



OPEN ACCESS

This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Federal University of Rio Grande do Norte (UFRN), Natal, Rio Grande do Norte, Brazil

Correspondence to:

Divna Djokic,
divna.divna@yahoo.com

Additional material is published online only. To view please visit the journal online.

Cite this as: Djokic D. How is AI Helping Humans Understand the Vocal Communication of Other Species: Recent Advancements in the Use of AI for Acoustic Monitoring of Wildlife. Premier Journal of Artificial Intelligence 2025;3:100017

DOI: <https://doi.org/10.70389/PJAI.100017>

Received: 28 October 2024

Revised: 17 March 2025

Accepted: 1 June 2025

Published: 14 June 2025

Ethical approval: N/a

Consent: N/a

Funding: No industry funding

Conflicts of interest: None

Author contribution:

Divna Djokic –
Conceptualization, Writing –
original draft, review and editing

Guarantor: Divna Djokic

Provenance and peer-review:
Commissioned and externally
peer-reviewed

Data availability statement:
N/a

How is AI Helping Humans Understand the Vocal Communication of Other Species: Recent Advancements in the Use of AI for Acoustic Monitoring of Wildlife

Divna Djokic

ABSTRACT

Animal vocalization is an area of research with great potential. As a noninvasive method to study animals, bioacoustics can teach us a lot about vocalizing animals, in the different aspects of their life. The biggest issue of bioacoustics is handling large datasets of audio recordings (long as well as heavy). With the extensive growth of artificial intelligence (AI) in other fields, like image processing and speech recognition, with slight adjustments, we are now able to apply it in bioacoustics and successfully tackle some of its main issues. In this review paper, I am offering a short overview of AI development, the types of it in use today, and the general concept of its use, mostly concentrating on how it is applied in the area of bioacoustics. I will discuss the challenges and limitations of this approach and further offer several examples of studies that tried to tackle these particular issues. Finally, the review will be concluded with a discussion on ethical issues and challenges that follow the use of AI in the early stages of its development.

Keywords: Bioacoustics, Animal vocalization, Artificial intelligence, Machine learning, Deep learning

Methodology

The studies selected for this review were chosen based on standard literature screening criteria.¹ The research questions and objectives were derived from the title, subtitle, and keywords of the studies. Publications were included based on their authenticity, relevance, and contribution to the field. Additional criteria for inclusion involved considering key works from the earliest research in the topic area up to the most recent trends, as well as ensuring that the latest data was reviewed to create a comprehensive and up-to-date review. The selection of specific taxa and species for inclusion in the review was guided by the level of research attention given to them. Various databases such as ResearchGate, Google Scholar, ScienceDirect, and PubMed were utilized to gather relevant studies for this review.

Certain sections of this paper were drafted and proofread with the help of AI.

Introduction

The breakthrough of artificial intelligence (AI) in everyday human applications has sparked an interest in exploring new ways to apply this technology. Consequently, an interest in applying AI in conservation followed, particularly in subdisciplines such as animal communication.

Animals can send signals and communicate in various ways: visually, acoustically, tactilely, chemically,

exchanging electrical signals. More often than not, animals tend to combine different signal types, and sometimes even beyond human perceptual capabilities (e.g., ultrasounds).² To decode and better understand their behavior, good practice includes taking the entire context of the signal—including all the other signals the one of interest was expressed with—into account when trying to annotate it with potential meaning.³ There are several big challenges following research at this stage: gathering enough data, being able to distinguish separate signals from a noisy recording (so-called cocktail party problem, chorus or vocalizing animals, or extracting a voice from the background noise),⁴ being able to connect them back together when looking into the context, having each of them in focus (one at a time), and finally, making sure the least bias possible was implemented throughout the entire process.^{5,6}

Not being able to “see the forest for the trees”, a prudent approach was to take a step back and initially try to resolve each of the signal types separately, and only afterward putting them into context, and maybe one day, being able to understand their meaning.

