



OPEN ACCESS

EDITED BY

Achaz von Hardenberg,
University of Chester, United Kingdom

REVIEWED BY

Bushuyev Sergey,
Kyiv National University of Construction and
Architecture, Ukraine

*CORRESPONDENCE

Divya Lakshmi S.

✉ divyabalu19@gmail.com

RECEIVED 07 October 2025

REVISED 27 November 2025

ACCEPTED 08 December 2025

PUBLISHED 09 January 2026

CITATION

S. DL and Kumar NS (2026) Cognitive alignment as a pathway to collaborative environmental sound AI in ecological monitoring. *Front. Ecol. Evol.* 13:1720295. doi: 10.3389/fevo.2025.1720295

COPYRIGHT

© 2026 S. and Kumar. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](#). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Cognitive alignment as a pathway to collaborative environmental sound AI in ecological monitoring

Divya Lakshmi S.^{1,2*} and N. Suresh Kumar³

¹Department of Computer Science and Engineering, Kalasalingam Academy of Research and Education, Krishnankoil, Tamil Nadu, India, ²Department of Computer Applications, Marian College Kuttikkanam Autonomous, Idukki, India, ³Computer Science and Engineering, Kalasalingam Academy of Research and Education, Krishnankoil, Tamil Nadu, India

KEYWORDS

cognitive alignment, co-listening AI, ecological monitoring, environmental sound classification (ESC), human-AI collaboration

Introduction

As global biodiversity declines, continuous acoustic monitoring has emerged as a non-invasive and scalable approach to track ecological change across landscapes. By capturing and analysing the sounds of wildlife, weather, and human activity, ecologists can gain real-time insight into ecosystem health and species dynamics. Yet, while artificial intelligence (AI) has accelerated the detection and classification of environmental sounds, it often lacks the interpretive sensitivity required for ecological decision-making.

Artificial intelligence has become an indispensable tool in ecology, reshaping how scientists detect, classify, and interpret environmental sounds. From bioacoustics sensors tracking biodiversity to deep-learning systems monitoring urban noise, environmental sound classification (ESC) technologies have expanded our capacity to “hear” the living world (Sharma et al., 2022). However, as these technologies advance, a growing disconnect has emerged between what machines detect and what ecologists understand.

The majority of ESC models are still optimized for performance measures—F1-scores, precision, and recall—instead of for interpretability, context, or ecological relevance (Haider et al., 2023; Rasmussen et al., 2024). Such systems recognize statistical patterns extremely well, yet they are often nothing more than closed, opaque black boxes, unrelated to the thinking that underlies ecological interpretation. A model should be able to recognize a bird call properly yet cannot produce its ecological meaning—whether it announces mating behaviour, territorial behaviour, or stress of the environment (Kohlberg et al., 2024). Without interpretability, Eco physiologically derived by AI risks being scientifically proper but Eco physiologically superficial.

This article is in favour of moving from an automated approach to “co-listening” through cognitive alignment - creating AI systems that are capable of listening with ecologists, rather than simply for them. Cognitive alignment describes the similarity between models’ internal representations and explanations and those of human ecological cognition (Kvsn et al., 2020). Aligned systems must allow for mutual

intelligibility - the ability of humans and AI to provide each other with interpretive processes, feedback and co-adaptive learning by elapsed time. The next sections outline the current limits on ESC, provide a conceptual framework for cognitive alignment and present paths toward design of “co-listening” systems that can combine human and machine understanding to create an understanding of ecology.

Cognitive alignment does not imply that ecological interpretation can be fully reduced to explicit rules. Rather, it requires anchoring AI representations in ecologically meaningful latent concepts—such as species traits, call types, behavioural contexts, and habitat-level acoustic indices—so that the model’s internal structures correspond to how ecologists reason about sound. This process can be enabled through interactive concept-refinement interfaces that allow experts to promote, demote, merge, or redefine ecological concepts inside the model. In this way, cognitive alignment becomes a pathway for shared interpretive grounding rather than simply a visualisation of hidden layers.

