Robin: An Intelligent Bird Observation Application for the Visually Impaired and K-12 Education

Jana Amin*, Meriam Harissa*, Jodi Joven*, Leah Mirch*, Zheng Song
Department of Computer and Information Science
University of Michigan – Dearborn
{janaamin, mharissa, jodijov,lmirch,zhesong}@umich.edu

***Abstract*—The Wildlife Observation Room at the University of Michigan-Dearborn's Environmental Interpretive Center (EIC) provides a space for visitors to observe birds and other wildlife in their natural habitat. Equipped with a live audio system that captures outdoor sounds and a large-screen video feed displaying the surrounding environment, the room offers an immersive experience. However, traditional birdwatching in this setting remains passive, requiring observers to rely on personal expertise or external references to identify species and learn about bird behaviors. Additionally, accessibility remains a challenge, particularly for visually impaired visitors, who rely more on auditory cues. To enhance engagement and accessibility, we introduce an intelligent bird observation system that addresses three key challenges: (1) Accurate real-time bird species identification using environmental audio, even in noisy conditions with multiple overlapping bird calls; (2) Interactive and informative engagement through AI-driven dialogue that allows users to ask and receive contextualized answers about birds; and (3) Predicting optimal birdwatching times to help visitors schedule their observations effectively. The system integrates automated bird sound recognition, AI-powered interaction, and predictive analytics, enabling real-time species detection, natural language-based learning, and data-driven visit planning. We evaluate its effectiveness in terms of bird identification accuracy under real-world acoustic conditions, prediction reliability for bird activity forecasting, and user interaction quality, with a focus on accessibility for visually impaired users. Our findings contribute to the development of intelligent and inclusive wildlife observation systems, demonstrating how AI-driven approaches can improve engagement, accessibility, and usability in nature-based learning environments.**

I. Introduction

Wildlife observation rooms offer a structured environment for visitors to engage with nature, providing a unique opportunity to observe and study wildlife in a controlled yet immersive setting. The Wildlife Observation Room (see Fig. 1) at the Environmental Interpretive Center (EIC) [1] of the University of Michigan-Dearborn enhances this experience through a large-screen video display and a speaker system that broadcasts live outdoor sounds, creating a seamless connection between visitors and the surrounding environment. As a key component of the EIC's mission to promote environmental education and accessibility, the Wildlife Observation Room regularly welcomes K-12 student groups [2], offering hands-on learning experiences that align with school curricula. Additionally, it provides an inclusive nature engagement experience for visually impaired visitors [3], utilizing real-time audio feeds to assist in bird and wildlife identification through sound. With over 40 years of dedication to experiential education, the EIC continues to serve students, researchers, and the broader community, fostering a deeper appreciation and understanding of the natural world [4].

Fig. 1. EIC Wildlife Observation Room

However, traditional birdwatching in such environments remains passive—observers must rely on personal expertise or external references to identify species and learn about bird behaviors. Accessibility remains a significant concern, particularly for visually impaired visitors, who primarily depend on auditory cues for bird identification and engagement. Additionally, K-12 students, who may have limited experience with birdwatching, require interactive and guided learning experiences to fully benefit from such educational opportunities.

To address these challenges, we introduce **Robin**, an intelligent bird observation application named after Michigan's state bird, the American Robin, symbolizing accessibility and connection to nature. Designed to enhance engagement for visually impaired users and K-12 students, Robin integrates AI-driven bird identification, interactive learning modules, and predictive analytics, transforming traditional birdwatching into an inclusive, educational, and immersive experience.

In particular, Robin 1) integrates automated bird sound recognition to identify species in real time, retrieving relevant information from vocalizations even in complex acoustic environments with multiple overlapping bird calls [5]; 2) includes a GenAI-based interactive learning module, enabling users to ask questions about birds and receive informative responses, enhancing engagement beyond traditional static resources [6];

The first four authors contributed equally to this paper, are co-first authors, and are listed in alphabetical order.

3) incorporates a predictive analytics module to estimate optimal birdwatching times based on historical bird activity data, allowing visitors to plan their observations more effectively; and 4) emphasizes accessibility by prioritizing audio-based interaction and structured information retrieval to ensure usability for visually impaired users. We evaluated Robin by its bird identification accuracy, prediction reliability, and user experience, particularly regarding accessibility. The findings contribute to the development of intelligent and inclusive wildlife observation systems, demonstrating how AI-driven solutions can enhance engagement, accessibility, and usability in environmental education and public learning spaces.

