Transparency in the Wild: Navigating Transparency in a Deployed AI System to Broaden Need-Finding Approaches

Violet Turri*
vmturri@sei.cmu.edu
Carnegie Mellon University
Software Engineering Institute
Pittsburgh, Pennsylvania, USA

Katelyn Morrison*
kcmorris@andrew.cmu.edu
Carnegie Mellon University
Pittsburgh, Pennsylvania, USA

Katherine-Marie Robinson
kmrobinson@sei.cmu.edu
Carnegie Mellon University
Software Engineering Institute
Pittsburgh, Pennsylvania, USA

Collin Abidi
cabidi@sei.cmu.edu
Carnegie Mellon University
Software Engineering Institute
Pittsburgh, Pennsylvania, USA

Adam Perer
adamperer@cmu.edu
Carnegie Mellon University
Pittsburgh, Pennsylvania, USA

Jodi Forlizzi
forlizzi@cs.cmu.edu
Carnegie Mellon University
Pittsburgh, Pennsylvania, USA

Rachel Dzombak
rdzombak@andrew.cmu.edu
Carnegie Mellon University
Pittsburgh, Pennsylvania, USA

ABSTRACT

Transparency is a critical component when building artificial intelligence (AI) decision-support tools, especially for contexts in which AI outputs impact people or policy. Effectively identifying and addressing user transparency needs in practice remains a challenge. While a number of guidelines and processes for identifying transparency needs have emerged, existing methods tend to approach need-finding with a limited focus that centers around a narrow set of stakeholders and transparency techniques. To broaden this perspective, we employ numerous need-finding methods to investigate transparency mechanisms in a widely deployed AI-decision support tool developed by a wildlife conservation non-profit. Throughout our 5-month case study, we conducted need-finding through semi-structured interviews with end-users, analysis of the tool's community forum, experiments with their ML model, and analysis of training documents created by end-users. We also held regular meetings with the tool's product and machine learning teams. By approaching transparency need-finding from a broad lens, we uncover insights into end-users' transparency needs as well as unexpected uses and challenges with current transparency mechanisms. Our study is one of the first to incorporate such diverse perspectives to reveal an unbiased and rich view of transparency needs. Lastly, we offer the FAccT community recommendations on broadening transparency need-finding approaches, contributing to the evolving field of transparency research.

*Both authors contributed equally to this research.

FAccT '24, June 03–06, 2024, Rio de Janeiro, Brazil

ACM ISBN 979-8-4007-0450-5/24/06
https://doi.org/10.1145/3630106.3658985

CCS CONCEPTS

• **Computing methodologies → Computer vision**; • **Human-centered computing → Empirical studies in collaborative and social computing**; • **Applied computing**;

KEYWORDS

Transparency Mechanisms, Need-Finding, AI-Supported Decision-Making, Case Study, Computer Vision

ACM Reference Format:
Violet Turri, Katelyn Morrison, Katherine-Marie Robinson, Collin Abidi, Adam Perer, Jodi Forlizzi, and Rachel Dzombak. 2024. Transparency in the Wild: Navigating Transparency in a Deployed AI System to Broaden Need-Finding Approaches. In *The 2024 ACM Conference on Fairness, Accountability, and Transparency (FAccT '24), June 03–06, 2024, Rio de Janeiro, Brazil.* ACM, New York, NY, USA, 21 pages. https://doi.org/10.1145/3630106.3658985

1 INTRODUCTION

As the field of AI matures, calls for transparency in AI systems from the responsible AI (RAI) research and practitioner community have become more prevalent. To date, research on delivering transparency for end-users has typically focused on explainable AI (XAI) methods. XAI aims to provide explanations for how models arrive at outputs for individual or classes of inputs. While XAI methods can be helpful for tasks such as discovering spurious patterns [60], recent work in XAI literature has shown that XAI methods may not advance model understanding [2], lack actionability [30], or are potentially harmful during high-stakes decisions [29, 62]. Furthermore, recent research within the FAccT community has argued that transparency is much more than explainability [21, 32], or at least the current approaches to explainability [50]. Therefore, we align our work with the perspective from Liao and Vaughan [48], which states: "*Transparency is fundamentally about supporting appropriate human understanding, and this understanding is sought by different stakeholders with different goals in different contexts.*"

While transparency has become an important and debated topic in RAI research, AI practitioners have limited guidance on selecting, implementing, and maintaining transparency mechanisms for real-world systems. In practice, alignment between user needs and computational techniques for generating explanations is not always clear [35, 45, 47]. Although numerous studies suggest how to take a human-centered design approach to transparency need-finding (*e.g.*, [10, 19, 20, 28, 31, 34, 37, 46, 52, 58]), few leverage methods outside of the scope of interviews or user studies of the AI system's end-users. Triangulating varied data sources, however, has been shown to be a crucial aspect of building a comprehensive understanding of a problem space [65].

To explore gaps between transparency research and practice, we conducted an ongoing case study to implement transparency mechanisms for a deployed AI decision-support tool. We partnered with Wild Me, a wildlife conservation non-profit, which offers wildlife researchers AI-enabled decision support tools for identifying individual animals from uploaded images (*e.g.*, [4, 6]). Wild Me has offered a variety of transparency mechanisms within its Wildbook platforms over the past few years (*e.g.*, saliency maps visualizing the regions of the image that contributed the most to the prediction [59]). Other mechanisms outside of conventional XAI include match scores and model documentation. Wild Me introduced these transparency mechanisms to follow RAI practices and to give end-users methods for understanding and collaborating with their models. While the Wild Me team felt these mechanisms had been helpful to end-users, they wanted to reassess how they deliver transparency as they shift from interpretable models to deep learning models. Therefore, our case study goal was to identify the uses and challenges of the current transparency mechanisms and uncover stakeholders' transparency needs. Our need-finding indicates that users have unanswered questions about model behaviors and would like to provide transparency to additional stakeholders. To the best of our knowledge, this case study is the first to retrospectively analyze how transparency mechanisms have held up in a real-world deployment over several years using broader need-finding methods.

Our case study focuses on incorporating various data sources and connecting with a broad ecosystem of stakeholders to support our analysis. Specifically, we expand upon current need-finding approaches by analyzing Wildbook's community forum and user-made training documentation to provide unique insights into transparency challenges and needs. We also conduct in-depth need-finding with the ML and product teams, not just the end-users, and run experiments on Wildbook's MiewID model. By employing these additional methods, we identify tensions and overlaps across stakeholder needs, learn about technical limitations, and evaluate the efficacy of currently deployed transparency mechanisms. We conclude by discussing how insights from our case study can shape new approaches to transparency need-finding. The lessons we learned during this case study are presented to provoke discussion within the FAccT community about how to identify and expand both the problem and solution spaces for AI transparency in decision-support tools. We present additional background in Section 2, an overview of our methods in Section 3, analyses from interviews, experiments, and forum posts in Section 4, and discussion geared towards the FAccT community in Section 5. We summarize our work and contributions in Section 6.

2 BACKGROUND

2.1 Processes for Implementing Transparency

In the last five years, many XAI toolkits (*e.g.*, [1, 3, 7, 8]) have been developed to support AI practitioners in implementing explanations within their systems. However, determining which techniques to use or their alignment with stakeholder needs is an open challenge. Several works aim to address this gap by defining processes for setting XAI requirements and mapping from these needs to specific techniques (*e.g.*, [10, 19, 20, 28, 31, 34, 37, 40, 46, 52, 58, 66]). Approaches used across prior work include conducting interviews with end-users (*e.g.*, [20, 46, 66]), prototyping potential solutions with end-users (*e.g.*, [20, 37, 40, 66]), or presenting scenarios to stakeholders (*e.g.*, [20, 27]). Several processes and taxonomies map transparency needs specifically to XAI methods for individual inferences (*e.g.*, [41, 61]). While these processes and taxonomies are critical steps toward operationalizing transparency in deployed AI systems, they tend to narrow need-finding approaches to focus on end-users and limit solutions to XAI.

2.2 Transparency Case Studies

Several works explore transparency mechanisms in AI decision-support tools for specific domains, such as medicine (*e.g.*, [14, 16, 22, 64, 69]) and wildlife conservation (*e.g.*, [40]). Most similar to our case study context, Cai et al. [14] conduct need-finding with pathologists when collaborating with AI to find similar images, then develop a model and design an interface. Corti et al. [22], Xie et al. [69] and Kim et al. [40] present various explanation prototypes to relevant domain stakeholders to elicit design goals and recommendations for explainability techniques for their specific context. Our work expands upon these contributions by examining transparency mechanisms that have already been deployed instead of focusing on producing prototypes for entirely new mechanisms and/or systems. Furthermore, these works primarily focus on targeting end-users with varying levels of domain and AI expertise. Instead, our work targets an ecosystem of distinct stakeholders, such as end-users and ML/product team members.

2.3 Expanding Definitions of Transparency Solutions

As Suh et al. [63] and Radensky et al. [57] point out, local explanations, or explanations for individual inferences, are not the only way to incorporate transparency into AI decision-support tools. Similarly, Alqaraawi et al. [2] concluded after evaluating the efficacy of saliency maps with end-users that it is necessary to explore transparency at a global level. Model developers implicitly leverage various global transparency methods to understand model performance, such as red-teaming [17] and slice discovery methods [12]. Experiments conducted by model developers can result in information about global model behaviors that can be communicated via methods such as model cards [70]. Similar formal documentation guidelines have been developed for documenting data sources and annotation methods [56]. Suh et al. [63] provides guidelines on how to communicate the capabilities and limitations of a model to subject matter experts. Within the FAccT and CHI communities, researchers are designing and evaluating model auditing techniques that allow everyday users to discover the strengths and weaknesses of machine learning models [12, 36, 38]. In the context of our case study, we draw parallels between Wild Me's community forum and everyday-user model auditing. We also consider the match scores in the system to be a transparency mechanism because they aim to make model behaviors more understandable.

3 CASE STUDY OVERVIEW

We detail methods used during our 5-month case study below. Our methods are designed to understand end-users through a variety of data sources in addition to interviews, such as community forum posts and training documents. Additionally, we met repeatedly with the ML and product teams to capture their experiences with existing transparency mechanisms, ML models, and/or unmet end-user transparency needs. Figure 1 summarizes Wild Me users' workflow using Wildbooks, and Figure 2 summarizes our need-finding methods.

3.1 Wild Me's AI Decision-Support Tool: Wildbooks

Wild Me, founded in 2008, offers AI decision-support platforms called Wildbooks to assist end-users in identifying individual animals from input images. Each Wildbook supports a different species or group of species [4]. Wildbooks employ a pipeline involving object detection and matching algorithms to identify individuals from images. First, uploaded images are sent through an object detection model, which draws bounding boxes around localized animals (A1 in Figure 1). Users may accompany their image upload with metadata such as location, time, and custom keywords (A2 in Figure 1). Users can add or remove bounding boxes manually to ensure their accuracy. Next, users input the annotated images through one or more computer vision algorithms (A1 in Figure 1). Users are then presented with a list of the top-k closest individuals and image matches (part B in Figure 1).

Users can click on each of the suggested IDs within the ranked list to view their uploaded image alongside potential matches. To assist users in confirming matches, users are shown "match scores" for each ranked potential match (B2 in Figure 1) and given the option to view a saliency map of their input image and potential match image (B1 in Figure 1). These saliency maps, presented as heatmaps, highlight areas of importance in the model's matching process. When users want to view the saliency map for a potential match, the saliency maps appear in a new tab window. Users confirm all matches themselves. According to the documentation provided by Wild Me, "the 'score' indicates how similar the uploaded image is to potential matches within the Wildbook database". Match scores are computed differently depending on the model, with models employing a range of metrics such as cosine similarity to identify potential matches from within WildBook's private or public databases.

Wildbooks support several different computer vision models, such as Hotspotter (an interpretable model) [23] and MiewID (a deep learning model) [67]. Users can communicate with other users or the product or ML team about workflows, bugs, and feature requests via their online community forum. Moving forward, Wild Me plans to develop and deploy multi-species MiewID deep learning models for more species and communities. As part of this transition, the Wild Me team is interested in reflecting on existing transparency mechanisms within Wildbooks and considering ways to replicate the interpretability afforded by models like Hotspotter as they rely on more advanced deep learning methods.

