilchey, K.G.** 2011. A review of citizen science and community-based environmental monitoring: issues and opportunities. *Environmental Monitoring and Assessment*, *176*, pp. 273–291. DOI: https://doi.org/10.1007/s10661-010-1582-5

Cui, Y., Song, Y., Sun, C., Howard, A., Belongie, S. (2018) Large scale fine-grained categorization and domain-specific transfer learning. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* 4109–4118. DOI: https://doi.org/10.1109/CVPR.2018.00432

Deguines, N., Princé, K., Prévot, A.C. and **Fontaine, B.** (2020) Assessing the emergence of pro-biodiversity practices in citizen scientists of a backyard butterfly survey. *Science of the Total Environment*, 716, p.136842. DOI: https://doi.org/10.1016/j.scitotenv.2020.136842

Dellermann, D., Ebel, P., Söllner, M. and **Leimeister, J.M.** (2019) Hybrid Intelligence. *Business & Information Systems Engineering*, 61(5), pp. 637–643. DOI: https://doi.org/10.1007/s12599-019-00595-2

Dillenbourg, P. (1999) What do you mean by collaborative learning? In: *Collaborative-Learning: Cognitive and Computational Approaches*. Elsevier, Oxford, pp. 1–19.

Ellis, R. (2011) Jizz and the joy of pattern recognition: Virtuosity, discipline and the agency of insight in UK naturalists' arts of seeing. *Social studies of science*, 41(6), pp.769–790. DOI: https://doi.org/10.1177/0306312711423432

Engelbart, D.C. 1962. *Augmenting human intellect: A conceptual framework*. Menlo Park, CA 21. DOI: https://doi.org/10.21236/AD0289565

Haklay, M. (2013) Citizen science and volunteered geographic information: Overview and typology of participation. In Sui, D., Elwood, S., and Goodchild, M. (Eds.), *Crowdsourcing Geographic Knowledge* (pp. 105–122). Dordrecht: Springer. DOI: https://doi.org/10.1007/978-94-007-4587-2_7

Holquist, M. (2002) *Dialogism : Bakhtin and His World*, Routledge, London: Routledge. DOI: https://doi.org/10.4324/9780203425855

Hopkins, G.W., and **Freckleton, R.P.** (2002) Declines in the numbers of amateur and professional taxonomists: implications for conservation. *Animal Conservation*, 5, 245–249. DOI: https://doi.org/10.1017/S1367943002002299

Irwin, A. (1995). *Citizen science: A study of people, expertise and sustainable development*. London: Routledge.

Jeong, H., and **Hmelo-Silver, C.E.** (2016) Seven affordances of computer-supported collaborative learning: How to support collaborative learning? How can technologies help?. *Educational Psychologist*, 51(2), pp. 247–265. DOI: https://doi.org/10.1080/00461520.2016.1158654

Kingma, D.P. and **Ba, J.L.** (2014) Adam: A method for stochastic optimization. In *Proceedings of the 3rd International Conference for Learning Representations*, San Diego. DOI: https://doi.org/10.48550/arXiv.1412.6980

Koch, W., Hogeweg, L., Nilsen, E.B., and **Finstad, A.G.** (2022) Maximizing citizen scientists' contribution to automated species recognition. *Scientific Reports*, 12(1), p. 7648. DOI: https://doi.org/10.1038/s41598-022-11257-x

Lave, J. and **Wenger, E.** (1991) *Situated learning: Legitimate peripheral participation.* Cambridge: Cambridge University Press. DOI: https://doi.org/10.1017/CBO9780511815355

Licklider, J.C.R., 1960. Man-computer symbiosis. *IRE transactions on human factors in electronics* 4–11. DOI: https://doi.org/10.1109/THFE2.1960.4503259

Ludvigsen, S., Lund, K., and **Oshima, J.** (2021) A conceptual stance on CSCL history. In Cress, U., Rosé, C., Wise, A.F., & Oshima, J. (Eds.), *International Handbook of Computer-Supported Collaborative Learning* (pp. 45–63). Cham: Springer International Publishing. DOI: https://doi.org/10.1007/978-3-030-65291-3_3

Martineau, M., Conte, D., Raveaux, R., Arnault, I., Munier, D., and **Venturini, G.** (2017) A survey on image-based insect classification. *Pattern Recognition*, 65, pp. 273–284. DOI: https://doi.org/10.1016/j.patcog.2016.12.020

Memmert, L. and **Bittner, E.** (2022) Complex problem solving through human-AI collaboration: Literature review on research contexts. In *Proceedings of the 55th Hawaii International Conference on System Sciences*. DOI: https://doi.org/10.24251/HICSS.2022.046