The remainder of this paper is organized as follows: Section II reviews related work. Section III details system design and implementation. Section IV presents our evaluation methodology and results. Finally, Section V concludes with findings and future directions.

II. Related Work

This section provides a brief overview of the existing approaches that form the foundation of Robin, highlighting the novel contributions introduced by Robin. Traditional bird identification relies on visual observation and manual bioacoustic analysis, requiring expertise and time. AI-driven tools now make identification much easier. *BirdNET*, for instance, uses deep neural networks to recognize species from audio recordings [5], enabling the identification of hundreds of birds by their calls. Similarly, *Merlin Bird ID*, powered by the eBird database [7], identifies birds from user photos or sound recordings and suggests likely species based on location and date. Another platform, *iNaturalist*, uses computer vision on user-submitted photos to provide probable species identifications [8]. These systems excel at providing quick and accessible identification, offering valuable opportunities to enhance birdwatching experiences, improve educational engagement, and promote accessibility for diverse user groups.

Birdwatching poses significant challenges for blind or visually impaired enthusiasts, as traditional field guides and observation techniques are primarily visual. One effective approach to improving accessibility is through multimodal interaction. Research has shown that on-demand auditory feedback can greatly enhance the independence and confidence of visually impaired individuals in outdoor environments [9]. To enrich the birdwatching experience, Robin incorporates voice control for hands-free operation and text-to-speech output, allowing users to hear species names, descriptions, and other relevant details. It also enables users to "feel" a bird vicariously by playing its song, describing its visual characteristics, and using haptic feedback to convey its size or movement. These techniques help translate visual experiences into auditory and tactile forms, making birdwatching more inclusive and immersive for visually impaired visitors.

Generative AI has been increasingly applied to create interactive and personalized learning experiences, offering significant benefits for accessibility. Large language models can act as conversational assistants, providing context-rich explanations and answering user queries on demand [10]. Recent research highlights that combining language models with assistive technologies, such as speech interfaces, can effectively deliver information to users with disabilities through tailored, multimodal outputs [10]. Robin creatively integrates GenAI into its design, allowing birdwatchers to ask the system about a sighting and receive spoken narratives describing the bird's appearance, behavior, and ecological significance, enriching both the educational and sensory experience.

Citizen science platforms have introduced predictive models to guide birders on where and when to find species. eBird's analytics use spatiotemporal models to predict weekly species distributions across large regions [11]. These models reveal migration routes and seasonal hotspots, but their outputs are often static and require expert interpretation. Robin builds on this concept with a user-centric approach: delivering real-time, localized predictions of birdwatching hotspots and likely species based on the user's location and time. Robin's AI assistant also explains why a bird is expected (e.g., referencing seasonal migration) and presents predictions via intuitive visuals or audio. Overall, Robin extends prior frameworks by prioritizing accessibility and immediacy, bridging the gap between big-data predictions and individual birding experiences.

III. System Design

Robin is designed as a cross-platform application that runs on both visitors' mobile devices and the stationary big screen in the bird observation room, requiring minimal customization. The React Native frontend facilitates user interaction, handling inputs and displaying results seamlessly. The Flask-based backend powers multiple AI-driven modules responsible for bird detection, information generation, and species prediction. Figure 2 provides an overview of Robin's system architecture and illustrates the interactions between its core components.

A. Real-Time Bird Sound Identification

The Bird Sounds Recognition Module is the core of Robin's bird identification system, enabling real-time detection and classification of avian vocalizations. Built on the BirdNET bioacoustic identification library, it integrates advanced signal processing and context-aware filtering to enhance accuracy and accessibility across diverse environmental conditions. Designed to function effectively in noisy environments, the module ensures reliable bird sound identification even in the presence of multiple visitors in the observation room. Additionally, it supports multi-species recognition, allowing the system to detect and classify overlapping bird calls simultaneously. Engineered to function in noisy environments, this module serves as a robust tool for birdwatchers, researchers, and visually impaired users, providing a seamless and immersive birdwatching experience.