3.2 Need-Finding Methods

Meetings with ML Team [MT]. We attended weekly hour-long meetings (a total of nine meetings) with the Wild Me ML team as participant observers to understand the technical underpinnings of the Wild Me models and systems. Meetings were not recorded; however, three members of the author team took notes throughout the meetings. Meeting insights were discussed with the entire author team to derive implications and themes. During the meetings, members of the ML team would walk through experimental results, such as comparisons of a MiewID model's performance (*i.e.*, rank-x and mAP scores) for each species within a multi-species dataset. These meetings were attended sporadically by additional researchers at the intersection of ML and conservation work. Our interactions with the ML team led us to perform our own training and testing of the MiewID model to further explore model behaviors.

Investigation of the Tool's Machine Learning Models [ML]. To build a deeper understanding of Wild Me's deep learning models, we trained and tested a MiewID model for beluga whales[1] on the "Where's Whale-Do?" beluga whale dataset[2]. Our investigation gave us first-hand insight into the training, testing,

[1]We trained the MiewID model using code provided in Wild Me's MiewID GitHub repository: https://github.com/WildMeOrg/wildbook-ia
[2]https://lila.science/datasets/beluga-id-2022/

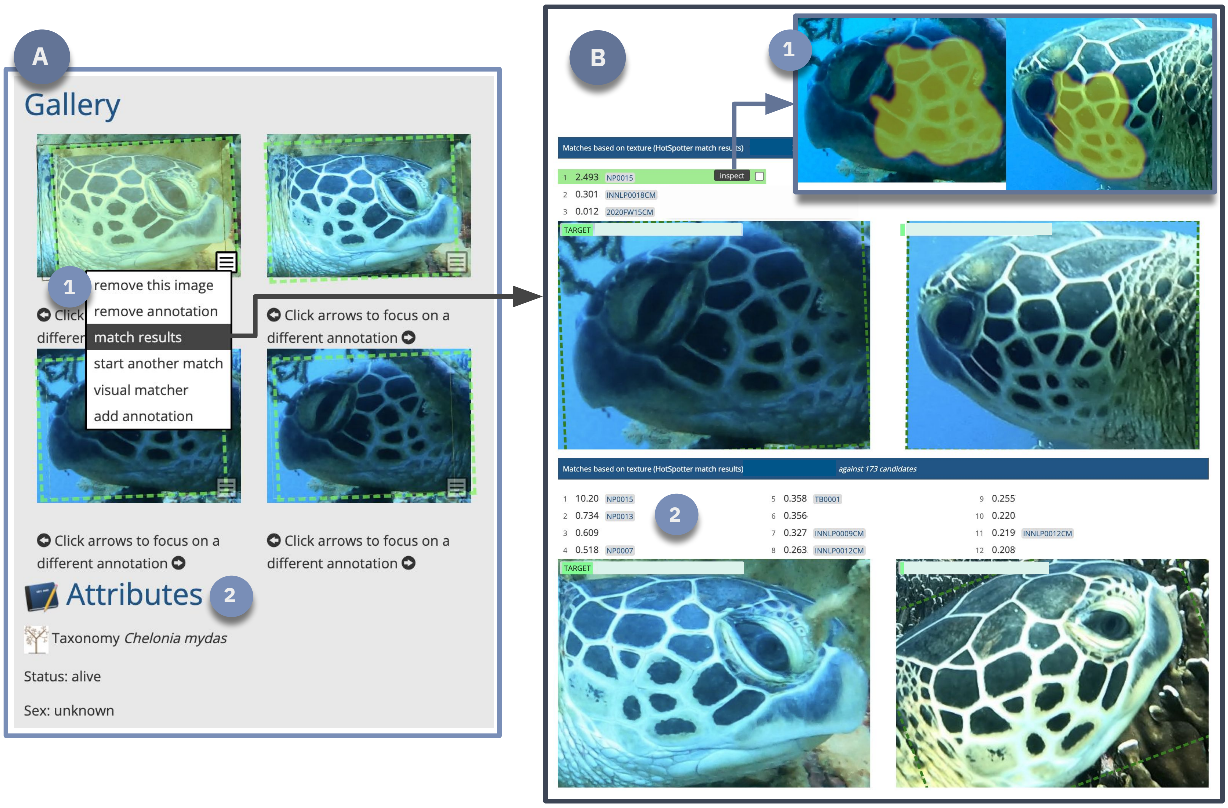

Figure 1: Workflow and features represented from the Internet of Turtles Wildbook. For every encounter ("*a sighting of a single animal at a specific location and time*"), users are shown the image/s and can add metadata, such as the sex. Users can attempt to match the unidentified individual (A1) to another already in the database. Matched results appear in a new tab (B) and will show separate results for each unique image in the encounter and/or each unique model available for that species. (B1) Users can inspect a match result, which reveals a saliency map to rationalize the model's result. (B2) Users can also rely on the similarity "score" to determine the best match result.

and deployment pipeline used by Wild Me, as well as how the current transparency mechanisms are implemented. We chose only to explore the MiewID model because WildMe intends to use this model as their primary model moving forward.

Meetings with the Product Team [PT]. As suggested by Liao et al. [46], we incorporated the designers and product team early on in the need-finding process. To learn about Wild Me's perspectives on end-user behaviors and needs, we held six one-hour-long meetings with the Wild Me product team, which consists of Wild Me's lead product manager, customer experience engineer, and director. Meetings were not recorded; three members of the author team took notes during these meetings. Meeting notes were discussed with the rest of the author team afterwards. At the beginning of our case study, the product team onboarded us on a Wildbook platform and gave us an overview of the system. After our onboarding, these meetings shifted to learning about existing transparency mechanisms and rationalizations behind current design choices. We also asked the product team about their ideas for improving the system based on their knowledge of the users' experiences. Through these meetings, we were informed of training materials developed by Wildbook end-users, which we reviewed as part of our need-finding.

Analysis of Community Forum Posts [CF]. We scraped and analyzed a year's worth of posts (from October 2022 through October 2023) from Wild Me's community forum. The community forum (pictured in Figure 3) is a public online resource for users to post and ask questions. While many topics are discussed on the forum, for the purpose of this study, we limited the scope to posts related to the "Feature Requests" and "Bug Reports" tags to uncover user's transparency needs. After this filter, we scraped 362 posts. We conducted open and axial coding analysis to analyze transparency uses, challenges, and needs [68]. More details about the coding process, including codes and example quotes, can be found in Appendix A.1.

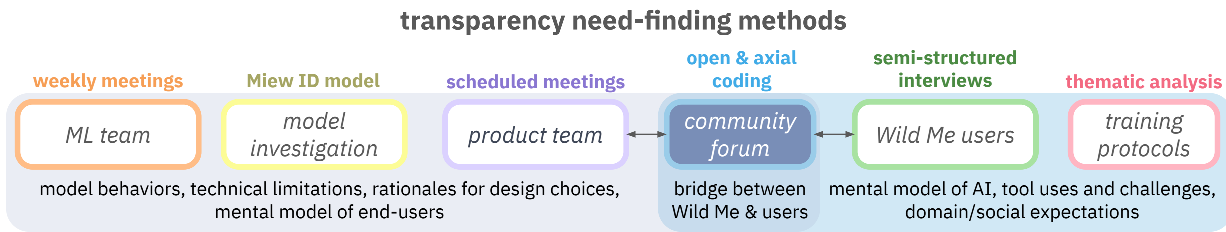

Figure 2: Our need-finding methods consisted of meeting the ML and product teams and experimenting with the MiewID model. These three sources provided information related to model behaviors, technical limitations, and rationales for technical and design choices. We interviewed end-users and analyzed their training protocols to understand their mental models, use cases, and expectations for Wildbooks. Lastly, we analyzed the community forum posts as a window into interactions across and within stakeholder groups.

Interviews with End-Users [EU]. As suggested by previous literature (*e.g.*, [46, 66]), we conducted semi-structured interviews with experienced Wildbook end-users. During the interviews, participants were asked to detail their motivation for using Wildbooks and their experiences and perspectives on current transparency mechanisms. We leveraged a contextual design inquiry approach by having users walk us through their workflow with Wildbook, from collecting data to confirming potential matches [5]. We also asked modified questions from Liao et al. [46] to elicit users' transparency challenges and limitations. We recruited from a list of participants among the top 100 most active profiles on the community forum leaderboard with activity in the past year or among the top active profiles for the past couple of weeks. We also made a community forum post with our recruitment call and sent a recruitment call in the monthly Wild Me newsletter. Participation was restricted to those in the United States who were at least 18 years old. Interviews lasted for one hour over Zoom. Interview transcripts were analyzed through affinity clustering.

Analysis of Training Documents [TD]. Inspired by Cai et al. [15]'s work identifying the onboarding needs of clinicians using AI systems, we analyzed user-made training documents from four different communities. As part of the onboarding process, research teams share training documentation via Google Drive. Wild Me's official training documentation is available on their site[3] and on their YouTube channel [4]. Examples of training materials include video walk-throughs for using the system and written protocols for image collection or matching. By going through the training material, we identified all potentially relevant quotes related to model use and transparency mechanisms. Common themes were identified by clustering the quotes in an affinity diagram.

4 FINDINGS

Our approach to need-finding enabled us to surface nuanced perspectives, opportunities, and limitations related to Wild Me's existing transparency mechanisms. To validate the need to broaden approaches to transparency need-finding, we present how each need-finding method helps us understand transparency solutions within the context of Wildbooks. Instead of highlighting findings from each method separately, we group relevant findings from each source into distinct themes: uses, challenges, and needs. First, we present meta-analyses of the our need-finding methods. Then, we highlight how different stakeholders use existing transparency mechanisms in subsection 4.1. In subsection 4.2, we highlight specific challenges and limitations with the tool's existing transparency mechanisms. Lastly, in subsection 4.3, we connect insights from the need-finding methods to end-user needs not currently addressed by existing transparency mechanisms.

Need-finding Meta-Analyses. We interviewed five end-users: a citizen scientist who uses the Sharkbook, a Marine Mammal researcher who uses the Flukebook, a conservation researcher using the Whiskerbook, a teaching professor using the Internet of Turtles book, and a graduate student using the Giant Sea Bass Wildbook. As seen in Table 1, some users leveraged Wildbooks solely to maintain a catalog of the species. By contrast, others used Wildbooks to understand or influence conservation policies. All of these users have been working with Wildbooks for at least three years and have been manually identifying species in their domain for even longer.

Regarding the community forum posts, we ended up with 71 codes after an open and axial coding of the 362 posts. Most of the forum posts were not related to the transparency of the models but instead were about feature requests (33% of posts) and confusion about the system's design/workflow or how-tos (41% of posts). The remaining posts were about errors (15% of posts), transparency of the system (less than 5% of posts), or miscellaneous topics (6%). The resulting codebook can be found in Appendix A.1.

The four training materials created by end-users were produced for different audiences with a variety of formats. For example, the Internet of Turtles interviewee's training material is geared towards undergraduates collecting field data and uploading images to the Wildbook. This single-document training protocol does not go into detail about the process of matching individuals because undergraduate students are

[3] https://docs.wildme.org/product-docs/en/wildbook/getting-started-with-wildbook/
[4] https://www.youtube.com/@wildme3451

Wildbook	User Role	Years of Use	Use Case
Sharkbook	Citizen Scientist	9 years	Maintain a record of locations that seven-gill sharks visited and how frequently they visited those locations to have
Flukebook	Marine Mammal Researcher	10 years	Create an archival catalog of sperm whales that includes a mapping between photo ID and genetic ID
Giant Sea Bass	Graduate Student	7 years	Working with the local government to produce a population density estimation to help maintain endangered species management
Internet of Turtles	Teaching Professor	4 years	Identify the impact of increased civilization in coastal regions by working with local sea turtle hospitals
Whiskerbook	Conservation Researcher	3 years	Calculating population growth over time to determine the effect of conservation policies on the snow leopard population

Table 1: A summary of the five end-users interviewed, each using the tool for a different animal: sharks (Sharkbook), whales (Flukebook), sea bass (Giant Sea Bass Wildbook), sea turtles (Internet of Turtles), and snow leopards (Whiskerbook). We summarize their main role in their organization and motivation to use the Wildbook.

tasked solely with collecting data for upload. However, the training materials from the Giant Sea Bass Wildbook interviewee, the Whiskerbook Interviewee, and an African Carnivore Wildlife end-user all target users who are collecting field data and identifying individual matches. These detailed training protocols, consisting of numerous documents and videos, include content about interpreting individual match scores and determining which models to use. The training protocols for these three Wildbooks also detail how to annotate images prior to input into a matching algorithm.