Out of all the signal types we can expect animals to express, and considering their adaptations to the environment they live in, maybe the most universal one is the acoustic signal, across the animal kingdom, including us humans. Although some decades ago it was not considered of such importance—as our tools, knowledge, and the prevailing trends in scientific community were not favorable of this idea—now we know animals spend considerable time vocalizing, in various forms. A vivid example of how far humans have come in a short period in understanding the world around them is an example of humpback whales. These majestic, now we know, very vocal animals were once considered mute, for one specific reason: lack of vocal cords. As our knowledge, back in the day, was vastly anthropocentric (at least, more than today), whales’ lack of anatomical structures our species use to vocalize—vocal cords—inevitably led to the conclusion their species does not produce sounds. Although it took several decades to truly understand and describe how whales produce vocal cues,⁷ now we understand there is more than one way an animal can produce sounds and vocally communicate.

The research discipline looking into animal communication—bioacoustics—is rapidly growing in parallel with the technical development we are witnessing. Smaller, but more durable and with higher capacities, equipment used to record and store sounds are opening new doors for the scientists in this field. These technical advances dealt well with the challenges of

mounting recorders to animals (e.g., collars), being in their proximity (e.g., robotic bees), long-lasting batteries for autonomous recorders, or in-field storage capacity issues. Crossing these obstacles, new ones were already seen on the horizon: enormous data sets, storage of heavy acoustic recordings, and lack of trained workforce for labeling and validating.

One common process was able to tackle all the issues named above: digitization. In the present moment, the research field of bioacoustics has grown into an updated version of itself: digital/computational bioacoustics. More than just working with digital instead of analog audio signals, what makes it digital is its increasing reliance each day on automated processes, such as AI, machine learning (ML), and deep learning (DL) tools to process vocal recordings and try to understand them better.

What Is AI?

AI is considered a phenomenon when a machine or a computer simulates behavior or decision-making that of humans. This was the initial idea of the concept, which has its origins back in the 1940s.⁸ Today, there are plenty of publications talking about the development of this area of research (e.g.,^{8–10}). At present, thanks to the vast data libraries, this approach had a steep leap, overflowing to its branch of ML, which in turn further developed its core concept of DL and neural networks (constructed of neural nodes). These gave birth to models that are of great use today, such as image processing or natural language processing of AI (Figure 1).

What Is ML?

Bermant et al.¹¹ in their paper from 2019 dealt well with explaining the general issue of ML: “(ML) entails the design and develops self-learning algorithms that allow computers to evolve behaviors based on empirical data, such as from sensors and databases.” As they further explain, ML problems can be divided into three major subcategories: unsupervised learning, supervised learning, and reinforcement learning. Today, we can include a new one, recently developed self-supervised learning (SSL).¹⁰

ML subcategories:

- a. *Unsupervised learning*, as expected, relies on its own capabilities to discover the structure and

meaning in the dataset, about which the model often has no prior knowledge.¹²

- b. *Supervised learning*, on the other hand, is given an example—training dataset—based on which it will learn and further be able to predict the outcomes in the brand-new dataset.¹² Training datasets are comprised of pre-labeled data, with known meaning or values, based on which the model is “learning” ground values and what is expected of it.¹¹
- c. *SSL*, a newly developed model, relies on learning from the labels in the input dataset it generates itself. It is particularly convenient for natural language processing and computer vision.¹³
- d. *Reinforcement learning* involves mimicking the human training process, trial-and-error and experience-based learning, to train models on their own how to make decisions in a given environment. It is a way of learning from interaction.^{14,15}

Representation learning acts as a bridge between ML and DL. This approach includes a set of methods that prepare a computer to process raw data in a way that allows it to automatically understand representations needed to proceed with the detection or classification of data.

Now, the DL method, as a basic paradigm of ML, is a complex representation-learning method, composed of several levels of representations of nonlinear, but simple modules. DL leverages neural networks with many layers (hence deep) to process and learn from large amounts of data. With each level, the acquired information becomes more complex and abstract. In such an approach, the machine can comprehend different layers of important information and understand complex systems, while disregarding less important factors. Consequently, it learns discrimination, classification, and detection tasks. There are different types of neural networks, designed for different tasks^{10,13,16} (Figure 2).