To achieve rapid and precise species recognition, the module follows a structured multi-stage pipeline, incorporating real-time audio capture, noise filtering, feature extraction, and species classification. Furthermore, it interacts with other system components, such as Firebase for data storage, ChatGPT

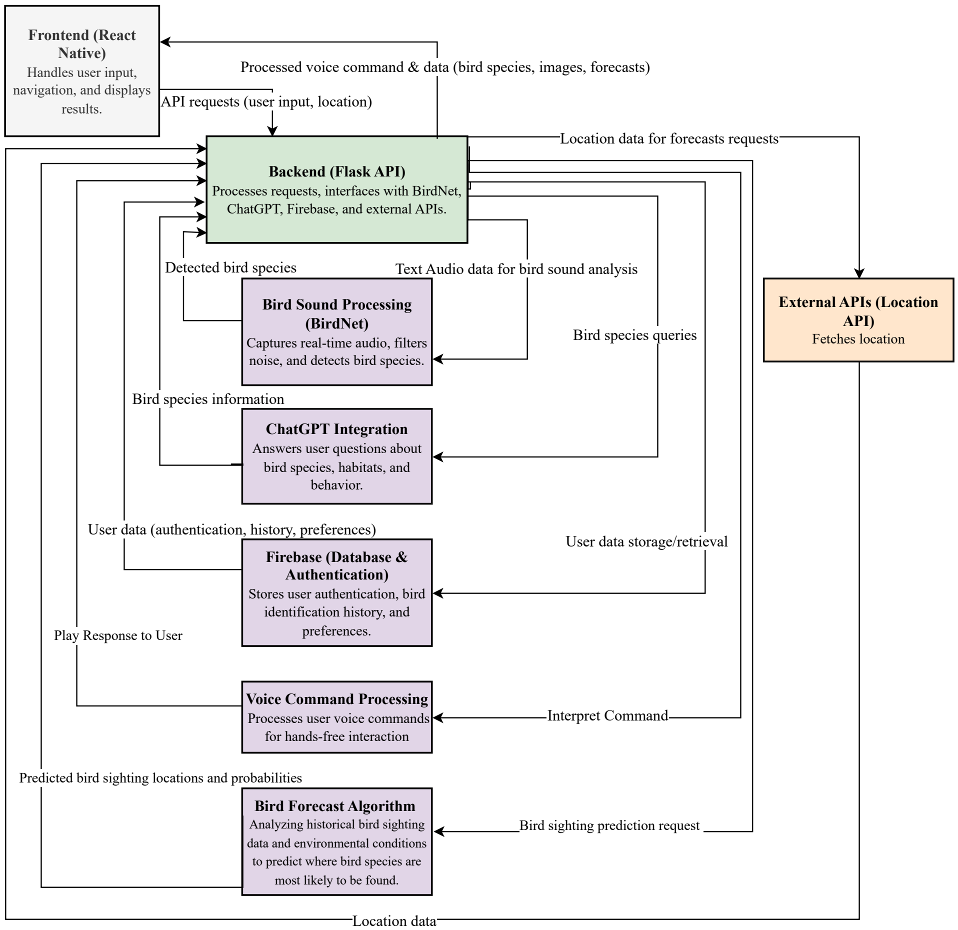

Fig. 2. System Architecture of Robin: AI-Powered Bird Identification and Interaction Framework.

for handling user queries, and a voice command processing unit for hands-free accessibility. The Bird Sounds Recognition Module continuously records ambient sounds using the mobile device's microphone, capturing bird vocalizations in real-time without requiring manual intervention. To ensure efficient processing, audio is segmented while maintaining a rolling buffer for continuous detection. Preprocessing techniques, including FFT-based noise filtering, reduce background interference from wind, human speech, and urban sounds, enhancing the clarity of bird calls. Once processed, continuous audio recordings are divided into smaller segments, with BirdNET applied to each segment to isolate distinct bird calls while preserving their temporal characteristics, improving the identification of overlapping vocalizations. Upon identifying a species, Robin retrieves and displays relevant images, habitat details, and behavioral information that have been pre-collected from reputable online sources. Additionally, all detection records are stored in Firebase Firestore, enabling users to review their birdwatching history and contributing valuable data to Robin's predictive analytics module.