4.1 Usage of Existing Transparency Mechanisms

Transparency mechanisms have been integrated within Wildbooks for several years, and saliency maps have been implemented for at least three years. As a result, we were given the unique opportunity to learn how well popular transparency mechanisms have worked in practice. Our findings indicate that both mechanisms (saliency maps and match scores) can provide value to end-users and the Wild Me team. These mechanisms are informative to developers and help the product team support end-users' transparency needs. We detail specific uses observed using each need-finding method for both transparency mechanisms below; Table 2 provides a high-level summary of uses.

4.1.1 *Saliency Map Uses.*

ML Team [MT], [CF]. Our discussions with the ML team taught us that saliency maps are an essential part of their model training and evaluation workflow. The ML team uses saliency maps to help convey model performance issues during their meetings. For example, during one session, one of the ML engineers walked through a graphic depicting examples of correct and incorrect model predictions with their accompanying saliency maps to communicate aspects of a model's performance to the rest of the ML team. During another meeting, an ML engineer showed saliency maps with ambiguous trends from one of the newer models still under development to discuss how to better explain that model's predictions. ML engineers also leveraged saliency maps within the community forum to address a user's comment about errors in image segmentation. Overall, the ML engineers rely on the saliency maps as a diagnostic tool for understanding model behavior and as a way to communicate these behaviors to others.

Product Team [PT]. According to the lead product manager and customer experience engineer, saliency maps are used by their team and within their products to build trust among end-users. The product team also discussed prior user research involving saliency maps which indicated that maps are important for cultivating trust and adoption. The team also shared Figma prototypes of new designs for Wildbook's interface which indicate that saliency maps will be used moving forward.

End-users [EU], [TD]. Through interviews with end-users, we learned that end-users view the system as a way to narrow the search space for potential matches. However, users still manually match individuals by visually analyzing their unique patterns, markings, or spots (Figure 1, part B). Matching methods vary by species and research team based on pre-AI workflows for identifying individuals. Many users identify and validate key points across images without using saliency maps at all, instead zooming in on the image for closer manual inspection. For example, the Internet of Turtles interviewee said, "*Sometimes I will zoom in on the picture...so I haven't used the [saliency maps] that much.*" While many users are inclined to match individuals manually, some users, such as the WhiskerBook interviewee, leverage saliency maps as a starting point for identifying shared features across individuals. Figure 1 highlights how the saliency maps from two images cover the same scales on a turtle, which the user can then use to confirm the match. Conversely, the Giant Sea Bass Wildbook interview uses platform-specific visualizations that highlight specific points instead of regions.

While end-users may not rely heavily on saliency maps during the matching process, maps can be used to communicate information about how Wildbooks work to external stakeholders and newer users. For example, the Whiskerbook and Internet of Turtles interviewees have both leveraged saliency maps when sharing results with outside stakeholders (*e.g.*, local policymakers, government officials, and research audiences) to justify their use of the system and the validity of their research. Our Internet of Turtles interviewee also used saliency maps as part of the onboarding process to demonstrate how the models work.

4.1.2 Match Score Uses.

ML Team [MT]. Match scores derived from Wildbook models are used to determine the top-k closest matches to an input image. While match scores could hypothetically be analyzed during the training process as an indicator of model calibration, our meetings with the ML team indicated that scores are not consulted within their current workflow.

Product Team [PT]. While the lead product manager and customer experience engineer feel that match scores can be confusing for end-users to interpret, the product team also shared that many end-users develop their own intuition for interpreting match scores to assist in their matching process. The product team has considered removing the match scores to encourage users to rely solely on the rankings instead of the score. However, they've opted to keep match scores because they have learned from prior user research that scores have been helpful to many power users.

End-users [EU], [CF], [TD]. Many experienced users have developed their own systems for interpreting match scores. For example, when the Internet of Turtles interviewee "*...sees a [match score] above 1...there's a pretty good likelihood...that's a match.*" Alternatively, if the interviewee sees a score below 1: "*I'll look at it the [potential match], but I know that there's not really anything there.*" Across training materials produced by end-users, we saw that higher match scores were described as indicating a *higher likelihood* of a match. In the training material for the Giant Sea Bass Wildbook, for example, users are advised that when scores are "*...greater than 200 [it's] a very high probability of a match... 2 is the threshold for potential matches.*" Similarly, the training video from the Whiskerbook Interviewee discusses the match score as "*...the ranking of how well the images matched to each other.*" Like saliency maps, match scores are just one factor in navigating potential matches, as all interviewees stated they still manually match the result just to be sure. For example, the Internet of Turtles interviewee described the match scores as "*...a very generic starting point...*". Questions about match scores on the community forum further corroborate that they are used to some extent during the matching process.

4.2 Challenges with the Existing Transparency Mechanisms

Aside from use cases, we also learned about current challenges related to transparency mechanisms. Due to model and system constraints, these challenges are often connected to design and technical choices made by the Wild Me team. We saw that transparency mechanisms can unintentionally introduce confusion to end-users within the matching process. We detail specific challenges observed from each stakeholder for both transparency mechanisms below; Table 2 provides a summary of challenges.

4.2.1 Challenges with Saliency Maps.

ML Team [MT], [ML]. We learned from the ML team that while saliency maps for interpretable models produce visualizations that highlight matching features across the input and potential match images, saliency maps for their deep learning models can produce unusual gridlike patterns that are hard to understand. During our own experiments with the MiewID model, we struggled to produce reasonable saliency maps using a different visualization approach. As a result of the misalignment between saliency methods and the MiewID model, saliency maps have become less usable over time in this context. Additionally, the ML team voiced that generating saliency maps can be computationally inefficient, which only adds to the already long processing time needed to identify matches. These challenges suggest that alternative transparency solutions to saliency maps may need to be explored moving forward.

Through our experience using the MiewID code, we learned that the current system utilizes GradCAM++ to generate saliency maps [18]. Discussions with the ML team revealed that the team found this visualization method to be one of the faster techniques. However, challenges in applying traditional saliency map techniques may arise as the Wildbook models are for information retrieval rather than classification which means these models produced embeddings rather than classification scores.

Product Teams [PT], [CF]. The product team faces challenges in informing end-users of the saliency map capability. While saliency maps are included in a number of demos and across documentation related to Wildbooks, the product team has seen on the community forum and within their user research that end-users are often unaware of how to access these features. Further, the tab for accessing the saliency map is labeled the "inspect tool", which can lead to confusion about what the tool is and the information it provides. Discussions on the community forum between the customer experience engineer and end-users revealed that saliency maps are discarded 1-2 weeks after user uploads are matched, at which point the button for inspecting the saliency map disappears. This decision was made as a result of technical constraints related to storage. The product team has had to field numerous questions and complaints related to the unavailability of saliency maps in the community forum.

Transparency Mechanism	**Uses** (Section 4.1)	**Challenges** (Section 4.2)
Saliency Maps	• [MT] Identify model behaviors, such as blindspots or spurious patterns • [MT], [CF] Communicate unexpected model behaviors • [PT] Incorporates saliency maps in the tool to help build end-users' trust • [EU], [TD] Consult while confirming a match as a second source • [EU], [TD] Communicate with external stakeholders and new users	• [ML], [MT] Saliency maps for some models are ambiguous or uninterpretable • [MT] Need to prioritize computationally efficient saliency maps • [EU] The discoverability of the saliency maps feature is low • [EU] Need more clarity on the goal of saliency maps • [EU], [CF], [PT] Need more guidance on where and when to find the saliency maps • [EU], [CF], [TD] Saliency maps can be wrong or convincingly misleading
Match Scores	• [PT] Incorporates match scores in the tool for power users to consult while reviewing match results • [EU], [CF], [TD] Consult to reason about the recommended matches "confidence", "accuracy", or "quality"	• [MT], [ML] Score ranges and meanings not standardized across models • [PT] Presenting how the match scores in a user-friendly way • [CF] Unclear on how to interpret and compare match scores

Table 2: Wild Me's tool currently offers two transparency mechanisms: saliency maps and match scores. This table provides a link between the need-finding methods and the key uses of and challenges with the transparency mechanisms. Insights are mapped to the community forum analyses [CF], end-users' training documents [TD], end-user interviews [EU], meetings with the product team [PT], meetings with the ML team [MT], and model experiments [ML].

End-Users [EU], [CF], [TD]. While the Whiskerbook and Internet of Turtles interviewees were aware of and used the saliency maps feature before, community forum posts and discussions with the Flukebook and Sharkbook interviewees reveal that many users do not share the same experiences. Many users are unaware of the existence of saliency maps, struggle to access them, and/or rely on alternate visualizations. For instance, the Sharkbook interviewee was unaware of the feature and uncertain of its meaning after being shown the feature during an interview. Examples of platform-specific alternatives to saliency maps used by interviewees include "trailing edge" visualizations for whale flukes and "spot-based" visualizations for giant sea bass.

Users who are aware of saliency maps mentioned pain points related to their use including the expiration of saliency maps. To quote a post from a community forum user, "*I'd like to see the inspection results, but I'm getting this error 'inspection image unavailable (likely outdated)' on every attempt.*" Saliency maps can also inaccurately highlight mismatched elements across images, as warned by the WhiskerBook training materials, making saliency maps difficult to use or potentially misleading during manual matching. Per the Whiskerbook interviewee's experience: "*Sometimes it is wrong, and it's not useful.. sometimes the match is not correct, and it will match on a pattern that is not the same.*"

4.2.2 Challenges with Match Scores.

ML Team [MT], [ML]. According to the ML team, the meaning of match scores varies across algorithms. For example, the MiewID algorithm uses cosine distances between embeddings to identify the top-k matches. The scores shown to end-users are the cosine distances between the target embeddings and (pre-computed) match candidate embeddings. Through our close investigation of the MiewID model, we learned that the embeddings are generated by training an EfficientNet model [42] with an ElasticArcFace module. During evaluation, the ElasticArcFace module is removed, and embeddings are extracted from the final layer of the EfficientNet. This differs from the Hotspotter algorithm, which uses a variation on Local Naive Bayes Nearest Neighbor (LNBNN) to calculate potential matches. The ranges of match scores and the information they communicate will therefore vary across algorithms and trained model instances.

Product Teams [PT], [CF]. The product team frequently responds to questions from end-users about how to interpret match scores. Similar questions were seen in the community forum, where users often ask what match scores represent. Because match scores can mean very different things, conveying their meaning to end-users can be challenging. The product team responds to user questions about match scores on the community forum with responses akin to, "*All three algorithms have relatively inscrutable scores. They're not probabilities and not especially meaningful but are rather distance metrics of one form or another and will vary from the trained instance of the algorithm to another trained instance.*" Another definition in a similar community forum post reads: "*The more visual texture similarity, the higher the score.*" Despite the product team adding a statement about this caveat in their own training documentation, they continue to encounter questions about match scores on the forum.

End-Users [CF]. Based on our analysis of the community forum, end-users are hesitant about interpreting match scores. One user on the community forum asked, "*Regarding 'Match score' in Wild Me Documentation, it is stated that, 'The match score represents the numeric value returned from the algorithm.' But what exactly are these numeric values? What do they represent: percentage, a score from 0 to 1, or 0 to 10? Or something else?*" Many end-users incorrectly interpret the match scores as probabilities. On the community forum, users ask questions such as "*Is the highest score 1?" and "The CurvRank v2 has 2 results [with a score] around 5. If the highest rank [score is] 10, [do] the results have a confidence of 50 percent?*" and "*FinFindR has 2 results around 0.2. What is the maximum score? What is the confidence of the matching [Whale X]?*" Ultimately, users struggle to interpret match scores and are unable to compare scores across algorithms since these scores may fall within different ranges.