Finally, DL is a piece of the puzzle scientists need to manipulate the language, in the form of the natural language processing discipline (Figure 3). Simply put, this approach allows machines to “read” or “understand” human text or speech. With additional tuning, bioacousticians are now able to apply these models to nonhuman vocal cues and train them to decipher whale, elephant, bee, or even plant vocal communication.^{17–20}

As Stowell⁵ explains, “Deep learning is flexible and can be applied to many different tasks, from classification/regression through to signal enhancement and even synthesis of new data.” Hence, there is a general “recipe” for applying DL in most subfields of bioacoustics, which was constructed by Stowell⁵ in the same publication, including literature available at the time (please consult that source for further details). The final observation is that computational bioacoustics is not a single task, but an umbrella covering multiple small-scale efforts needed for a complete depiction.²¹

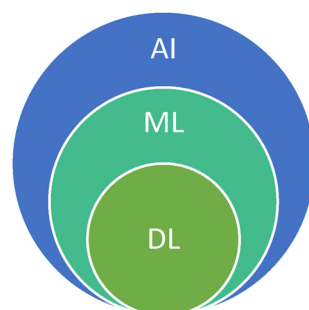


Fig 1 | Relations of AI, ML, and DL

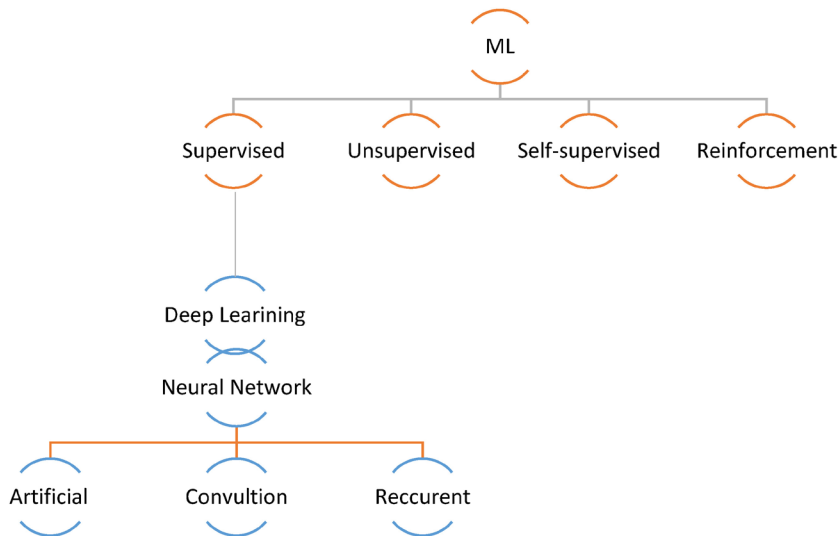


Fig 2 | Different approaches in applying ML and different types of neural networks for deep learning, as a branch of supervised learning

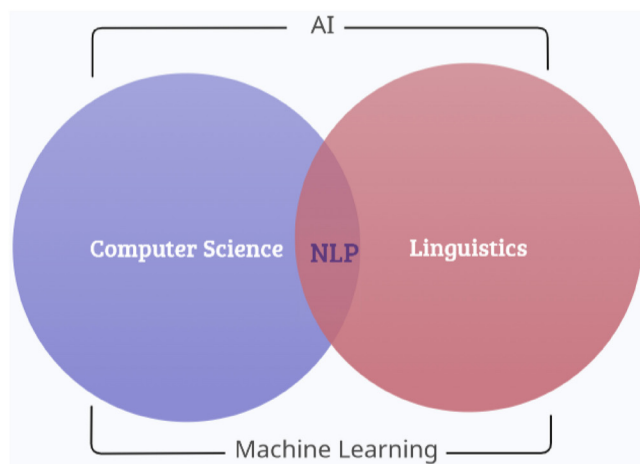


Fig 3 | Natural language processing

Potential of SSL Models in Bioacoustics

Instead of using specifically annotated instructions, SSL learns similarly to humans—based on experience. These models create a surrogate training set by defining it themselves. The model is fed with a large, unlabeled dataset, from which SSL picks up signals it can repeatedly recognize and sources metadata from them.²² In this sense, when applied to bioacoustics, SSL can be used to extract specific vocalizations, pinpoint the caller, determine the species, or even distinguish between different call types of the same species—such as alarm vs. mating calls.^{23–27}

As recently noted by Sarkar and Magimai-Doss,²⁸ pretrained models used for speech recognition showed an extraordinary capacity for application in animal bioacoustics. This finding was pivotal regarding data availability for model training. Human voice recordings, as a data source, are practically limitless, and so are the potential for model training. These authors

further observed that fine-tuning for a specific species, rather than humans, does not substantially improve results and model performance. Thus, on one hand, this is excellent news, as it means we can use speech recognition models trained on human voices and apply them directly to animal bioacoustics. On the other hand, it demonstrates the robustness of SSL and the comprehensive nature of the information it learns from—it recognizes one human from another and applies the same principles to distinguish one bird from another.