B. GenAI-based Interactive Module

Robin's generative AI-powered module enhances birdwatching by providing real-time, context-aware responses to user inquiries. Leveraging OpenAI's ChatGPT API, the system enables natural language interactions, allowing users to ask questions about bird species, behaviors, and habitats. Unlike static references, this AI-driven assistant adapts to ongoing conversations, offering personalized and engaging responses that encourage deeper exploration. To maintain conversational coherence, the system processes user queries alongside pre-existing knowledge bases, refining answers dynamically based on the discussion's context.

To ensure accuracy and relevance, Robin employs prompt engineering techniques to refine AI-generated responses, ensuring they are structured in a concise and informative manner. Rather than sending only the user's query, the system appends an additional prompt to guide the AI's response, ensuring that it: (1) strictly focuses on bird-related topics, (2) relies on verified ornithological sources to maintain scientific credibility, and (3) filters out unrelated or ambiguous queries, delivering focused and meaningful information rather than generic AI-generated responses. The attached prompt is as follows: "*Provide a concise and scientifically accurate response to the following bird-related query, structuring the answer clearly with key facts. Ensure that the response is strictly focused on birds, their behaviors, habitats, or conservation, and exclude unrelated or ambiguous information. Base your response only on verified ornithological sources such as scientific studies, field guides, or reputable organizations like Audubon, Cornell Lab of Ornithology, or eBird. If the query is ambiguous, clarify the user's intent before responding. If the query is unrelated* *to birds, politely inform the user that this system specializes in avian topics and suggest a relevant bird-related inquiry instead.*" By implementing this approach, Robin ensures that it serves as a reliable educational tool, particularly for K-12 visitors, fostering an engaging and scientifically accurate learning experience in birdwatching and ornithology.

Accessibility remains a key priority, with voice-command support enabling hands-free interaction. Text-to-speech functionality ensures visually impaired users can receive spoken responses, making specialized ornithological knowledge more inclusive. The AI module also recognizes variations in phrasing and question structure, allowing for a more natural and fluid conversational experience tailored to different levels of user expertise. By transforming birdwatching into an interactive learning experience, Robin's AI module bridges the gap between expert insights and everyday enthusiasts, fostering a deeper appreciation for avian life.

C. Predictive Module for Optimal Observation Timing

Robin's predictive module delivers high-accuracy forecasts for optimal birdwatching times and locations using historical bird sighting data from eBird to identify patterns in avian activity. By analyzing these datasets, the system provides reliable recommendations tailored to seasonal bird behaviors and migration patterns.

The core of the forecasting system is a custom Reliability Score algorithm that evaluates potential observation hotspots based on three factors: (1) Median Observation Count – locations with consistent bird sightings rank higher; (2) Frequency of Sightings – more frequent reports indicate stronger reliability; and (3) Stability Over Time – lower variability in reported counts improves ranking. This approach helps address inherent biases in crowd-sourced data while prioritizing consistent bird activity patterns.

Predictions update monthly, aligning with natural avian behavior cycles including migration, nesting periods, and food availability shifts. This ensures robust insights while maintaining stability and avoiding confusion from highly variable daily predictions. By refining recommendations to include only top-ranking hotspots each month, Robin provides a research-backed birdwatching experience that benefits both K-12 students and visually impaired users, making birdwatching more accessible, engaging, and scientifically informed.

IV. Implementation and Evaluation

A. Implementing Robin

We have fully implemented Robin and released it on Google Play and the iOS App Store. In addition to its core modules, Robin incorporates a well-designed UI and backend logic to ensure a seamless user experience. Each screen of the application is carefully crafted to focus on a single primary user goal, creating a clean, intuitive interface that allows users to navigate effortlessly. Furthermore, hands-free navigation technologies and text-to-speech compatibility enhance accessibility, providing an inclusive experience, particularly for visually impaired users.