4.3 Broader End-User Needs

Our analyses of end-user interviews, community forum posts, and training documents helped us form an understanding of end-users' mental models of AI and their workflows throughout the entire Wildbook pipeline. Talking to the ML and product team helped us situate those needs within the limitations of the system design and models. Based on our analyses, we learned that while current transparency mechanisms are targeted toward the match exploration stage, users have transparency needs across stages of the pipeline leading up to and following the match process. This is also seen in related works such as Corti et al. [22].

4.3.1 Distribution of Acceptable Model Inputs [PT], [MT], [CF], [TD]. We identified unmet transparency needs related to the image annotation process throughout our data sources. Before users can view potential match results, they must oversee the upload and annotation process. While the annotation process is automated using object detection models for some Wildbooks, users sometimes need to manually add or correct annotations due to model failures. During the annotation process, users also have to assign a viewpoint of the animal to the image (*e.g.*, top, left).

Discussions in the community forum reveal that users grapple with uncertainties about which attributes of model inputs are likely to cause failures during the annotation process. Numerous users have expressed frustration after trying to bulk upload a set of images only to learn that the type, quality, and/or size of their images were not accepted by the system. For example, one community forum post discusses the unexpected importance of crop ratio, "*The Hotspotter algorithm does not seem to be able to match against images other than with the same crop aspect ratio. So if there are no images with the exact same crop aspect ratio, there are no matches.*" Similarly, discussions on the community forum suggest that researchers expect viewpoints to be heavily factored into the AI-enabled matching process, as the viewpoint is critical to their manual matching processes. Conversely, the Whiskerbook training documents indicate that users believe the algorithm matches against all images, not just those with the same labeled viewpoints, so annotation issues will not impact matching. Interestingly, concerns about model inputs and annotation failures were not brought up during the interviews with end-users, although they were prevalent on the community forum.

The process of curating acceptable model inputs is further complicated by end-users' pre-AI notions of image quality. Through the product team, we learned that wildlife conservation researchers grade the quality of their images for individual identification using image quality and individual distinctiveness scores [33]. According to the ML team, some users include these scores within the metadata for their images, but their systems for rating images vary widely across users and domains. The concept of distinctiveness scores may influence end-users' mental model of high- versus low-quality images and may lead to misaligned definitions of image quality between the end-user and system specifications. End-users are left to make informed guesses about how their annotations impact their matches.

4.3.2 Explore and Compare Algorithmic Performance [EU], [TD]. Performance metrics for Wildbook models are not generally available to end-users. As a result, users are left to fill in the blanks on model performance through their own analyses and/or develop protocols about which algorithms to use based on their experiences. For example, a group of Whiskerbook users conducted quantitative analyses of model performance for different k-values and published these findings for the snow leopard research community [9]. This information can then be used to inform user protocols for reviewing potential matches. The Whiskerbook training material, for example, advises Whiskerbook users to consult multiple models. By contrast, the Flukebook interviewee consults Wildbook models consecutively, analyzing results from the first model and looking at the second model only if a match is not already found. Alternatively, the Giant Sea Bass Wildbook interviewee has determined differences between the two models from "trial and error": "*...I3S [one model] I feel like is more precise, but they give less options...maybe there's a potential chance for like 5% error where Groth [another model] I feel like is probably like a*

97% chance of having it within the 20 matches." These findings showcase how users have compensated for not being provided information about model performance (*i.e.*, through trial and error or through their own comparison experiments). Users are willing to go to great lengths to conduct these analyses because they are essential to their work.

4.3.3 Human-Human-AI Collaboration Workflows [PT], [CF]. Although experienced users typically review and finalize matches on their own, many research teams employ an offline system for validating matches made by newer users or handling challenging matches. Some Wildbooks, such as the Giant Sea Bass Wildbook and Sharkbook, invite the community to take part in uploading their own images and suggest potential matches for the experts to approve. These human-human-AI collaboration workflows, in which a user reviews AI suggestions and collaborates with other users to make a final decision, often take place behind the scenes of the system's workflow. This process aims to maintain the quality of the animal ID database and prevent errors from propagating, such as an individual animal incorrectly being identified as two separate entities throughout the database. According to the product team and "Feature Requests" on the community forum, users have asked for the Wildbooks to offer features that accommodate their collaborations. On the community forum, users have requested a new kind of user role with limited access privileges "*...allowing them to record an encounter, consult the encounters, launch a match, and see the results but not to validate this match or to assign an individual to a photo.*" To finalize the match, "*...[we] would then let only selected and more experienced users perform this critical last assignment step.*" Alternatively, users on the community forum have also asked for "peer review" functionality, in which case "*matches validated by the contributors are in an 'unapproved' state by default, and there is then a validation step when someone to whom we have assigned this role can edit it in an 'approved' state.*" While the interviews with end-users indicated that users participate in human-human-AI collaboration for difficult matches, they did not reveal details to the extent found on the community forum.

5 REFLECTIONS & INSIGHTS ON NAVIGATING TRANSPARENCY IN DEPLOYED AI SYSTEMS

Throughout our 5-month case study with Wild Me, we utilized diverse data sources to paint a varied picture of current transparency mechanisms. We uncovered from interviews with end-users and analyzing user-created training materials that user workflows often involve manual review of AI outputs and complex review structures. We learned that the ML team uses saliency maps to investigate model behaviors and communicate those behaviors to others. However, through our experiments with MiewID, we saw first-hand how these maps can become challenging for deep learning models. By contrast, the product team uses existing mechanisms to build trust among end-users, although analyzing the community forum posts expose how users can also struggle to interpret or find these mechanisms.

Our methods enabled us to identify tensions between stakeholders, which we can mitigate by prioritizing functional needs and finding dual-use solutions. For example, the ML team's challenge of providing computational efficiency ultimately needed to take precedence over other stakeholder concerns and constrained the solution space. However, investing in methods for interpreting match scores, for instance, could have positive implications for the ML developer and end-user workflows while remaining computationally inexpensive.

While interpreting our reflections, it should be taken into consideration that our research setting may be unique in that we have direct access to a variety of stakeholders and data sources, such as the community forum. Additionally, we are working with a deep learning model that performs an information retrieval task, and the nature of the system and user workflows may be somewhat specific to our case study. However, we do our best to provide generalizable reflections for the broader FAccT community without overextending their implications. We situate our findings within related literature and the FAccT community by reflecting on our approach below.

Need-finding with the ML team and the deployed models provides necessary insight into technical limitations and challenges. While previous transparency need-finding frameworks have included model developers in their process (*e.g.*, [46]), the model developers are leveraged primarily to implement transparency solutions. In our work we positioned technical considerations to be front and center by considering the ML teams' perspectives to be on the same level as end-users. Being hands-on with the machine learning model used within the system, as opposed to only hearing from the ML teams about how the system works, exposed us to additional model behaviors and technical limitations related to transparency that we would not have otherwise discovered. For example, using ML models for information retrieval within the system means that saliency maps need to be generated in a way that differs from classification contexts. Overall, our meetings with the ML team taught us about model behaviors that end-users may be unaware of.

While this method exposes valuable and relevant insights, gaining access to the ML team and models may not always be possible, given the researchers' networks or data-sharing policies. Additionally, the insights that arise can mislead researchers if this method is prioritized more than an end-user-centric method. We encourage transparency researchers to conduct need-finding closely with the ML team and the deployed models early on in the transparency need-finding process to gather the necessary information to fuel transparency solutions, such as behavior descriptions of the model [13] and onboarding modules [16]. Ultimately, some form of internal support from the ML or development team should be a requirement for deploying transparency mechanisms.

Community forums for AI tools can expose valuable crowdsourced information about transparency needs and challenges. Although analysis of large-scale crowdsourced data can be time-consuming, and a small portion of the data may be specific to transparency, we found the community forum extremely valuable in identifying needs across the wider user base. Previous work has also found end-users to share model blindspots and failures they encounter online, such as on Twitter [44]. The value of leveraging end-users' crowd-sourced insights for improving model transparency and fairness has been emphasized by several recent workshops and studies (*e.g.*, [11, 25, 26, 43]). As shown by our findings and previous transparency research, analyzing content from end-users beyond semi-structured interviews proves vital to identifying and understanding transparency needs. As such, we recommend that AI practitioners consider building community forums for their users. We also recommend that transparency researchers value insights from community platforms or other social media platforms as these can be useful for triangulating user data [65].

Exploring how end-users compensate for unmet transparency needs helps identify potential transparency solutions. A vital step in the need-finding process is the observation stage, allowing researchers to understand users' current workflows and identify their needs without explicitly asking users what those needs are. Very few existing processes for transparency need-finding suggest leveraging observational or ethnographic studies to understand users' current workflow with the AI system [10]. Furthermore, none has suggested analyzing community forum posts or user-designed training documentation. While interviewing end-users, we discovered that many users were creating their own training documentation on how to collaborate with the system and interpret the predicted match results. Cai et al. [15] observed clinicians having a similar need for users to be onboarded to the AI system, corroborating our findings. We also learned that users rely on the community forum as a way to expose the transparency of the models they work with. From understanding model blindspots in the community forum to comparing the models in training documentation, end-users have created numerous means to compensate for unmet transparency needs. Outside of the matching process, users have questions about model failure modes, image quality requirements, and algorithmic performance. The information end-users are seeking by their own means can be addressed by building on ideas suggested from previous FAccT contributions, such as model cards [24, 51] and data cards [56]. Furthermore, a user study with non-expert analysts highlights the significant value added by providing an interactive tool that offers information about a model's uses and blindspots [24]. While we identified users' needs to align with various transparency solutions, such as model cards (*e.g.*, [51]) or onboarding methods (*e.g.*, [16, 53, 54]), many transparency taxonomies fall short of suggesting solutions beyond local explanations [61]. We encourage the FAccT community to consider incorporating contextual inquiry-inspired approaches within transparency need-finding to broaden the transparency solution space. Ultimately, incorporating contextual inquiry or observational studies in the transparency need-finding process can help researchers and practitioners uncover how end-users compensate for currently unmet transparency needs. In the absence of training materials, researchers may benefit from working with end-users and the developer team to create documentation to see how transparency needs or uses arise.

Transparency solutions must be adaptable to complex human-human-AI collaboration workflows. As revealed by many of our interviewees, end-users often work within a larger team with members with varied levels of experience matching species within their domain. End-users may hand more challenging matches to senior members of their team for review. This is not the only place where human-human-AI collaboration workflows exist. For example,radiology students and residents may collaborate with AI to help them localize and/or grade tumors and then pass the patient case to their attending for validation or further investigation [39]. With the exception of works such as McNeese et al. [49] and Munyaka et al. [55], few papers investigating human-AI teaming explicitly acknowledge human-human-AI collaborations. To our knowledge, no previous works recognize the need to adapt transparency mechanisms to such workflows. From our analysis of the community forum and discussion with the product team, we identified the need for transparency solutions to be adaptable to complex human-human-AI collaboration workflows within Wildbooks. Future research is needed to explore transparency mechanisms in the context of human-human-AI collaboration scenarios. One such solution could be to consider collaborative saliency maps, where users can annotate regions of the map they believe are convincing or misleading for their human collaborators to review. Another solution could be to consider offering users the option to explicitly identify which model they used to identify a match and a short rationalization for their decision. In terms of broadening transparency need-finding approaches, we recommend that researchers identify not only the end-users' workflow but also the processes that exist before and after the end-user interacts with the data or model.