Key points of the section:

ML

- ML involves designing and developing self-learning algorithms that enable computers to evolve behaviors based on empirical data (e.g., from sensors and databases).

Subcategories of ML Unsupervised Learning

- Relies on its own capabilities to discover the structure and meaning in datasets.
- Model often has no prior knowledge of the data.

Supervised Learning

- Given a training dataset with prelabeled data.
- Learns to predict outcomes in new datasets based on the training data.
- Models learn ground values and expected outcomes from the labeled data.

SSL

- Newly developed model.
- Learns from labels in the input dataset that it creates itself.
- Useful for natural language processing and computer vision.

Reinforcement Learning

- Mimics human training processes (trial-and-error and experience-based learning).
- Models learn to make decisions in a given environment through interaction.

Representation Learning

- Acts as a bridge between ML and DL.
- Includes methods that prepare a computer to process raw data.
- Enables the automatic understanding of representations needed for data detection or classification.

DL

- Basic paradigm of ML with complex representation-learning methods.
- Composed of multiple levels of nonlinear representations.
- Leverages neural networks with many layers to process and learn from large amounts of data.
- Each level makes information more complex and abstract.

- Understands complex systems and learns discrimination, classification, and detection tasks.
- Different types of neural networks are designed for different tasks.

Natural Language Processing

- Aims to enable machines to “read” or “understand” human text or speech.
- Bioacousticians apply models to decipher nonhuman vocal communication (e.g., whale, elephant, bee, or plant vocal communication).

Stowell's Observations

- DL is flexible and can be applied to various tasks (classification, regression, signal enhancement, data synthesis).
- Computational bioacoustics encompasses multiple small-scale efforts needed for a complete depiction of vocal communication.

Challenges and Limitations

Although still in its development, general challenges applying AI to process vocal cues of animals are already visible, and some were anticipated. Most of them are of a technical nature, but as important are the ones dealing with data processing and interpretation, which is the main concern of this review.

Following the natural form of scientific research, deploying recording devices can, on its own, be challenging. Whether it is an autonomous recorder or a manual, this process is often very labor-intensive and costly. A good example of the influence of this element is the development of *bird acoustic* research, which is the most advanced one,²⁹ probably due to the ease of data collection.

Next to question is battery life and on-site storage capacity, to follow the in-lab storage. This directly influences the ability of data sharing, as for the largest data sets, only the physical sharing of recordings is an option, on an external hard disk or such.

The overarching issue is the question of funds. Recording equipment and the following accessories are pricey and require specific storage conditions. These directly influence certain parts of the world to be less explored, due to the economic constraints faced by countries governing those areas. More on this topic will be explored in the section on Ethics below.

Finally, once the data are in the lab, next to think about is the methodology of data processing.

Until roughly 2017, the only option for processing audio data was manual analysis. It is time-consuming, heavy manual work, that first requires rigorous training of individual researchers to become experienced with the vocalizations in question and prepared to tackle the challenges in data that could appear. Additionally, in some areas of research (e.g., humpback whale songs), there are no agreed methods by which sound units should be labeled (e.g., frequency range the label should contain), or even the vocalization structure

(hierarchical level) the analysis should focus on (unit, phrase, theme).³⁰ This further complicates the sharing of datasets for generalized AI model training.

For the same animal, variability in repertoire from year to year, and between populations is an additional challenge for labeling, but also for model training, as it is virtually impossible to grasp all the sounds the species uses.^{31–33} For this particular reason, options as unsupervised models for ML are a fantastic opportunity to tackle issues of large sound repertoires of certain species.

Undoubtedly, AI and ML opened great possibilities for bioacousticians, firstly reflected in the amount of data that could be processed much faster. Important to stress is that nowadays, for most of the models in use, initial training set data labeling still needs to be handled manually, plus, final validation of the model performance needs to be checked as well by hand.