The misalignment problem

Over the past decade, ESC systems have achieved remarkable technical progress. Early approaches relied on engineered features such as mel-frequency cepstral coefficients (MFCCs) and classifiers like random forests or support vector machines (Toffa and Mignotte, 2021). Deep-learning architectures—convolutional, recurrent, and transformer-based—now dominate the field, delivering state-of-the-art results on datasets like ESC-50, UrbanSound8K, and DCASE (Jahangir et al., 2023).

Despite these advances, most architectures remain opaque. Their decision processes are difficult to interpret, and *post-hoc* visualization methods such as Grad-CAM yield insights that are limited or ecologically irrelevant. *Context collapse* further occurs when isolated audio clips are analysed without temporal, behavioural, or environmental context (Zinemanas et al., 2021).

Distributional drift compounds this issue: ecological soundscapes evolve with seasons, habitats, and weather (Patchipala, 2023), leading to poor model generalization beyond training conditions. Human–AI interaction is similarly one-sided—ecologists provide annotations for training, yet deployed systems rarely accept ongoing feedback or correction.

Consequently, a cognitive gap persists. Models “hear” statistically while ecologists “listen” contextually. Without shared interpretive grounding, predictions may be accurate yet cognitively alien, eroding trust and limiting ecological understanding.

Toward cognitive alignment

Cognitive alignment offers a conceptual as well as practical answer to the misalignment that is being observed. It refers to the extent to which the reasoning, representations, and uncertainty estimates of an artificial intelligence system have been made

consistent with ecological interpretive models (Rane et al., 2024). A coherent model of an ecological soundscape classifier must be able to go beyond mere categorisation of acoustic phenomena and give a description of the rationale behind it in a way that can be understood by experts in the domain.

The efficacious communication thus requires a restructuring of the automatization of the classification process involved in ecological soundscape classifiers (ESC) into a collaborative interpretive process, which involves a mutually co-operative co-listening activity between machine and human.

1. **Soundscape Contribution:** Raw environmental records are accompanied by contextual metadata of variables in terms of time, place, habitat and weather conditions.
2. **Representation of Models:** The Acoustic features are changed, resulting in probabilistic forecasts of uncertainty based on quantitative features.
3. **Ecologically valid Interface:** The utility visualisations of the model are presented in the form of ecologically valid representations of ecologically valid activation or attention-map representations that enhances the interpretation of the visuals.
4. **Human Feedback:** Ecologists analyse the output of the model, discover invalid classifications, and record arguments like redundant vocalization.
5. **Model Adaptation:** Active or incremental learning to adapt the system with elite information is influencing the system to predict better.
6. **Iteration:** Due to the process of co-adaptation between the human reviewers and the AI model, the framework will converge to a more precise and common interpretive alignment.