6 CONCLUSION

Current transparency need-finding approaches often focus primarily on end-users and use this information to determine XAI solutions. Although conversations with end-users are an essential component of understanding user needs, users may have challenges that are missed or overlooked without a broader context and that cannot be addressed by XAI solutions alone. This work sets out to broaden transparency need-finding approaches within FAccT and the larger RAI research community by demonstrating what multi-stakeholder engagement looks like in the context of a 5-month case study centered around a widely-deployed AI-decision support tool. By leveraging a diverse set of need-finding methods (*i.e.*, meetings with the ML and product teams, experiments with an ML model, analysis of community forum posts, interviews with end-users, analysis of user-created training documents), we investigate the utility of currently deployed transparency mechanisms and provide insights about end-users' unique transparency needs. Our discussion highlights recommendations for transparency researchers to consider when approaching need-finding in practice. We intend to experiment with a variety of transparency solutions moving forward such as reframing match scores, providing algorithmic performance information to end-users, or building explanations for human-human-AI interactions. While our work does not cover all possible approaches for identifying transparency needs, we invite researchers from FAccT and beyond to leverage such approaches as a stepping stone to adopting more varied need-finding methods in their work.

7 RESPONSIBLE AI RESEARCH IMPACT STATEMENT

Ethical Considerations. All authors have completed the CITI certification and our study has received IRB approval. However, it is important to acknowledge that the community forum is a public forum, even if you are not a user of the tool. Specific individuals could be revealed if one were to search for their direct quotes on the forum. While none of the comments would bring harm to an individual if they were linked back to that individual, this still is an ethical consideration to take seriously in transparency work.

Authors' Positionality. The authors span a wide range of research topics aside from FAccT and HCI and hold positions across research, industry, and government, providing a unique perspective on the lifecycle of developing AI systems and transparency mechanisms. The author team is collaborating with Wild Me as an external research team. Wild Me provides us access to their ML team, product team, and Wildbook platforms. They did not provide us direct access to any end-users. In exchange, we provide recommendations and prototypes for improving the transparency of their system.

Adverse and Unintended Impacts. While we hope to raise awareness of wildlife conservation as a space that needs more HCI and FAccT involvement, possible unintended impacts could arise. For example, improving these tools to be more user-friendly and accessible could make it easier for malicious actors to take advantage of the knowledge (e.g., poachers).

ACKNOWLEDGMENTS

We would like to thank the Wild Me Lab of Conservation X Labs for their collaboration on this work.

This material is based upon work funded and supported by the Department of Defense under Contract No. FA8702-15-D-0002 with Carnegie Mellon University for the operation of the Software Engineering Institute, a federally funded research and development center.

REFERENCES

[1] Chirag Agarwal, Satyapriya Krishna, Eshika Saxena, Martin Pawelczyk, Nari Johnson, Isha Puri, Marinka Zitnik, and Himabindu Lakkaraju. 2022. OpenXAI: Towards a Transparent Evaluation of Model Explanations. https://arxiv.org/abs/2206.11104v3

[2] Ahmed Alqaraawi, Martin Schuessler, Philipp Weiß, Enrico Costanza, and Nadia Berthouze. 2020. Evaluating saliency map explanations for convolutional neural networks: a user study. In *Proceedings of the 25th international conference on intelligent user interfaces*. 275–285. https://doi.org/10.1145/3377325.3377519

[3] Vijay Arya, Rachel K. E. Bellamy, Pin-Yu Chen, Amit Dhurandhar, Michael Hind, Samuel C. Hoffman, Stephanie Houde, Q. Vera Liao, Ronny Luss, Aleksandra Mojsilović, Sami Mourad, Pablo Pedemonte, Ramya Raghavendra, John Richards, Prasanna Sattigeri, Karthikeyan Shanmugam, Moninder Singh, Kush R. Varshney, Dennis Wei, and Yunfeng Zhang. 2019. One Explanation Does Not Fit All: A Toolkit and Taxonomy of AI Explainability Techniques. https://arxiv.org/abs/1909.03012v2

[4] Tanya Y Berger-Wolf, Daniel I Rubenstein, Charles V Stewart, Jason A Holmberg, Jason Parham, Sreejith Menon, Jonathan Crall, Jon Van Oast, Emre Kiciman, and Lucas Joppa. 2017. Wildbook: Crowdsourcing, computer vision, and data science for conservation. *arXiv preprint arXiv:1710.08880* (2017). https://arxiv.org/abs/1710.08880

[5] Hugh Beyer and Karen Holtzblatt. 1999. Contextual design. *interactions* 6, 1 (1999), 32–42. https://dl.acm.org/doi/fullHtml/10.1145/291224.291229

[6] Drew Blount, Shane Gero, Jon Van Oast, Jason Parham, Colin Kingen, Ben Scheiner, Tanya Stere, Mark Fisher, Gianna Minton, Christin Khan, et al. 2022. Flukebook: an open-source AI platform for cetacean photo identification. *Mammalian Biology* 102, 3 (2022), 1005–1023. https://doi.org/10.1007/s42991-021-00221-3

[7] Angie Boggust, Benjamin Hoover, Arvind Satyanarayan, and Hendrik Strobelt. 2022. Shared interest: Measuring human-ai alignment to identify recurring patterns in model behavior. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. 1–17. https://doi.org/10.1145/3491102.3501965

[8] Angie Boggust, Harini Suresh, Hendrik Strobelt, John Guttag, and Arvind Satyanarayan. 2023. Saliency Cards: A Framework to Characterize and Compare Saliency Methods. In *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency*. 285–296. https://doi.org/10.1145/3593013.3593997

[9] Eve Bohnett, Jason Holmberg, Sorosh Poya Faryabi, Li An, Bilal Ahmad, Wajid Rashid, and Stephane Ostrowski. 2023. Comparison of two individual identification algorithms for snow leopards (Panthera uncia) after automated detection. *Ecological Informatics* 77 (2023), 102214. https://doi.org/10.1016/j.ecoinf.2023.102214

[10] Garrick Cabour, Andrés Morales, Élise Ledoux, and Samuel Bassetto. 2021. Towards an Explanation Space to Align Humans and Explainable-AI Teamwork. https://doi.org/10.48550/arXiv.2106.01503 arXiv:2106.01503 [cs].

[11] Ángel Alexander Cabrera, Abraham J Druck, Jason I Hong, and Adam Perer. 2021. Discovering and validating ai errors with crowdsourced failure reports. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW2 (2021), 1–22. https://doi.org/10.1145/3479569

[12] Ángel Alexander Cabrera, Erica Fu, Donald Bertucci, Kenneth Holstein, Ameet Talwalkar, Jason I Hong, and Adam Perer. 2023. Zeno: An interactive framework for behavioral evaluation of machine learning. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–14. https://doi.org/10.1145/3544548.3581268

[13] Ángel Alexander Cabrera, Adam Perer, and Jason I Hong. 2023. Improving human-AI collaboration with descriptions of AI behavior. *arXiv preprint arXiv:2301.06937* (2023). https://doi.org/10.1145/3579612

[14] Carrie J Cai, Emily Reif, Narayan Hegde, Jason Hipp, Been Kim, Daniel Smilkov, Martin Wattenberg, Fernanda Viegas, Greg S Corrado, Martin C Stumpe, et al. 2019. Human-centered tools for coping with imperfect algorithms during medical decision-making. In *Proceedings of the 2019 chi conference on human factors in computing systems*. 1–14. https://doi.org/10.1145/3290605.3300234

[15] Carrie J Cai, Samantha Winter, David Steiner, Lauren Wilcox, and Michael Terry. 2019. " Hello AI": uncovering the onboarding needs of medical practitioners for human-AI collaborative decision-making. *Proceedings of the ACM on Human-computer Interaction* 3, CSCW (2019), 1–24. https://doi.org/10.1145/3359206

[16] Carrie J Cai, Samantha Winter, David Steiner, Lauren Wilcox, and Michael Terry. 2021. Onboarding Materials as Cross-functional Boundary Objects for Developing AI Assistants. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–7. https://doi.org/10.1145/3411763.3443435

[17] Stephen Casper, Tong Bu, Yuxiao Li, Jiawei Li, Kevin Zhang, Kaivalya Hariharan, and Dylan Hadfield-Menell. 2023. Red Teaming Deep Neural Networks with Feature Synthesis Tools. In *Thirty-seventh Conference on Neural Information Processing Systems*. https://arxiv.org/abs/2302.10894

[18] Aditya Chattopadhay, Anirban Sarkar, Prantik Howlader, and Vineeth N Balasubramanian. 2018. Grad-CAM++: Generalized Gradient-Based Visual Explanations for Deep Convolutional Networks. In *2018 IEEE Winter Conference on Applications of Computer Vision (WACV)*. IEEE. https://doi.org/10.1109/wacv.2018.00097

[19] Haomin Chen, Catalina Gomez, Chien-Ming Huang, and Mathias Unberath. 2022. Explainable Medical Imaging AI Needs Human-Centered Design: Guidelines and Evidence from a Systematic Review. https://doi.org/10.48550/arXiv.2112.12596 arXiv:2112.12596 [cs, eess].

[20] Douglas Cirqueira, Dietmar Nedbal, Markus Helfert, and Marija Bezbradica. 2020. Scenario-Based Requirements Elicitation for User-Centric Explainable AI. In *Machine Learning and Knowledge Extraction*, Andreas Holzinger, Peter Kieseberg, A Min Tjoa, and Edgar Weippl (Eds.). Springer International Publishing, Cham, 321–341. https://doi.org/10.1007/978-3-030-57321-8_18

[21] Eric Corbett and Emily Denton. 2023. Interrogating the T in FAccT. In *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency*. 1624–1634. https://doi.org/10.1145/3593013.3594104

[22] Lorenzo Corti, Rembrandt Oltmans, Jiwon Jung, Agathe Balayn, Marlies Wijsenbeek, and Jie Yang. 2024. "It Is a Moving Process": Understanding the Evolution of Explainability Needs of Clinicians in Pulmonary Medicine.. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*. https://doi.org/10.1145/3613904.3642551

[23] Jonathan P Crall, Charles V Stewart, Tanya Y Berger-Wolf, Daniel I Rubenstein, and Siva R Sundaresan. 2013. Hotspotter—patterned species instance recognition. In *2013 IEEE workshop on applications of computer vision (WACV)*. IEEE, 230–237. https://doi.org/10.1109/wacv.2013.6475023

[24] Anamaria Crisan, Margaret Drouhard, Jesse Vig, and Nazneen Rajani. 2022. Interactive model cards: A human-centered approach to model documentation. In *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*. 427–439. https://doi.org/10.1145/3531146.3533108

[25] Wesley Hanwen Deng, Boyuan Guo, Alicia Devrio, Hong Shen, Motahhare Eslami, and Kenneth Holstein. 2023. Understanding Practices, Challenges, and Opportunities for User-Engaged Algorithm Auditing in Industry Practice. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–18. https://doi.org/10.1145/3544548.3581026

[26] Wesley Hanwen Deng, Michelle S Lam, Ángel Alexander Cabrera, Danaë Metaxa, Motahhare Eslami, and Kenneth Holstein. 2023. Supporting User Engagement in Testing, Auditing, and Contesting AI. In *Companion Publication of the 2023 Conference on Computer Supported Cooperative Work and Social Computing*. 556–559. https://doi.org/10.1145/3584931.3611279

[27] Upol Ehsan, Q Vera Liao, Michael Muller, Mark O Riedl, and Justin D Weisz. 2021. Expanding explainability: Towards social transparency in ai systems. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–19. https://doi.org/10.1145/3411764.3445188

[28] Upol Ehsan, Q. Vera Liao, Samir Passi, Mark O. Riedl, and Hal Daume III. 2022. Seamful XAI: Operationalizing Seamful Design in Explainable AI. https://doi.org/10.48550/arXiv.2211.06753 arXiv:2211.06753 [cs].