Mugford, J., Moltchanova, E., Plank, M., Sullivan, J., Byrom, A., and **James, A.** (2021) Citizen science decisions: A Bayesian approach optimises effort. *Ecological Informatics*, 63, 101313. DOI: https://doi.org/10.1016/j.ecoinf.2021.101313

Palmer, M.S., Huebner, S.E., Willi, M., Fortson, L., and **Packer, C.** (2021) Citizen science, computing, and conservation: How can "Crowd AI" change the way we tackle large-scale ecological challenges? *Human Computation*, 8, pp. 54–75. DOI: https://doi.org/10.15346/hc.v8i2.123

Peeters, M.M.M., van Diggelen, J., van den Bosch, K., Bronkhorst, A., Neerincx, M.A., Schraagen, J.M., and **Raaijmakers, S.** (2021) Hybrid collective intelligence in a human–AI society. *AI & Society*, 36, pp. 217–238. DOI: https://doi.org/10.1007/s00146-020-01005-y

Pimm, S.L., Jenkins, C.N., Abell, R., Brooks, T.M., Gittleman, J.L., Joppa, L.N., Raven, P.H., Roberts, C.M., and **Sexton, J.O.** (2014) The biodiversity of species and their rates of extinction, distribution, and protection. *Science*, 344. DOI: https://doi.org/10.1126/science.1246752

Pocock, M.J.O., Roy, H.E., Preston, C.D., and **Roy, D.B.** (2015) The Biological Records Centre: A pioneer of citizen science. *Biological Journal of the Linnean Society*, 115, pp. 475–493. DOI: https://doi.org/10.1111/bij.12548

R Core Team. (2021) *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. http://www.R-project.org/

Rafner, J., Dellermann, D., Hjorth, H.A., Verasztó, D., Kampf, C.E., Mackay, W.E., and **Sherson, J.** (2022a) Deskilling, upskilling, and reskilling: A case for hybrid intelligence. *Morals + Machines*, pp. 25–39. DOI: https://doi.org/10.5771/2747-5174-2021-2-24

Rafner, J., Gajdacz, M., Kragh, G., Hjorth, A., Gander, A., Palfi, B., Berditchevskiaia, A., et al. (2022b) Mapping citizen science through the lens of human-centered AI. *Human Computation*, 9, pp. 66–95. DOI: https://doi.org/10.15346/hc.v9i1.133

Reverberi, C., Rigon, T., Solari, A., Hassan, C., Cherubini, P., Antonelli, G., Awadie, H., et al. (2022) Experimental evidence of effective human–AI collaboration in medical decision-making. *Scientific Reports*, 12(1), pp. 1–10. DOI: https://doi.org/10.1038/s41598-022-18751-2

Rezwana, J. and **Maher, M.L.** (2022) Understanding user perceptions, collaborative experience and user engagement in different human-AI interaction designs for co-creative systems. In *Proceedings of the 14th Conference on Creativity and Cognition* (pp. 38–48). Association for Computing Machinery. DOI: https://doi.org/10.1145/3527927.3532789

Riesch, H. and **Potter, C.** (2014) Citizen science as seen by scientists: Methodological, epistemological and ethical dimensions. *Public Understanding of Science*, 23(1), pp. 107–120. DOI: https://doi.org/10.1177/0963662513497324

Sambasivan, N. and **Veeraraghavan, R.** (2022) The deskilling of domain expertise in AI development. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (Article 587, pp. 1–14). New York, NY: Association for Computing Machinery. DOI: https://doi.org/10.1145/3491102.3517578

Sharma, N., Colucci-Gray, L., van der Wal, R., and **Siddharthan, A.** (2022) Consensus building in on-line citizen science. *Proceedings of the ACM on Human-Computer Interaction*, 6, pp. 1–26. DOI: https://doi.org/10.1145/3555535

Sharma, N., Greaves, S., Siddharthan, A., Anderson, H.B., Robinson, A., Colucci-Gray, L., Wibowo, A.T., et al. (2019) From citizen science to citizen action: Analysing the potential for a digital platform to cultivate attachments to nature. *Journal of Science Communication*, 18. DOI: https://doi.org/10.22323/2.18010207

Shirk, J.L., Ballard, H.L., Wilderman, C.C., Phillips, T., Wiggins, A., Jordan, R., McCallie, E., et al. (2012) Public participation in scientific research: A framework for deliberate design. *Ecology and Society*, 17(2). DOI: https://doi.org/10.5751/ES-04705-170229

Siddharthan, A., Lambin, C., Robinson, A.-M., Sharma, N., Comont, R., O'Mahony, E., Mellish, C., and **Van der Wal, R.** (2016) Crowdsourcing without a crowd. *ACM Transactions on Intelligent Systems and Technology*, 7, pp. 1–20. DOI: https://doi.org/10.1145/2776896