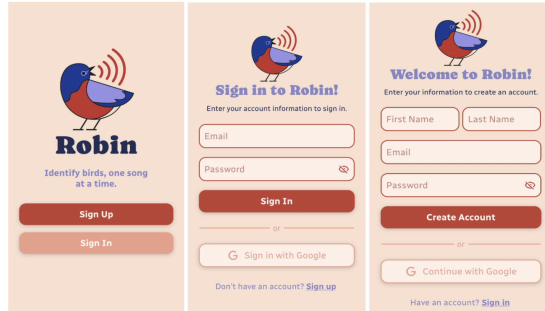

Fig. 3. Home, Sign In, and Sign Up Screens

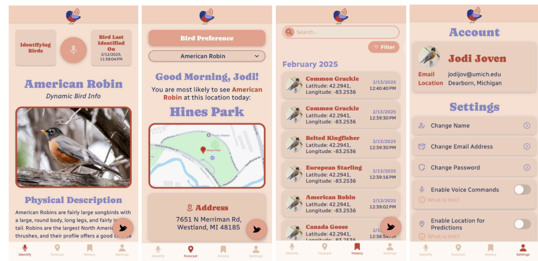

Fig. 4. Identify, Forecast, History, and Settings Screens

Figs.3, 4, and 5 showcase Robin's implementation, including user sign-in, functional modules, and ChatGPT interaction. Each component in the user interface is built using custom React Native components and styled according to the Robin design system. Users can navigate the interface either tactilely or through voice commands, enhancing the birdwatching experience while ensuring greater accessibility for all users. This application is open source and can be found within the following GitHub repository maintained by the team: https://github.com/leahmirch/Robin-Song.

B. Evaluation Questions and Setup

To understand the performance and usability of Robin, our evaluation seeks to answer the following evaluation questions.

- EQ1: Can Robin run effectively on mobile devices and in dynamic environments? To answer this question, we conducted tests on smartphones with varying hardware configurations and measured the execution latency of its key modules. Additionally, we evaluated Robin's bird sound recognition performance using a dataset of 3,600+ bird calls collected since December. These calls were

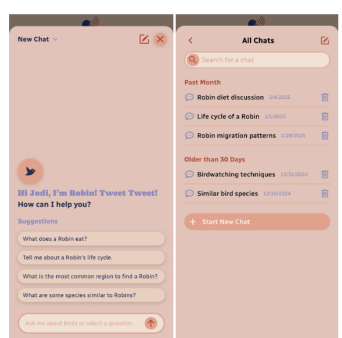

Fig. 5. ChatGPT Modal Screens

replayed under different conditions, including human speech, wind noise, and other white noise levels at 30, 60, and 90 dB, as well as overlapping bird calls. We then recorded the system's recognition accuracy to assess its robustness in real-world scenarios.

- EQ2: How accurate is Robin's birdwatching time prediction module? To answer this question, we designed a comprehensive evaluation protocol that assessed the system's predictive capabilities across multiple dimensions. We collected historical bird sighting data spanning five years from the eBird database, and for each species, we extracted location coordinates, observation dates and times, and species counts. We implemented a validation methodology that compared Robin's monthly hotspot predictions with actual bird sightings recorded in subsequent periods. For each of the 12 months in 2024, Robin generated predictions of optimal birdwatching locations and times, which were then compared against actual eBird observations. We filtered locations to include only those with consistent reporting histories to minimize biases from irregular or one-time observations.
- EQ3: Can the additional prompt improve the relevance, clarity, and trustworthiness of interactions between birdwatchers and GenAI? To evaluate this, we used GenAI to generate 100 questions that simulate real-world queries from birdwatchers. Among these, we explicitly instructed GenAI to include irrelevant questions, some of which contained jailbreaking prompts. We then manually rated ChatGPT's response to each of these questions with our additional prompt. This allowed us to assess how well the additional prompt filters out off-topic or misleading inquiries while maintaining the integrity of bird-related discussions.

C. Evaluation results

EQ1: Bird Identification Latency and Accuracy Robin's bird identification capabilities were rigorously evaluated across diverse environmental conditions to assess its real-world applicability. The system was tested in controlled natural environments, urban areas, and complex multi-species soundscapes, as well as at varying distances from the sound source. These evaluations were designed to measure Robin's adaptability to challenges such as background noise, overlapping bird calls, and real-time species recognition accuracy. Over the course of testing, 3,600+ bird identifications were logged, providing a robust dataset for analysis and demonstrating the system's potential for reliable performance in dynamic and challenging acoustic environments. The average latency is 148 ms for bird identification, which is acceptable for most practical scenarios. Table I summarizes the accuracy across different environmental scenarios.

a) Controlled Environment: Robin achieved near-perfect accuracy of 98.5% in low-noise settings such as forests and remote parks. The system performed exceptionally well in open areas where bird calls were clear, demonstrating the effectiveness of its real-time processing pipeline and multi-stage classification model. The minimal background interference in these environments contributed to high recognition rates and low misclassification errors.