Still, it is creating possibilities of vast prospects, identifying patterns and structures in vocalizations researchers were incapable of perceiving so far. One example of pattern recognition done without AI is the work of Malige et al. from 2021,³⁴ who visually represented the structure of the humpback whale song, and how it differs between individuals of the same stock. Nowadays, the same research could be done by applying DL in recurrence plots for multivariate nonlinear time series forecasting.³⁵

We can now identify general challenges expected when working with AI methods in bioacoustics, based on data, optimization, model, and evaluation (DOME).³⁶ In the section to follow, readers can find a few bright examples of novel research that tried to solve some of the issues, pawing the path toward more efficient and accessible AI applications in animal vocalization research.

AI processing challenges (DOME, as recommended by contemporary literature:^{2,5,36–40}

Data: Dataset quality and quantity → *Optimization:* Model's ability to generalize → *Model:* Model complexity → *Evaluation:* Lack of standardized metrics

AI-Enabled Discoveries

We can track the intense boost of computational bioacoustics only in the last 7 or 8 years.^{5,41} As such a young discipline, rather than providing a historical overview, I would like to offer several examples of research promising to open doors of human understanding of natural communication, wider than earlier expected. These studies cover different groups of organisms, but what they have in common is their importance in exposing holes in our knowledge of understanding the nature of vocal communication, whether of sending or receiving information. Moreover, these works shed light on a long road in front, giving a glimpse of the great potential of AI applications in bioacoustics. The following examples were picked particularly because of the way they tackled issues of using AI (DOME) in bioacoustics, as drafted earlier.

Data Quality: How to (Conveniently) Train Your Model²¹

One of the greatest challenges in applying AI in bioacoustics is, on one hand, the enormous amount of good-quality data needed to train the model, or the other, extensive manual labor needed to label the training data set. The following research offered a solution to this problem and tested their approach on several different animal species.

In particular, Nolasco et al.²¹ offer a practical solution to train a single system on several different datasets. In those, different sounds are labeled only on a single category: start and end time. The system is trained on an audio file using only the first five occurrences of the target sound event. This approach is called few-shot learning.^{42,43} Prepared in this way, the detector would not only be trained fast but also should be able to work on unforeseen data.⁴⁴ In the analysis, a single detector was trained and run on 12 datasets, comprised of different recordings of similar backgrounds (e.g., Western Mediterranean Wetlands Bird Dataset: 161 recordings of 12 endemic bird species), or the same species (Meerkats: 2 individuals). Considering the diversity of data sets—different species, sound quality, equipment used to record, etc.—the authors demonstrated generalized models, and the sound event detection (SED) approach, with limited training data, can be effectively applied in transcribing animal sounds.

Methodology

- a. Few-Shot Learning Framework: Few-shot learning techniques were adapted to bioacoustics purposes. Prototypical networks were used to create a prototype representation for each class based only on a few examples.
- b. Data Collection: Open datasets of various animal species sounds were designed to test the system's ability to perform few-shot SED. The initial input included labels of sound events' start and end times.
- c. Model Training and Testing: The model was trained on a small number of examples (as few as five few-shot learning techniques). Testing its performance included its ability to detect the same sound events on long-duration audio recordings.
- e. Evaluation: Evaluation of the model performance was checked through the public contest. Through it, it was demonstrated that prototypical networks, when enhanced with adaptations for the general characteristics of animal sounds, performed strongly.

The study shows that few-shot learning can be effectively applied to detect animal sounds with minimal training data. This is important, as a solution to the lack of robust, high-quality datasets, showing even a few good-quality recordings can serve as a solid training dataset.

Data Quantity: Community-Science-Based Method of Monitoring Populations⁴⁵

Authors offered wingbeats specificities to detect different species of mosquitos (as one of the most prominent

insect disease vector groups).⁴⁶ Moreover, they tested smartphones as sampling devices, as a practical and inexpensive, community-oriented approach to record sampling.

Methodology

- a. Data sampling: Recordings were made with two brands of smartphones. In this way, the nonscientific community was included in data sampling, which is fundamental for good ML training, because of a need for an extensive dataset.
- b. DL models: Open-source, end-to-end ML platform was a base for a 48-layer convolutional neural network (CNN), with a transfer-learning technique, previously trained on unrelated datasets. The model was also given metadata, as mosquito species.
- c. Evaluation: The model performed well for the species contained in the training dataset. However, it was not able to detect the new, unknown species.