[29] Upol Ehsan and Mark O Riedl. 2021. Explainability pitfalls: Beyond dark patterns in explainable AI. *arXiv preprint arXiv:2109.12480* (2021). https://arxiv.org/abs/2109.12480

[30] Upol Ehsan, Koustuv Saha, Munmun De Choudhury, and Mark O Riedl. 2023. Charting the Sociotechnical Gap in Explainable AI: A Framework to Address the Gap in XAI. *Proceedings of the ACM on Human-Computer Interaction* 7, CSCW1 (2023), 1–32. https://doi.org/10.1145/3579467

[31] Malin Eiband, Hanna Schneider, Mark Bilandzic, Julian Fazekas-Con, Mareike Haug, and Heinrich Hussmann. 2018. Bringing Transparency Design into Practice. In *23rd International Conference on Intelligent User Interfaces (IUI '18)*. Association for Computing Machinery, New York, NY, USA, 211–223. https://doi.org/10.1145/3172944.3172961

[32] Florian Eyert and Paola Lopez. 2023. Rethinking Transparency as a Communicative Constellation. In *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency*. 444–454. https://doi.org/10.1145/3593013.3594010

[33] Nancy Friday, Tim D Smith, Peter T Stevick, and Judy Allen. 2000. Measurement of photographic quality and individual distinctiveness for the photographic identification of humpback whales, Megaptera novaeangliae. *Marine Mammal Science* 16, 2 (2000), 355–374. https://doi.org/10.1111/j.1748-7692.2000.tb00930.x

[34] Umm-e Habiba, Justus Bogner, and Stefan Wagner. 2022. Can Requirements Engineering Support Explainable Artificial Intelligence? Towards a User-Centric Approach for Explainability Requirements. https://doi.org/10.

48550/arXiv.2206.01507 arXiv:2206.01507 [cs].

[35] Sungsoo Ray Hong, Jessica Hullman, and Enrico Bertini. 2020. Human factors in model interpretability: Industry practices, challenges, and needs. *Proceedings of the ACM on Human-Computer Interaction* 4, CSCW1 (2020), 1–26. https://doi.org/10.1145/3392878

[36] Jinbin Huang, Aditi Mishra, Bum Chul Kwon, and Chris Bryan. 2022. ConceptExplainer: Interactive explanation for deep neural networks from a concept perspective. *IEEE Transactions on Visualization and Computer Graphics* 29, 1 (2022), 831–841. https://doi.org/10.1109/tvcg.2022.3209384

[37] Weina Jin, Jianyu Fan, Diane Gromala, Philippe Pasquier, and Ghassan Hamarneh. 2022. EUCA: the End-User-Centered Explainable AI Framework. https://doi.org/10.48550/arXiv.2102.02437 arXiv:2102.02437 [cs].

[38] Nari Johnson, Ángel Alexander Cabrera, Gregory Plumb, and Ameet Talwalkar. 2023. Where Does My Model Underperform? A Human Evaluation of Slice Discovery Algorithms. *arXiv preprint arXiv:2306.08167* (2023). https://doi.org/10.1609/hcomp.v11i1.27548

[39] Ja Ae Kim, Kenneth Mascola, and Yoan K Kagoma. 2021. Re-Envisioning On-Call Resident Workflows: Impact on Resident Experience. https://www.rsna.org/-/media/Files/RSNA/Practice-Tools/Quality-improvement/Quality-improvement-reports/2021/Re-Envisioning_On-Call_Resident-QDP-QI-20_RSNA_.ashx Accessed: January 22, 2022.

[40] Sunnie SY Kim, Elizabeth Anne Watkins, Olga Russakovsky, Ruth Fong, and Andrés Monroy-Hernández. 2023. "Help Me Help the AI": Understanding How Explainability Can Support Human-AI Interaction. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–17. https://doi.org/10.1145/3544548.3581001

[41] Sunnie SY Kim, Elizabeth Anne Watkins, Olga Russakovsky, Ruth Fong, and Andrés Monroy-Hernández. 2023. Humans, AI, and Context: Understanding End-Users' Trust in a Real-World Computer Vision Application. In *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency*. 77–88. https://doi.org/10.1145/3593013.3593978

[42] Brett Koonce and Brett Koonce. 2021. EfficientNet. *Convolutional Neural Networks with Swift for Tensorflow: Image Recognition and Dataset Categorization* (2021), 109–123. https://link.springer.com/chapter/10.1007/978-1-4842-6168-2_10

[43] Michelle S Lam, Mitchell L Gordon, Danaë Metaxa, Jeffrey T Hancock, James A Landay, and Michael S Bernstein. 2022. End-User Audits: A System Empowering Communities to Lead Large-Scale Investigations of Harmful Algorithmic Behavior. *Proceedings of the ACM on Human-Computer Interaction* 6, CSCW2 (2022), 1–34. https://doi.org/10.1145/3555625

[44] Rena Li, Sara Kingsley, Chelsea Fan, Proteeti Sinha, Nora Wai, Jaimie Lee, Hong Shen, Motahhare Eslami, and Jason Hong. 2023. Participation and Division of Labor in User-Driven Algorithm Audits: How Do Everyday Users Work together to Surface Algorithmic Harms?. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–19. https://dl.acm.org/doi/full/10.1145/3544548.3582074

[45] Q Vera Liao, Daniel Gruen, and Sarah Miller. 2020. Questioning the AI: informing design practices for explainable AI user experiences. In *Proceedings of the 2020 CHI conference on human factors in computing systems*. 1–15. https://doi.org/10.1145/3313831.3376590

[46] Q. Vera Liao, Milena Pribić, Jaesik Han, Sarah Miller, and Daby Sow. 2021. Question-Driven Design Process for Explainable AI User Experiences. https://doi.org/10.48550/arXiv.2104.03483 arXiv:2104.03483 [cs].

[47] Q Vera Liao, Hariharan Subramonyam, Jennifer Wang, and Jennifer Wortman Vaughan. 2023. Designerly understanding: Information needs for model transparency to support design ideation for AI-powered user experience. In *Proceedings of the 2023 CHI conference on human factors in computing systems*. 1–21. https://doi.org/10.1145/3544548.3580652

[48] Q. Vera Liao and Jennifer Wortman Vaughan. 2023. AI Transparency in the Age of LLMs: A Human-Centered Research Roadmap. arXiv:2306.01941 [cs.HC] https://doi.org/10.1162/99608f92.8036d03b

[49] Nathan J McNeese, Beau G Schelble, Lorenzo Barberis Canonico, and Mustafa Demir. 2021. Who/what is my teammate? Team composition considerations in human–ai teaming. *IEEE Transactions on Human-Machine Systems* 51, 4 (2021), 288–299. https://doi.org/10.1109/thms.2021.3086018

[50] Tim Miller. 2023. Explainable AI is Dead, Long Live Explainable AI! Hypothesis-driven Decision Support using Evaluative AI. In *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency*. 333–342. https://doi.org/10.1145/3593013.3594001

[51] Margaret Mitchell, Simone Wu, Andrew Zaldivar, Parker Barnes, Lucy Vasserman, Ben Hutchinson, Elena Spitzer, Inioluwa Deborah Raji, and Timnit Gebru. 2019. Model cards for model reporting. In *Proceedings of the conference on fairness, accountability, and transparency*. 220–229. https://doi.org/10.1145/3287560.3287596

[52] Sina Mohseni, Niloofar Zarei, and Eric D. Ragan. 2020. A Multidisciplinary Survey and Framework for Design and Evaluation of Explainable AI Systems. https://doi.org/10.48550/arXiv.1811.11839 arXiv:1811.11839 [cs].

[53] Hussein Mozannar, Jimin J Lee, Dennis Wei, Prasanna Sattigeri, Subhro Das, and David Sontag. 2023. Effective Human-AI Teams via Learned Natural Language Rules and Onboarding. *arXiv preprint arXiv:2311.01007* (2023). https://proceedings.neurips.cc/paper_files/paper/2023/file/61355b9c218505505d1bedede9da56b2-Paper-Conference.pdf

[54] Hussein Mozannar, Arvind Satyanarayan, and David Sontag. 2022. Teaching humans when to defer to a classifier via exemplars. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 36. 5323–5331. https://doi.org/10.1609/aaai.v36i5.20469

[55] Imani Munyaka, Zahra Ashktorab, Casey Dugan, J Johnson, and Qian Pan. 2023. Decision Making Strategies and Team Efficacy in Human-AI Teams. *Proceedings of the ACM on Human-Computer Interaction* 7, CSCW1 (2023), 1–24. https://doi.org/10.1145/3579476

[56] Mahima Pushkarna, Andrew Zaldivar, and Oddur Kjartansson. 2022. Data cards: Purposeful and transparent dataset documentation for responsible ai. In *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*. 1776–1826. https://doi.org/10.1145/3531146.3533231

[57] Marissa Radensky, Doug Downey, Kyle Lo, Zoran Popovic, and Daniel S Weld. 2022. Exploring the role of local and global explanations in recommender systems. In *CHI Conference on Human Factors in Computing Systems Extended Abstracts*. 1–7. https://doi.org/10.1145/3491101.3519795

[58] Gesina Schwalbe and Bettina Finzel. 2023. A Comprehensive Taxonomy for Explainable Artificial Intelligence: A Systematic Survey of Surveys on Methods and Concepts. *Data Mining and Knowledge Discovery* (Jan. 2023). https://doi.org/10.1007/s10618-022-00867-8 arXiv:2105.07190 [cs].

[59] Ramprasaath R Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, and Dhruv Batra. 2017. Grad-cam: Visual explanations from deep networks via gradient-based localization. In *Proceedings of the IEEE international conference on computer vision*. 618–626. https://doi.org/10.1109/iccv.2017.74

[60] Sahil Singla and Soheil Feizi. 2022. Salient ImageNet: How to discover spurious features in Deep Learning?. In *International Conference on Learning Representations*. https://openreview.net/forum?id=XVPqLyNxSyh

[61] Timo Speith. 2022. A review of taxonomies of explainable artificial intelligence (XAI) methods. In *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*. 2239–2250. https://doi.org/10.1145/3531146.3534639

[62] Simone Stumpf, Adrian Bussone, and Dympna O'sullivan. 2016. Explanations considered harmful? user interactions with machine learning systems. In *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems (CHI)*. https://www.yumpu.com/cs/document/view/55465112/explanations-considered-harmful-user-interactions-with-machine-learning-systems

[63] Ashley Suh, Gabriel Appleby, Erik W Anderson, Luca Finelli, Remco Chang, and Dylan Cashman. 2023. Are Metrics Enough? Guidelines for Communicating and Visualizing Predictive Models to Subject Matter Experts. *IEEE Transactions on Visualization and Computer Graphics* (2023). https://doi.org/10.1109/tvcg.2023.3259341

[64] Sana Tonekaboni, Shalmali Joshi, Melissa D McCradden, and Anna Goldenberg. 2019. What clinicians want: contextualizing explainable machine learning for clinical end use. In *Machine learning for healthcare conference*. PMLR, 359–380. http://proceedings.mlr.press/v106/tonekaboni19a/tonekaboni19a.pdf

[65] Data Source Triangulation. 2014. The use of triangulation in qualitative research. In *Oncol nurs forum*, Vol. 41. 545–7. https://onf.ons.org/pubs/article/233796/preview

[66] Michael Vössing, Niklas Kühl, Matteo Lind, and Gerhard Satzger. 2022. Designing transparency for effective human-AI collaboration. *Information Systems Frontiers* 24, 3 (2022), 877–895. https://doi.org/10.1007/s10796-022-10284-3

[67] WildMe. 2023. Image Analysis Pipeline - Wild Me Documentation. https://docs.wildme.org/product-docs/en/wildbook/introduction/image-analysis-pipeline/#matching-and-interpreting-embeddings-for-wildlife-id-miew-id-or-id

[68] Michael Williams and Tami Moser. 2019. The art of coding and thematic exploration in qualitative research. *International management review* 15, 1 (2019), 45–55. https://www.academia.edu/download/82465402/imr-v15n1art4.pdf

[69] Yao Xie, Melody Chen, David Kao, Ge Gao, and Xiang'Anthony' Chen. 2020. CheXplain: enabling physicians to explore and understand data-driven, AI-enabled medical imaging analysis. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–13. https://doi.org/10.1145/3313831.3376807

[70] Andrew Zaldivar, Ben Hutchinson, Elena Spitzer, Inioluwa Deborah Raji, Lucy Vasserman, M. Mitchell, Parker Barnes, Simone Sanoian McCloskey Wu, and Timnit Gebru (Eds.). 2019. *Model Cards for Model Reporting*. https://doi.org/10.1145/3287560.3287596

A COMMUNITY FORUM

The community forum (Figure 3) is open to the public, and all users have access to it. Figure 3 is further below due to formatting constraints. On the left-hand side, users can filter the form by category to quickly find the latest posts about a specific topic.