Silvertown, J. (2009). A new dawn for citizen science. *Trends in Ecology & Evolution*, 24, pp. 467–471. DOI: https://doi.org/10.1016/j.tree.2009.03.017

Spiesman, B.J., Gratton, C., Hatfield, R.G., Hsu, W.H., Jepsen, S., McCornack, B., Patel, K., et al. (2021) Assessing the potential for deep learning and computer vision to identify bumble bee species from images. *Scientific Reports*, 11(1), p. 7580. DOI: https://doi.org/10.1038/s41598-021-87210-1

Stahl, G., Cress, U., Ludvigsen, S., and **Law, N.** (2014) Dialogic foundations of CSCL. *International Journal of Computer-Supported Collaborative Learning*, 9(2), pp. 117–125. DOI: https://doi.org/10.1007/s11412-014-9194-7

Steyvers, M., Tejeda, H., Kerrigan, G., and **Smyth, P.** (2022) Bayesian modeling of human–AI complementarity. *Proceedings of the National Academy of Sciences of the United States of America*, 119, e2111547119. DOI: https://doi.org/10.1073/pnas.2111547119

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., and **Wojna, Z.** (2016) Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (pp. 2818–2826). DOI: https://doi.org/10.1109/CVPR.2016.308

Truong, M-X.A. and **Van der Wal, R.** (2024) Exploring the landscape of automated species identification apps: Development, promise, and user appraisal. *BioScience*, 74(9), pp. 601–613. DOI: https://doi.org/10.1093/biosci/biae077

Tejeda, H., Kumar, A., Smyth, P., and **Steyvers, M.** (2022) AI-assisted decision-making: A cognitive modeling approach to infer latent reliance strategies. *Computational Brain & Behavior*, 5, pp. 491–508. DOI: https://doi.org/10.1007/s42113-022-00157-y

Trausan-Matu, S., Wegerif, R., and **Major, L.** (2021) Dialogism. In Cress, U., Rosé, C., Wise, A.F., & Oshima, J. (Eds.), *International Handbook of Computer-Supported Collaborative Learning* (pp. 219–239). Cham: Springer International Publishing. DOI: https://doi.org/10.1007/978-3-030-65291-3_12

Trouille, L., Lintott, C.J., and **Fortson, L.F.** (2019) Citizen science frontiers: Efficiency, engagement, and serendipitous discovery with human–machine systems. *Proceedings of the National Academy of Sciences*, 116(6), pp. 1902–1909. DOI: https://doi.org/10.1073/pnas.1807190116

Van der Wal, R., Sharma, N., Mellish, C., Robinson, A., and **Siddharthan, A.** (2016) The role of automated feedback in training and retaining biological recorders for citizen science. *Conservation Biology*, 30, pp. 550–561. DOI: https://doi.org/10.1111/cobi.12705

Van Horn, G., Mac Aodha, O., Song, Y., Cui, Y., Sun, C., Shepard, A., Adam, H., et al. (2018) The iNaturalist Species Classification and Detection Dataset. In *Proceedings of the IEEE conference on computer vision and pattern recognition* 8769–8778. DOI: https://doi.org/10.1109/CVPR.2018.00914

Wegerif, R. (2008). Dialogic or dialectic? The significance of ontological assumptions in research on educational dialogue. *British Educational Research Journal*, 34, pp. 347–361. DOI: https://doi.org/10.1080/01411920701532228

Wegerif, R. (2024). Afterword: Dialogic space. *Theory Into Practice*, 63(2), pp. 239–250. DOI: https://doi.org/10.1080/00405841.2024.2309840

Wegerif, R. and **Major, L.** (2019) Buber, educational technology, and the expansion of dialogic space. *AI & Society*, 34, pp. 109–119. DOI: https://doi.org/10.1007/s00146-018-0828-6

Willi, M., Pitman, R.T., Cardoso, A.W., Locke, C., Swanson, A., Boyer, A., Veldthuis, M., and **Fortson, L.** (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

TO CITE THIS ARTICLE:
Sharma, N, Colucci-Gray, L, Lakeman-Fraser, P, Robinson, A, Newman, J, Van der Wal, R, Rueger, S and Siddharthan, A. 2024. Image Recognition as a "Dialogic AI Partner" Within Biodiversity Citizen Science—an empirical investigation. *Citizen Science: Theory and Practice,* 9(1): 35, pp. 1–16. DOI: https://doi.org/10.5334/cstp.735

Submitted: 16 February 2024 **Accepted:** 16 September 2024 **Published:** 09 December 2024

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