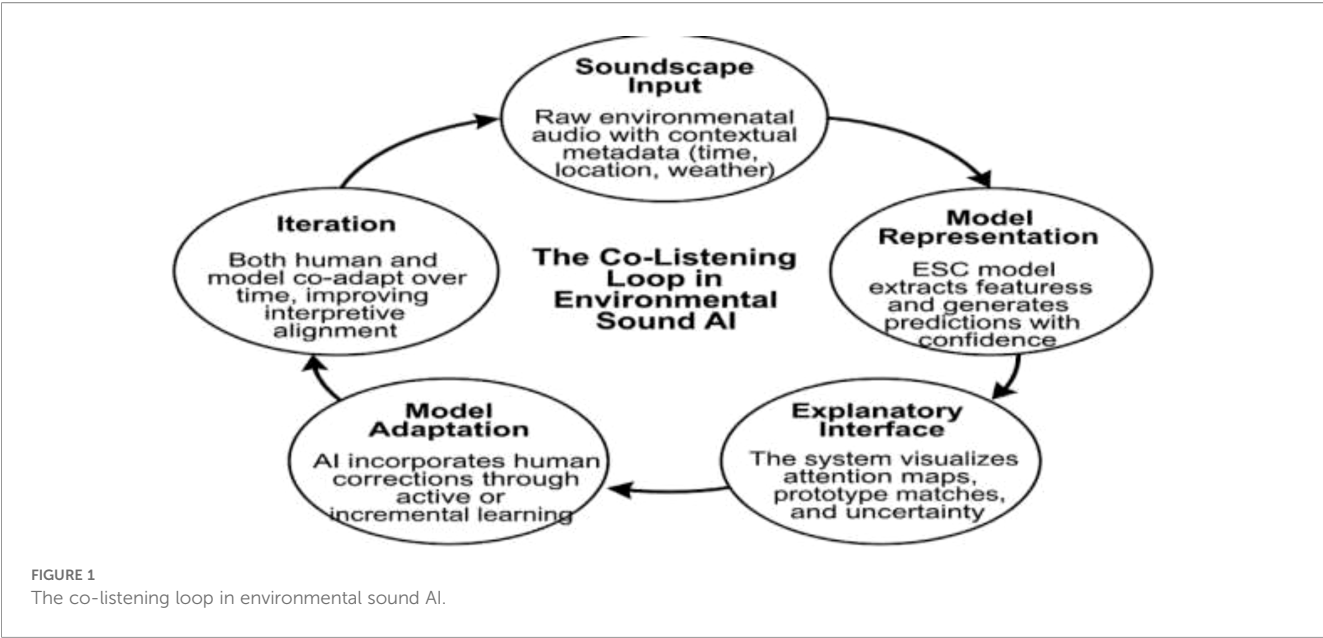
This cycle recasts ESC as an interpretive co-operation instead of a pipeline, allowing AI to engage ecological reasoning through contextual awareness and clear feedback.

Design of paths for cognitive correspondence

Cognitively harmonized environmental sound classification (ESC) systems are designed in such a way that a systematic refactoring of representation, context, feedback and uncertainty are required to ensure ecologically interpretable results.

Interpretability: The ecological partitions: the latent representations should capture ecological partitions such as species traits, types of calls or acoustic indices. Internal reasonability can be achieved through prototype-based and concept-bottleneck architectures (Zheng et al., 2025; Cheng et al., 2025) and can generate more realistic species-occurrence data and increase trust in the biodiversity estimate.

Context awareness: Environment (e.g. weather conditions and habitat properties) and time (e.g. diel cycles) metadata should be



included in models so as to form relationships between acoustic patterns and underlying ecology, constituting long-term habitat and biodiversity monitoring.

To strengthen context awareness in ESC systems, multimodal fusion architectures should be used to integrate ecological metadata with acoustic features. Early fusion models concatenate spectrogram-derived embeddings with structured variables such as time-of-day, weather indices, or habitat descriptors before entering a shared encoder. Late-fusion approaches process acoustic and contextual information in parallel streams and combine their latent representations for joint inference. Cross-modal attention mechanisms further allow contextual variables to dynamically weight acoustic features, supporting ecologically coherent representations within the model.

TABLE 1 Current limitations in environmental sound classification and corresponding cognitive alignment strategies for ecological insight.

Dimension	Current limitation	Alignment strategy	Expected benefit
Interpretability	Opaque decision processes	Prototype/concept bottleneck models	Transparent ecological reasoning
Context Awareness	Isolated sound clips	Integrate temporal and environmental metadata	Context-aware classification
Human Feedback	One-way training process	Co-listening loops with expert review	Adaptive, continuously improving models
Uncertainty	Overconfident predictions	Bayesian/evidential inference	Communicated ambiguity and trust
Domain Adaptation	Poor generalization	Continual learning with drift detection	Sustained performance across habitats

Human feedback Human-in-the-loop learning (Retzlaff et al., 2024) combines predictive output with conservation priorities, which allows models to optimize the detection of indicator species, and spend less resources on the annotation of rare species.

Uncertainty and adaptation: Calibrated confidence estimates can be provided based on Bayesian or evidential frameworks (Zhuo et al., 2023) when soundscapes are changing. Adaptive sampling is informed by the transparent quantification of uncertainty, and model robustness is maintained in dynamic ecosystems through incessant learning.