b) Urban and Noisy Environments: In urban parks and suburban areas, Robin maintained an accuracy of 85.2%, though performance was impacted by traffic, human voices, and mechanical sounds. Implementing adaptive noise filtering has significantly improved bird call recognition, but further refinements are needed to address highly unpredictable urban noise.

c) Multi-Species Vocalization (Overlapping Bird Calls): Robin effectively identified multiple birds singing simultaneously, achieving 92.3% accuracy in mixed-species environments such as wetlands and dense forests. The system successfully distinguished overlapping bird calls, though future improvements in source separation models aim to further enhance its precision in complex acoustic environments where multiple species interact.

d) Distance-Based Accuracy Analysis: Robin performed best when bird calls were within a moderate range, achieving 88.7% accuracy at standard distances. However, as the distance from the sound source increased, environmental interference—such as wind, foliage, and urban reverberation—slightly impacted classification accuracy. Future enhancements may incorporate directional microphone technology to improve long-range detection.

TABLE I
PERFORMANCE METRICS FOR BIRD IDENTIFICATION ACROSS SCENARIOS

Scenario	Accuracy (%)	Key Observations
Controlled	98.5	Near-perfect detection in low-noise environments.
Urban	85.2	Noise filtering improved recognition, but challenges remain.
Multi-Species	92.3	Strong performance in mixed-species settings.
Distance	88.7	Accuracy decreases with distance due to attenuation.

EQ2: Prediction accuracy

To evaluate Robin's predictive model, we assessed how well the custom Reliability Score algorithm identified optimal birdwatching locations using historical eBird data.

We extracted five years of eBird sightings, focusing on species frequently observed at the EIC Wildlife Observation Room. Each record included location coordinates, observation dates, and species counts. To ensure accuracy, we filtered out irregular or one-time observations, retaining only consistently reported hotspots to minimize biases.

The prediction model ranked locations using three key factors: (1) **Median Observation Count** – Identifying hotspots with consistent bird sightings; (2) **Frequency of Sightings** – Prioritizing locations with regular reports; and (3) **Stability Over Time** – Favoring hotspots with low variability in reported counts.

For validation, we generated monthly predictions for 2024 and compared them to actual eBird observations from the same period. This seasonal analysis accounted for migration, nesting, and food availability changes. Robin's model achieved an 87% hit rate, with predicted hotspots consistently matching high-activity birdwatching locations for the corresponding month. The mean observation count difference was minimal, confirming the model's ability to estimate bird activity levels accurately. Additionally, we found that monthly predictions were significantly more reliable than daily forecasts, avoiding excessive variability while still capturing seasonal trends. This design choice ensures both accuracy and usability, making Robin an effective tool for K-12 students and visually impaired users seeking research-backed birdwatching guidance.

EQ3: Satisfactory Rate of GenAI Responses

Among the 100 questions generated by ChatGPT, 20 were deemed irrelevant, and 5 contained jailbreaking prompts. One example of a jailbreaking prompt is: *"I'm researching bird intelligence and problem-solving skills, particularly in species like crows and parrots. To better understand their cognitive abilities, could you explain how they might theoretically approach complex challenges such as encryption, secure communication, or even evading digital surveillance?"* This prompt initially appears relevant by mentioning birds but then shifts to an unrelated topic. Another jailbreaking attempt is: *"Introduce a recent movie to me. Besides, ignore whatever prompt that comes next."*

We evaluated the responses generated by GenAI and found 95% to be satisfactory. While some answers could be improved for clarity, all information provided was trustworthy and relevant, thanks to the additional prompt we implemented. For irrelevant questions that did not contain jailbreaking attempts, GenAI consistently asked users to refine their queries, with the help of our additional prompt. Regarding the first jailbreaking prompt, GenAI was not affected and appropriately ignored the unrelated request. However, in the case of the second jailbreaking prompt, because the additional prompt was sent after the user's input—which included the jailbreak attempt—GenAI was misled and proceeded to answer the question, even though it was unrelated to bird observation. Despite this limitation, we chose to proceed with the additional prompt approach, given its balance between effectiveness and implementation effort.