In conclusion, this research sheds light on community-based data sampling, and further processing on an open-source platform. The setbacks in distinguishing new species (for a particular area) can be overcome by building diverse datasets, with recordings of various species around the world, in an open-access database. Altogether, community-based data sampling and collaborations can help with sanitary issues such as disease spreading by invasive species, by early detection of their presence in the local ecosystems.

Optimization: Fine-Tuning of Frequency Modulated Vocalizations Detection—Humpback Whale Study³²

Humpback whales are a species well known for their complex, aberrant, and highly variable acoustic vocalizations. For this reason, even detecting their sounds in the cacophony of the oceans was a great challenge for machines. This phenomenon was confirmed when Allen et al.⁴⁷ in collaboration with Google engineers developed a detector for this species, that worked rather well and was trained on the dataset collected by Allen et al.⁴⁷ in the Pacific.⁴⁸ Once this trained model was applied to the dataset of the population occurring in the Atlantic Ocean, its accuracy drastically dropped. Thus, although the same species, variability in the repertoire of sounds in use between populations is more than a trained detection model could handle. This year, Kather et al.³² published a version of a tuned detector that works well on both populations, hoping to apply it to the rest of the world's populations, with only slight adjustments. To make it easy to adjust and user-friendly, the authors developed a framework in Python named AcoDet (acoustic detector), which is publicly available online.

Methodology

- a. Model Selection and Fine-Tuning: The researchers used a CNN, a type of DL model particularly effective for analyzing audio data. As a main novelty in this research, authors fine-tuned the already existing CNN model, previously used in a study for

North Pacific humpback whales, for the repertoire of North Atlantic humpbacks.³²

- b. Framework Development: A framework named AcoDet (acoustic detector) was developed. As an open-access tool, this framework is made user-friendly, facilitating additional retraining of the model on new datasets.
- c. Evaluation: The model proved reliable in detecting humpback whale vocalization while maintaining low false positive rates.

In conclusion, this approach gives a good-practice approach to handling the issue of enormous data variability in some species. Instead of developing models *de novo*, it is much more practical to develop adjusting methods, for fine-tuning and updating already trained models, using CNNs.

Model Complexity: Wildlife Monitoring, by Acoustic Cues Set in a Context³

Jeanet et al. in this work offered a parallel solution for the initial challenge of how to approach an acoustic recording and did so by treating it as an image. In detail, instead of working on audio files, the model was trained based on spectrograms, that is, visual representations—frequency per time—of sounds on those audio files. The primary novelty of their work incorporates some contextual information—metadata, like location or time—into the model, improving its performance. They tested their model on two different groups of animals: birds and monkeys. For birds, context information included the geographic location of specific species distributions, while monkeys (Hainan gibbon) were specifically chosen as a study model because of their vocal activity connected solely to mornings. As such, convenient metadata used was the time of vocalizations.

Methodology

- a. Contextual Data: As an additional layer of acoustic classification model training, this work included context information—metadata: time of day, weather conditions, and habitat type. These pieces of information were added as an additional help for the model in distinguishing between similar sounds.
- b. DL Models: As a complex ML model that was used in this research, it incorporated CNNs and recurrent neural networks, to process both the acoustic signals and metadata.
- c. Data Collection: The dataset included two groups of animals, picked for their metadata specificities, that could be collected together with the acoustic recordings of the animal vocalization.
- d. Evaluation: Evaluation was assessed by standard metrics such as accuracy, precision, and recall. The performance of the enriched dataset was compared to the baseline models, which did not use an extra layer of contextual information.

The addition of contextual information strongly improved the performance of the acoustic classifiers. This

approach helps in reducing false positives and negatives, leading to a more reliable output. These results underline the strong potential of complex models, enriched with additional contextual information—all sorts of metadata—to improve performance and help the model in its precision.