A.1 Codebook with Examples

View the table on the following pages.

Axial Categories	Axial Coding Examples
Bulk Import	"This user's bulk import was sent to detection approx. 19hrs ago with no detections completed. I see that another user has uploaded a whole lot of batches in the past day+ but all of those batches appear to have completed detection so I'm not sure if there's still volume that's preventing this batch from getting through detection or if there's a different problem. There I've asked the user to send me the spreadsheet but I don't think the issue is there. Appreciate your help with this one."
How To	"Hello, I'm working with Sharkbook and was doing a regional check of our encounters to see if there were any match with other spots (our sharks seen anywhere else). After checking and finding a highly potential match, an individual merge was apparently launched between the two sharks… However, we wanted to do a more thorough analysis before doing it, and I don't know what happened (I misclicked maybe?) … Now the merge will be effective in a week, is there a way to cancel the merge while we do our check? Thank you in advance"
Please Explain	"Hello, i want to upload my Wildbook Standard Format XLS template in Amphibian and Reptile Wildbook. How do I to fill in Excel column "Encounter.measurementX" (cm; m) ? Thank you!"
Please add	"Hi, could you add Bulgaria to the list of countries, or to LocationID, under Black Sea? Thanks!"
Match Scores	"Hi! Regarding "Match score", in Wild Me Documentation it is stated that "The match score represents the numeric value returned from the algorithm". But what exactly are these numeric values? What do they represent: percentage, a score from 0 to 1, or 0 to 10? Or something else? Thanks!"
Merging IDs	"It seems that I have some individuals with the same name. For example twice LG-0061F, not sure how that happens. Is there a way to merge them?"
System error message	"Error message when trying to Inspect match results on multiple encounters Wildbook for Carnivores Wildbook for Carnivores African Carnivore Wildbook Error accessing Inspection image1920×894 121 KB. I should be able to inspect match result?
Metadata category	"We would like to add to individual metadata two dates: Birth and Death. Some kind of "Notes" box would also be very good. Birth and death dates are known for some individuals and data was in Norppagalleria, at the moment we don't have this in Codex, we can't see if the animal is confirmed dead. To the "notes" box we could mark for example individual's known pups that don't have Phs-code."
Email alerts	"I uploaded three new encounters and did not get automatic emails notifying me. What did you expect to happen? I usually get an email anytime a new encounter is uploaded What are some steps we could take to reproduce the issue? If this is a bulk import report, send the spreadsheet to services@wildme.org with the email subject line matching your bug report"
New Species	"Would it be possible to add long-finned pilot whales as an option to the list of recognized species in Flukebook? We are looking to try the general matching algorithm and the only options currently are short-finned pilot whales or unknown globicephala - neither of which would be accurate. Thanks!"
Using multiple images to conduct a match	"Very often we have to use the "start a new match" functionality : for example when an annotation is to be added or modified, or when we subsequently add an image to an encounter (this happens often, as images are extracted from videos), or after deleting a duplicate encounter. When you "start a new match", it can only be on one photo and it "dissociates" this photo from the others in the match results display. Thus, we no longer have access to the page with the match results for each photo one below the other, which is an extremely practical display. To not lose this great functionality, we should either: Be able to "start a new match" on all the photos of the encounter, exactly like when you first submitted the photos. Or always display the match results for each photo on the same page, whether the match was made at the same time when the photos were submitted or whether it was made later for one or more of the images. How would this functionality help you? This would avoid degrading the ergonomics (which is so great !) of the match results when we have to redo a match on one of the images, which often happens!"
Zooming capabilities	"The images on the matching panels can come through quite small, depending on how they're cropped – I'd love to be able to zoom in on certain areas of the photo. If I could zoom in on the 'thumbnail' it'd save me having to pop open the potential match encounter in another window to view the photo larger."
Data replaced by new upload	"I unintentionally replaced the historical data from Sighting 9, August 26 2019 with different photos. I wanted to test the algorithm by importing Full Frame photos with slightly different file names for one sighting from 2019. However, upon import, it seems to have displaced the historical data with the full frame images adopting the original file names (not the new ones) and clearing the previous ID #s for that sighting."
Customer Support needed to succeed	"Two bulk imports are stuck - could someone please run them through detection again?"

Axial Categories	Axial Coding Examples
Add to project from encounters	"I want the ability to add an encounter to a project from the encounter page (rather than going through the search and project management way which can be time consuming, especially when you don't seem to be able to use the encounter number to search by)."
Exporting Viewpoints	"I've had a question from Panthera about exports that I can't answer - are viewpoints exportable? Thanks!"
Server down	"The ACW Hydra server is down again"
Timeout indicator	"I would like to see a pop-up that says something like "You have been logged out due to inactivity (timeout), please log back in". There is a "timeout" which disconnects us from Whiskerbook after a certain time, but the page remains visually the same so we don't know we are no longer connected."
Alternate IDs	"I'd like to set an 'alternate ID'. I've tried adding the numerical ID to the MarkedIndividual.nickname field, hoping that might be linked to the Alternate ID field on the encounter page, but I see it isn't. Is there a way to accomplish this?"
Process taking longer than usual	"Hi, i have two assistants helping me run scan tasks on the whale sharks from Ningaloo, and recently they seem to be running very slowly. Today, 13th June, one of my assistants reported that even after an hour of the process running there was no movement."
Matching IDs	"I would appreciate having the option to select whether identifications should be matched against other datasets. What I mean is that it's not meaningful to compare the IDs of hyenas from Ghana to Senegal since there is absolutely zero connectivity between these two countries. Therefore, it could be valuable to have the ability to choose which sites, countries, or datasets we want to use for identification matches."
Metadata not displaying properly	"Hi. You can see that the é, à, è are written normally in the Excel export. On the other hand, when contributors add their comments to the encounter page ("Attributes → Additional comment") it goes to the same export colum, but the letters "é" "è" "à" and others weird french letters are not displayed correctly in the export (french letters : é, è, à, ç, ë, ê, ù, â, ê, î, ï, ô, ö, û, ü). See here with the encounter remarks "lien probable avec une autre fiche du même jour à 15h41 sur ce même site" and "peut-être un lien avec la vidéo du même jour à 3h21 sur ce même site" and "très mauvaise qualité". I expected the characters entered in the encounter to be displayed normally in the Excel export, the same way as when they are entered through a bulk import."
Export images	"We have a cowowrker who will be working on developing out ML for individual recognition. She needs the images themselves, and a table mapping each file to some individual ID. We can export the data we need just fine but getting the images has been the challenge. We dont have access to all the imags on a single computer or drive so we were hoping to use the URL to extract the data from GS. However, with the user permissions, this doesn't seem feasible. Is there any other way to do this?"
Access to project	"I would like to see all the meetings of Project "Mar de Tartaruga" so I can analyze the data for my research"
Historical data issue	"The ability to assign birth and death years to individuals, and for Flukebook to understand not to propose whales as matches for sightings that occur when they were not alive. This would streamline matching, since then Flukebook doesn't propose whales as matches outside of the years they were alive."
Incorrect annotation	"The annotation for the whale is completely off - it is around water rather than the whale. Also, the annotations for all of the matching candidates are also very strange. They are mostly going the wrong way across the whales."
Manual annotations	"I write you because I have entered a new encounter in Flukebook of blue whales. It is necessary to create a manual annotation for the algorithm identification, but in the manual annotation it is not possible to select the iaClass for this whale species."
Incorrect tag	"Hi !! Somebody submitted an encounter and I noticed it was tagged as lynx rufus. I deleted the encounter and submitted a new one with the same photos again and I did choose the attribute lynx lynx, but the pictures are again tagged as lynx Rufus. The pictures should be tagged lynx lynx."
Wording change	"Current text: "We will use your information to assist a global community of researchers and citizen scientists working to better understand and protect the world's biggest fish." I presume this was from whaleshark.org, but it could be updated for a more general response to something like: "We will use your information to assist a global community of researchers and citizen scientists working to better understand and protect the world's shark species.""
Visual Matcher	"We often know which 3 or 4 lynxes frequent an area of the mountain range. The software searches the entire database and sometimes does not find any relevant results. It would be extremely useful to have a kind of "visual matcher", to be able to compare the photo by eye with photos from the same viewpoint of this or that lynx which we know often frequents the area."

Axial Categories	Axial Coding Examples
Video	"Hi !! The pictures inside an encounter cas be deleted through the “hamburger menu” but it doesn’t appear on a video, so I can’t delete them."
Access to site without login account	"I’m working for an NGO in the Philippines and we are currently thinking of the best way to implement IoT tere. Is there any way for us to see the submissions of tourists/divers who would submit their encounters themself on IoT without an account, and add their submissions to our projects? Or do we have to submit their pictures ourself with the researcher account we have?"
Deleting encounters	"Have selected delete encounter under metadata but it is still appearing under unapproved encounters Internet of Turtles Internet of Turtles Internet of Turtles is a visual database of Turtle encounters and of individually catalogued Turtles."
Hot Spotter	"A user is reporting frequent examples of where the background is showing as “hotspotted” on the Inspect image comparison. I know background subtraction isn’t perfect but I thought I should check in to confirm, or not, whether background subtraction is actually being applied in Whiskerbook (and Wild North, for that matter), specifically to lynxes. Thanks!"
Sightings not loading	"No sightings showing up today. I have tried to search by owner, date and place and nothing shows up.
Relationships	"Kindly assist me in adding relationship especially for the mother and foal."
UX/UI Improvements	"As you go along and make matches between encounters, if you could conceal/hide those lines in the matching table that shows the ranked scores. Conversely, if we could click “not a match” and have that line/encounter hidden from the field of view in the matching/ranking table that would be helpful. We are digging deep into our photo data beyond the first 50 rankings and visibility on the screen layout is an issue. Concealing matches would be helpful to free up visual space as well as show us where we left off with the matching process. Concealing ‘not matching’ encounters would also help us visually pick up where we left off in the process."
Shared account access	"We have usually about 5 students working for short term in Saimaa ringed seal research, mainly helping us with seal photograph editing and submitting sightings to Codex. I was wondering, if it would be smarter or even possible to have 1 or more shared “student usernames”, so we wouldn’t have to create so many accounts for people that will not work with Codex much longer than couple months? So after one student finnishes, next one gets access to same account."
Manually change view points	"Is there a way to assign different viewpoints (of the annotation) for different images of the same indiviudal during the same encounter in the Amphibian and Reptile Wildbook? Let’s say, I find a fire salamander and I take a photo of the salamander from above, from one side and from below (the ventral side). Now, I want to upload all photos assigned to the same encounter, both to use these multiple images for identification and to share these images with my collaborateurs for research projects. Once, I have bulk imported my images, sent them to detection and identification and added them to my project, all images have an annotation assigned to them. However, by default the viewpoint of the annotation will be “up”. When I check match results in my project, the photos e.g., from below will also be matched against images from above. It is possible to change the viewpoint of the annotation manually?"
Un-merging IDs	I have a female leopard 0006 that was merged with two other individuals. 0003 is the same as 0006 but the images that were under 0003 do not appear under LF0006 and she still has her own page. I don’t understand why. I would like all the images to be under 0006. Then 0006 was merged with 0004 which is definitely a mistake. I would like to unmerge those individuals. Surprisingly though, the images of 0004 also did not appear under 0006."
Admin Access	"The library is maintained and used by marine biologists to collect and analyse shark encounter data to learn more about these amazing creatures. Apparently no one in our group has Admin access, when it rolled over to Sharkbook, we lost it. We need it for two of our researchers. We’re trying to add a an additional researcher and are unable to do so."
Unable to see matches	"When I click “view” under match results within the Icelandic Pilot Whale Identification Project, the images for neither the encounter being matched, nor the possible matches show up (even after waiting, logging out and back in, etc.). It appears to be that way for all of the remaining whales to match."
Merging sightings	"Hi there, I’m new in WildMe and need some help with the following. I uploaded 2 encounters, separately, to Sharkbook. These are of 2 different individuals. The encounters occurred in the same location at the same time, so they correspond to a Sighting. But they have been assigned different Sighting IDs. I guess this is because I uploaded them separately? How do I merge them to make them a Sighting with 2 encounters?"
Reassigning encounters	"Hi again, I made another rookie mistake. I’m working with Sharkbook. When uploading a new encounter my login expired so the encounter was not assigned to my account. How can I assign the encounter to my account now?"