Discussion

The way scientists use AI vs. old-school supervisory techniques is transformed by Cognitive Alignment and how they interact with Ecologists. The creation of a Collaborative Experience for humans and AI to build and utilize ESC Systems which can produce Real Time Ecological Interpretations and Iteratively Feedback during the Application Process occurs when Cognitive Alignment is integrated into an Environmental Sound Classification System. The creation of environmentally accurate, validated and scientific interpretation by Cognitively Aligned Models results from their ability to capture dynamically changing environmental conditions over time as well as behaviourally relevant information and contextual factors (e.g. Habitat Acoustic Diversity Levels, Specific Species Activity Cycles), thus allowing researchers to identify patterns and subsequently conduct biodiversity assessments and create effective conservation management strategies (McCrindle et al., 2021).

Transparency and interpretability from cognitively aligned systems will be able to increase the accuracy and reliability of the ecological modelling that occurs with the increased transparency and interpretability of system reasoning and uncertainty. The conservation data pipelines will then have increased trustworthiness due to their improved transparency and

interpretability. Additionally, by providing a transparent framework for reasoning, it is possible to allow ecologists to evaluate the limits of the model; this, in turn, provides an opportunity for the ecologist to validate the model in the field, as well as develop adaptive sampling strategies. As such, cognitively aligned systems are most beneficial for long-term monitoring projects and for developing management frameworks using reliable, science-based indicators of biodiversity.

Cognitive alignment is ethically beneficial as it distributes interpretation of data among many; AI enhances rather than replaces the professional expertise of a scientist, promotes an attitude of humility and ethical integrity for both researchers and policy makers of ecological and environmental work. Interpretability also enables collaboration between people who are outside of academics, enabling meaningful engagement by practitioners, policymakers and citizen science participants with the model outputs. In community-based monitoring, co-listening models enable the ability of local monitors to offer contextual feedback on the model and thus improve its applicability within the diversity of habitats and socio-ecological contexts (Figure 1).

Evaluating cognitive alignment remains an open challenge (Table 1). Potential indicators include overlap between human and model attention maps, expert correction rates, and qualitative satisfaction scores. Developing benchmark datasets annotated with expert rationales could provide measurable progress toward interpretive convergence.

To support transparent evaluation, we refine the Cognitive Alignment Score (CAS) into a modular benchmarking scheme with measurable indicators across four dimensions: (1) Representational alignment, quantified by metrics such as spatial overlap between expert-annotated and model-generated attention maps or rank correlations between reasoning traces; (2) Interpretive alignment, measured using explanation-validity rubrics and matches between predicted behavioural context and expert interpretations; (3) Adaptive alignment, captured by reductions in time-to-correction or decreases in repeated expert-flagged errors across feedback iterations; and (4) Uncertainty alignment, evaluated through calibration error and Brier scores relative to expert judgments of ambiguity. CAS provides a reproducible and extensible pathway for assessing whether humans and AI systems are progressively converging toward co-listening.

The next steps for future research will be to develop Open-Source Co-Listening Platforms, Develop Annotated Datasets with Context and to Perform Comparative Studies to Determine if Cognitively Aligned Systems Improve Ecological Inference Quality, Interpretability and Decision-Making Quality. These efforts will help make cognitive alignment an essential part of developing Responsible AI, Collaborative AI, and Ecologically Grounded AI.

A constructive outlook

Environmental Sound AI has reached a crossroads. Advances in technical capabilities allow for the identification of sounds within

the acoustic environment with greater sensitivity than ever before; however, they lack interpretive context which limits their ability to be used as valuable tools for understanding the environment. Cognitive alignment provides a method to develop systems which are able to collaboratively develop a shared understanding of the world through the combination of computational inference and ecological knowledge.

The long-term goal is not for AI to listen better than ecologists but to listen *with* them: to share the perceptual and cognitive work of understanding complex ecosystems. In doing so, AI becomes an interpretive partner that amplifies ecological reasoning and strengthens the foundations of conservation science.

Author contributions

DS: Writing – review & editing, Conceptualization, Writing – original draft. NK: Conceptualization, Writing – review & editing, Writing – original draft.