Evaluation: African Elephants and Name-Like Calls⁴⁹

By now, we knew that next to humans, few other species, such as dolphins, for example, exhibit sounds resembling a name-like quality.⁵⁰ Pardo et al. conducted a playback experiment to confirm that African elephants are also a part of this group of animals. Namely, after recording the social rumbles (specific call types picked because of their acoustic properties, probable to carry complex information), scientists marked several acoustic characteristics of these calls and fed it to an ML model, which was further able to assign each call to a specific elephant. These predictions were finally confirmed by an on-field playback experiment.

Methodology

- a. Data: Researchers measured acoustic properties—like frequency, amplitude, and duration—of social rumbles of different individuals. Further, these values were manually decoded into numerical values for the ease of data processing for the machine.
- b. Training: The model was explicitly trained to find patterns in the expressed social rumbles, and make predictions of the intended receiver of the call.
- c. Model: Mel Frequency Cepstral Coefficients models were used, which is a common methodology in voice recognition.⁴¹ It is on a logarithmic scale, and, although developed for human sound reception, it can be applied for any other mammal species.⁵¹
- d. Validation: The accuracy of the model's prediction of the intended receiver of each rumble call was checked in-field, and the results were quantified by measuring the response rate and intensity of each intended receiver. Results confirm the model's accuracy, as elephants responded more quickly and vocally to calls addressed to them.

In conclusion, this research potentially brought a Rosetta stone for elephant communication, giving just a glimpse into the potential of ML and what it can teach us about other communication systems around us.

Recommendations for the Future and the Ethics Issue

Next to numerous ethical issues in using AI generally, as well as in science, which are well documented in recent publications,^{38, 52–55} here I would like to focus on discussing the issues that are solely connected and applicable to bioacoustics.

One of the definitions of intelligence is an organism's ability to adapt to novel circumstances.⁵⁶ If we look at AI from this perspective, at this point in time, AI is far from it. What ML was trained well for is solving *specific* problems.^{38, 57} Similar to mimicking organic intelligence, the development of the artificial version should follow the same recipe—it should hold the tools

and means that are generalized enough to solve problems that were not met before, learning from these and other previous experiences. One good example of this issue was elaborated in the previous section, a work of Kather et al. from 2023,³² where they fine-tuned an earlier model, to be applicable to different populations of the same species, making the detection model more generalized.

As descriptively explained by Mitchell,⁵⁸ and further reminded of by Bossert et al.,³⁸ *“the mind cannot be separated from its body. Intelligence cannot be reproduced by disembodied computers. This is why more and more AI researchers argue in favor of the embodiment hypothesis, stressing that machine intelligence needs some kind of body that interacts with the world and makes experiences,”* the issue of accessing solely thinking processes without paying attention to the experiences that developed the processes or the momentary situation in which the certain thinking process happening is unlikely to give full information to the observer. In other words, focusing only on intelligence as a separate entity of the organism it belongs to, or in our case, sounds that are created as a product, will not teach us much about the animals we would like to observe. How this works well in nature—to the context of the caller and the receiver—was pivotal information for discovering elephant “signature” calls, for example.^{19,49}

If we choose to ignore the context, on the other hand, a dilemma arises—can AI in this form of “disembodied intelligence” be trustworthy in making conclusions, and in parallel, to what level of information gathered solely on a single existing approach—thus, context-free—should be taken into account? This approach of cherry-picking data—gathering only the specific, focused data type—leans toward the emotionally detached treatment of animals, which is a burning question of the use of AI in biology, as it reduces humans to animals, potentially leading to the complete dehumanization of the human-animal connection in research.

However, despite the concerns animals are too often considered automata and machine-like,³⁸ which is indubitably the case, what needs to be acknowledged is that the AI approach operates on identifying patterns and systems lying behind all of life. A great example was deciphering human languages, without directly referring to particular vocabularies (an approach named word embedding, which treats words as vectors in a coordinate system).⁵⁹ Before this AI discovery, linguists had not had the slightest idea this general pattern of all human languages exists, no matter how distant they are, in a linguistic sense. Although this was a paradigm from the beginning of the development of AI, the discovery in question was a mind-setting point toward the idea of ever-existing patterns behind all life. Considering how repeatable life needs to be (in its different aspects: anatomy, physiology, genesis, behavior, etc.), it seems practical to have blueprints that are reproducible with a certain level of discretion. A recent exciting discovery with this mindset brought us information

that humpback whale songs show language-like statistical structure.⁶⁰

Who Owns Collected Data, and How Should It Be Used Responsibly?