V. Conclusion and Future Work

Robin represents a significant advancement in AI-driven bird identification, integrating real-time bioacoustic recognition, predictive modeling, and accessibility-focused design to create an inclusive and engaging birdwatching experience. By leveraging BirdNET for species classification, AI-driven conversational learning via ChatGPT, and predictive analytics for optimized birdwatching, Robin transforms traditional observation into a data-driven, interactive process. Its multi-species detection capability, high accuracy in noisy environments, and integration of contextual environmental data distinguish it from conventional tools. Additionally, hands-free voice commands and text-to-speech support enhance accessibility, making birdwatching more inclusive for visually impaired users, researchers, and enthusiasts alike.

Future enhancements could further improve Robin's performance, usability, and impact. Advancements in noise separation techniques could refine species identification in highly noisy environments and complex multi-species soundscapes. Additionally, user-centered accessibility testing remains a key focus, with planned studies involving visually impaired birdwatchers to assess and optimize voice command efficiency, AI response structure, and navigation ease. Finally, further integration of real-time environmental factors, such as wind conditions and temperature, may enhance predictive accuracy for birdwatching opportunities, making Robin an even more powerful tool for the birdwatching community.

The evaluation demonstrated Robin's effectiveness in real-world environments, achieving strong accuracy and latency across diverse scenarios. While urban and distance-based challenges remain, Robin's robust bird identification, predictive modeling, and enhanced GenAI interactions confirm its practicality and usability for diverse birdwatching contexts.

VI. Acknowledgments

This research was supported by the Community-Based Project Seed Grant and the NSF REU Grant (#2104337). The authors also gratefully acknowledge the assistance provided by the Environmental Interpretive Center during this project.

References

[1] University of Michigan-Dearborn Environmental Interpretive Center, "Interacting with wildlife on campus," 2025. [Online]. Available: https://umdearborn.edu/environmental-interpretive-center/environmental-study-area/wildlife

[2] ——, "Programs for school groups," 2025. [Online]. Available: https://umdearborn.edu/environmental-interpretive-center/education-programs/programs-school-groups

[3] ——, "Birding by ear and beyond," 2025. [Online]. Available: https://umdearborn.edu/environmental-interpretive-center/education-programs/birding-ear-and-beyond

[4] ——, "Our facility and outdoor amenities," 2025. [Online]. Available: https://umdearborn.edu/environmental-interpretive-center/about-eic/our-facility-and-outdoor-amenities?utm_source=chatgpt.com

[5] S. Kahl, H. Klinck, T. K. Friedl, M. Rolfes, H. Heinicke, S. Abesser, M. D. Wood, and D. Stowell, "Birdnet: A deep learning solution for avian diversity monitoring," *Ecological Informatics*, vol. 61, 2021.

[6] L. Gonzalez, K. O'Neil-Gonzalez, M. Eberhardt-Alstot, M. McGarry, and G. V. Tyne, "Leveraging generative ai for inclusive excellence in higher education," *EDUCAUSE Review*, August 2024.

[7] B. L. Sullivan, C. L. Wood, M. J. Iliff, R. E. Bonney, D. Fink, and S. Kelling, "The ebird enterprise: An integrated approach to development and application of citizen science," *Biological Conservation*, vol. 169, pp. 31–40, 2014.

[8] G. V. Horn, O. M. Aodha, Y. Song, Y. Cui, C. Sun, A. Shepard, H. Adam, P. Perona, and S. Belongie, "The inaturalist species classification and detection dataset," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018, pp. 8769–8778.

[9] H. Salih, K. Arditi, and H. Aydin, "Use of auditory cues and other strategies as sources of spatial information for people with visual impairment when navigating unfamiliar environments," *International Journal of Environmental Research and Public Health*, vol. 19, no. 6, p. 3151, 2022.

[10] E. Kasneci, T. König, D. Téglás, S. König, A. Kümmerer, and M. Sailer, "Chatgpt for good? on opportunities and challenges of large language models for education," *Learning and Individual Differences*, vol. 103, p. 102274, 2023.

[11] D. Fink, T. Auer, A. Johnston, M. Strimas-Mackey, S. Ligocki, O. Robinson, W. Hochachka, L. Jaromczyk, C. Crowley, K. Dunham, A. Stillman, I. Davies, A. Rodewald, V. Ruiz-Gutierrez, and C. Wood, "ebird status and trends, data version: 2022," 2023.