Funding

The author(s) declared that financial support was not received for this work and/or its publication.

Conflict of interest

The author(s) declared that this work was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

- Cheng, X., Niu, Z., Jiang, Z., and Li, L. (2025). Enhancing bottleneck concept learning in image classification. *Sensors (Basel Switzerland)* 25, 2398. doi: 10.3390/s25082398
- Haider, U., Hanif, M., Kobayashi, H., Parajuli, L., Shimotoku, D., Rashid, A., et al. (2023). "Bioacoustics signal classification using hybrid feature space with machine learning," in *2023 15th International Conference on Computer and Automation Engineering (ICCAE)* (Piscataway, NJ, USA: IEEE). 376–380. doi: 10.1109/ICCAE56788.2023.10111384
- Jahangir, R., Nauman, M. A., Alroobaea, R., Almotiri, J., Malik, M. M., and Alzahrani, S. (2023). Deep learning-based environmental sound classification using feature fusion and data enhancement. *Computers Materials Continua* 74, 1069–1091. doi: 10.32604/cmc.2023.032719
- Kohlberg, A., Myers, C., and Figueroa, L. (2024). From buzzes to bytes: A systematic review of automated bioacoustics models used to detect, classify and monitor insects. *J. Appl. Ecol.* 10, 1198–1209. doi: 10.1111/1365-2664.14630
- Kvsn, R. R., Montgomery, J., Garg, S., and Charleston, M. (2020). Bioacoustics data analysis – A taxonomy, survey and open challenges. *IEEE Access* 8, 57684–57708. doi: 10.1109/ACCESS.2020.2978547
- McCrindle, B., Zukotynski, K., Doyle, T., and Noseworthy, M. (2021). A radiology-focused review of predictive uncertainty for AI interpretability in computer-assisted segmentation. *Radiology. Artif. Intell.* 3, 1–368. doi: 10.1148/ryai.2021210031
- Patchipala, S. G. (2023). Tackling data and model drift in AI: Strategies for maintaining accuracy during ML model inference. *Int. J. Sci. Res. Arch.* doi: 10.30574/ijrsra.2023.10.2.0855
- Rane, S., Bruna, P., Sucholutsky, I., Kello, C., and Griffiths, T. (2024). Concept alignment. doi: 10.48550/arXiv.2401.08672
- Rasmussen, J., Stowell, D., and Briefer, E. (2024). Sound evidence for biodiversity monitoring. *Science* 385, 138–140. doi: 10.1126/science.adh2716
- Retzlaff, C., Das, S., Wayllace, C., Mousavi, P., Afshari, M., Yang, T., et al. (2024). Human-in-the-loop reinforcement learning: A survey and position on requirements, challenges, and opportunities. *J. Artif. Intell. Res.* 79, 359–415. doi: 10.1613/jair.1.15348
- Sharma, S., Sato, K., and Gautam, B. (2022). "Bioacoustics monitoring of wildlife using artificial intelligence: A methodological literature review," in *2022 International Conference on Networking and Network Applications (NaNA)* (Cham, Switzerland: Springer). 1–9. doi: 10.1109/NaNA56854.2022.00063
- Toffa, O. K., and Mignotte, M. (2021). Environmental sound classification using local binary pattern and audio features collaboration. *IEEE Trans. Multimedia* 23, 3978–3985. doi: 10.1109/TMM.2020.3035275
- Zheng, C., Miller, T., Bialkowski, A., Soyer, H., and Janda, M. (2025). Supporting data-frame dynamics in AI-assisted decision making. *ArXiv*. doi: 10.48550/arXiv.2504.15894
- Zhuo, J., Wang, S., and Huang, Q. (2023). Uncertainty modeling for robust domain adaptation under noisy environments. *IEEE Trans. Multimedia* 25, 6157–6170. doi: 10.1109/TMM.2022.3205457
- Zinemanas, P., Rocamora, M., Miron, M., Font, F., and Serra, X. (2021). An interpretable deep learning model for automatic sound classification. *Electronics* 10, 850. doi: 10.3390/electronics10070850