Axial Categories	**Axial Coding Examples**
Account Access	"One of our research group member's username was accidentally inactivated a while ago. Now she would need need to start using Codex and I couldn't find a way to activate that username again. Can it be activated or removed so new can be made instead?"
No matchable detection	"Two different wild dog researchers have reported this issue - encounters with correct body annotations are showing as "no matchable detection" per this example. Here is the sighting link for the above encounter. Important to note that all encounters except the 2nd one in the list have "no matchable detection". These should all be able to be sent to identification. So far, I'm only aware of this problem with wild dog annotations. But it doesn't appear to be happening with all wild dog encounters/annotations, as in the sighting link above. No idea why. Your help is very much appreciated!"
Incorrect encounter matching	"In encounters with two sides, comparisons are being made with only one side and when the other side rotates, it appears as a new individual. Is thiss normal and how to proceed with the side that did not appear during the analyzes. I tried it on 2 different days and the same problem continued with the generated individual. I imagined that the two sides should appear, each with their possible compatibility together, on the same page."
Edit access	"I wanted to edit information for an encounter supporting the managing researcher, but could not edit the details. This is the case for all encounters assigned to one of the sublocations for Kenya. I think editing is not possible, because to the account attached to my email is listed in "Roles for Turtles" but not the individual sublocations. Could this be amended please?"
Automatic annotations aren't working	"I'm attempting to upload and annotate humpback whale flukes for matching purposes, but the flukes aren't being automatically annotated even after waiting for several days."
Old URL	"Hi there, since my presentation on Friday at the Leopard Conference, I've had a few enquiries from non-African leopard species to have access to Wildbook. I'm referring them to Whiskerbook but when I went to find the URL to refer them to, I noticed that the Wildbook Account request form on the Wild Me site doesn't have Whiskerbook in the list. I believe WB for Jaguars is what has now become Whiskerbook but that's not clear for new users looking to get a new account. Could someone fix this?"
Identification of individuals	"Hi, we want to avoid any risk of error in the identification of individuals who could "spread" over time. Indeed, if a collaborator is mistaken and identifies the lynx named "CHAMAR" on a photo when it is actually the lynx named "CHACOR", his misidentified pictured will serve as one of the references for future matches and we will again identify "CHACOR" as "CHAMAR" the next time. So we would need for example one of those two options: be able to assign a role to users allowing them to record an encounter, to consult the encounters, to launch a match and to see the results but not to validate this match or to assign an individual to a photo. We would then let only selected and more experienced users perform this critical last assignment step."
Delete photos	"Hi, I'm using flukebook. How can I delete wrong photos (or all photos) were uploaded during Bulk Import process?"
Candidate ranking	"I am wondering about why the tool says 'cannot start match'. I am matching mantas visually, but each time I need to sort through every encounter in the Yucatán to find a match (if it exists). I have found matches after sorting through many individuals that are very obviously not matches. I thought that the most similar candidates would be suggested first, but that doesn't seem to be happening. I am wondering if I am doing something wrong?"
incorrect email	"Hi, I tried emailing this email address auto@iot.wildlbook.org through gmail, and an automatic reply says this email addredd does not exist. This is the email address displayed when we successfully submit an encounter on the internet of turtles."
New Encounter	"The ability to specify when adding a new annotation that I want it to have a new encounter record. I'm trying to clean up some lion annotations in a bulk import and I'm finding that the system allows me to add more than 1 head part/annotation per encounter. This is a problem because the 2 heads belong to different lions. I believe that this is intentional functionality for Wildbooks where multiple photos are added to a single encounter however that creates an issue when there are also multiple animals in a single image - there's no way for me to force the system to create a new encounter for the separate individual. The lion detector misses lion heads fairly frequently, particularly in photos with more than 1 individual in them, which is VERY common, particularly in tourist images. If I'm not able to get the system to create a new encounter for every extra head annotation that I need to create, then that restricts me to only being able to ID 1 animal per photo, which severely limits our lion ID functionality. So some way to either have the system not allow multiple lion head parts per encounter or, preferably, to allow me, the user, to tell the system: when I add this new head part, put it in a new encounter record, not the old one, is what I'm looking for."

Axial Categories	**Axial Coding Examples**
Shapefile question	"Add the Default ID to the Shapefile export. The Shapefile export is very useful however the individual is identified by an alpha-numeric string without their default ID / name, which doesn't give the researcher the ID info they need to understand that data point. Hopefully this will also then populate the Sex column as well, which isn't populating currently. This would put all of the researcher's necessary datapoints into this export, and remove the need for additional cross-referencing and digging to determine which ID'd individual relates to which row in the export."
Automatic Annotations	"I am noticing two different IA classes in the automatically generated annotations: whale_orca and whale_orca+fin_dorsal. It looks like whale_orca is for when the dorsal fin and part of the body is visible while whale_orca+fin_dorsal appears to be for annotations of the dorsal fin only, but I was wondering if that interpretation is correct and/or if there is any documentation defining these categories."
Interest in research question	Dear WildMe operators, I am a 20-year-old student at the University of Groningen studying Global Responsibility and Leadership. I am currently exploring various topics for my bachelor thesis project. I am highly interested in doing research with crowdsource databases like those managed by Wild Me. What sparks my interest most is marine life and the species living under the water and I am thus mostly exploring the whale shark datasets. With this new topic, I would like to ask if there is any specific research that is wished to be done on crowdsource databases or the Wild Me in specific. I would love to be able to explore any collaboration possibilities. Looking forward to hearing from you!"
Match Results	"I will be quantifying the match success of flukebook for the killer whale population I am working with by uploading photos of known animals (with their IDs specified) to create a catalog to match against, and then uploading additional photos of known animals (without their IDs specified) to see where the correct whale is ranked in the match results. Is there a way that the match results can be exported in bulk to assess the suggested matches against my known IDs for those photos and to then bulk update the metadata for those photos with the correct IDs once the matching success has been assessed? For example, an ideal workflow I would picture is: Submit reference and test photos as described above. For all submitted photos put through the matching algorithms, export a table of IDs of possible match ID by their rank (for each algorithm) Upload an updated metadata sheet that includes the correct IDs for the test photos. Is anything like this possible? Or would I need to manually review the possible matches for each photo, manually record the correct ID's rank for each algorithm, and manually select the correct match (if found)? On a somewhat related note: is there documentation on how to export ID results once matching has been completed for a series of encounters?"
Add another animal to an encounter	"When I have finished uploading one encounter, it would be so helpful if there was an option to 'add another giraffe/photos to this encounter'. Which means that these photos/giraffe would be straight away added to that Occurence ID, with same coordinates and place. This would help save a lot of time."
Exporting data	"How do we export data from Codex to excel for analysis please?"
Information missing from record	"Something weird has happened to this Marked Individual record -there is lots of missing info: missing sections, missing encounter links and list, missing gallery, missing everything except the single ID. It used to have a merged ID as well, which, when I search for it, it takes me to the correct record: Does anyone have any idea what happened here? Without a log section, I can't see if the user did something to cause this or if something happened on the system side?"
Nicknaming an animal	"This is a new encounter I just submitted into Sharkbook. But when I go to "Nickname Me!" and went through the adoption payment process, I was not looped back to where you can nickname this individual. Instead I was brought to the gallery and prompted to select an individual. But I was not able to find it on the gallery just by scrolling through. I then tried typed in his ID in the search engine and found his page. I clicked "Nickname Me!" again, it prompted me to do the donation payment again. Due to this circle, my bank account has 3 pending charges now"
New language	"We're expecting to onboard new French-speaking users and so need to post French versions of our ToU. I've sent both updated versions to the services@wildme.org email address."
APIs access	"Get API access, this is already pretty much there in WBIA, it would really be nice to for research purpose. HTherefore, we would be able to programatically access the database of wildbook and take out images, and data of interest."
Viewing encounters	"When opening different sightings from individuals page, they should open in different tabs. This would help to open for example all sightings of one individual in a row of tabs in same window quickly."
New membership organizations	"Hi there, We've added a lot of new users lately and have a few new orgs to add to the membership list. When you have a chance please add the orgs below and please add my admin user ID to each of these orgs so I can add the correct users under each."
User Access Logs not displaying	"The user access log page does not appear to be loading correctly? We use that to make sure someone else isn't working on Unapproved Encounters at the same time so it is important to our work flow."

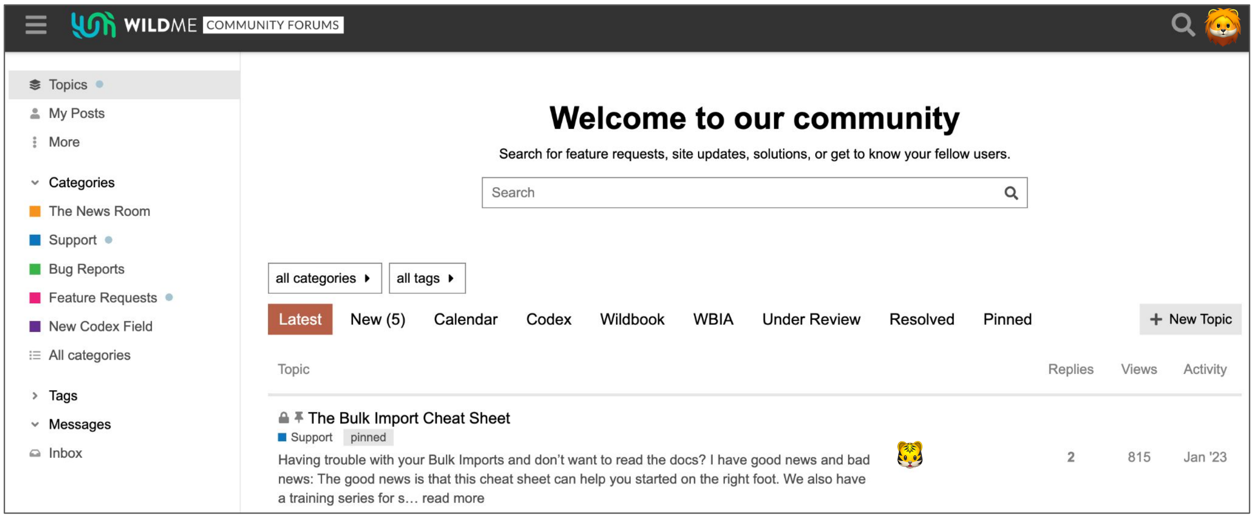

Figure 3: The home screen for Wild Me's community forum.

Axial Categories	Axial Coding Examples
Documentation	"Hi, We've had a wave of questions from new users about the export functionality in Wildbook. I've been trying to find documentation on the Wild Me Docs site but without success. If there isn't any documentation on that functionality, would it be possible to schedule a call to have someone at Wild Me walk us through the options?"
Viewpoints	"ACW users need to be able to search and filter based on viewpoint(s) detected. Currently, they can only use keywords to try to find # of animals with variations of left viewpoints, for example. However since keywords are based on the photo not the annotations in the photo, filtering using keywords for the purpose of understanding your viewpoint distribution, particularly of Marked Individuals, doesn't work. With this functionality, wsers want to be able to work out how many individuals they have that have at least one left or at least one right viewpoints. This affects the certainty of the total count of individuals in a dataset. Being able to filter to individuals with only lefts and no rights and vice versa, would allow the users to then assess whether or not any of these cases represent the same animal. For example: In a dataset of 700 individuals, we may have a minimum of 350 ID'd individuals with both a left and a right viewpoint assigned. Users want to know if they have 700 individuals or 350 or something in between. Being able to filter using viewpoints would facilitate this analysis."
Behaviour Descriptions	"Instead of now removing the annotation first and then adding annotation. Can't there be the option to change the existing annotation? Because often the annotation just needs to be adjusted a little bit."