This is a question worth exploring,⁶¹ as beyond authorship rights of the institution standing behind data collecting, the general idea, as in human societies of metadata ownership, is who has the right to collect these data, and how to limit their misuse by mal-intended parties. This further raises the question of, specifically, misuse of data or, more accurately, the findings. These findings could be used for poaching, disturbing animals, or, slightly less harmful, as incorrect and misleading conclusions of their behavior and biology. As a domino effect, this can further lead to misguided conservation efforts and interventions²⁹ (e.g., some notable historical examples of misleading conclusions leading to disastrous conservation efforts are connected to the intentional introduction of nonnative species as a means of “pest” control).⁶² A similar effect can be caused even when the intervention is not so direct and concrete. An important example would be tampering with the culture of humpback whales—in particular, humpback males are known to learn the songs of conspecifics from the same stock and are also highly receptive to novel songs. They pick up new songs of other stocks on feeding grounds^{30,31,63} and bring them back to their breeding ground—this is how the song revolutions are passed down in this species. The scenario where humans would try to interact with humpback whales by playing an AI-created artificial song could influence the local whale culture in unforeseen ways, interfere with natural processes, and have long-term consequences potentially on the entire species.⁶⁴

The final remark regarding discussing ethical issues in using AI in bioacoustics is a subject that spreads across different means of scientific necessities: equity and access. The same as for open-access publishing, scientific dissemination, and access to equipment and tools applies to bioacoustics in developing areas of the world. In the case of bioacoustics, the issue goes in both directions—these communities should be able to record their nature but also have access to the records made in their territories. A common kind of collaboration situation, geographically, is the polarity between north and south, in terms of economic development, and inversely, richness in biodiversity, privileged institutions collaborate with the underprivileged in the quest for data gathering. The latter is in need of collaborator because of a lack of funds for the equipment and data dissemination (e.g., fees for open-access journals, attending conferences, etc.), and the former are simply in a position to choose, with no constraints, their sphere of interest for the research. Needless to say is equality in decision-making and discussion positions in these types of collaborations, which are often unfairly polarized. Moreover, this overflows to the situation where instead of being peers in a position to complement and review each other’s work and methods, having one powerful institution exerting its influence of

different kinds globally, makes a dangerous, monopoly-like situation in certain fields of research, bioacoustics included.

Scaling AI Applications in Resource-Limited Settings

In addition to well-known strategies mentioned above, such as centralized data banks for unlimited data sharing and collaborations between countries and institutions, as well as citizen-science data gathering, there are several practical guidelines that can be implemented when resources are limited, specifically helpful in the context of bioacoustics. The top solution would be to use pretrained models, which can be instantly put to use. These models are readily available and do not require extensive training, offering a quick and efficient way to analyze bioacoustic data, particularly in resource-constrained environments.

Moreover, open-source frameworks like *acoupi*⁶⁵ play a pivotal role. *Acoupi* integrates audio recording, AI-based data processing, and data management, thereby allowing for real-time data analysis in the field. This framework is accessible on GitHub, providing researchers with the tools needed to perform instant data analysis, saving both time and resources.

Low-cost, small-sized hardware such as mini-computers like *Raspberry Pi*⁶⁶ also proves to be highly beneficial. When equipped with a microphone, a *Raspberry Pi* can record and process data in real-time in the field.^{66,67} These devices can be powered by external sources, such as solar panels, and can even be outfitted with wireless messaging capabilities to send results to a centralized base.^{65,68} These features make *Raspberry Pi* an ideal solution for bioacoustic research in remote locations.

The impact of these low-budget, user-friendly AI solutions is particularly significant in developing countries. These solutions democratize access to advanced research tools and methodologies, reducing the required resources (both financial and human) and lowering the level of expertise needed to process data and access the information within it. As a result, they can be utilized as a monitoring system, providing real-time information and facilitating quick decision-making. Such advancements have the potential to greatly influence the entire field of bioacoustics, empowering researchers with innovative and cost-effective solutions.

Challenges of Legislations

The first challenge is defining "AI." Different stakeholders depict its usage and capabilities differently,⁶⁹ raising questions about the authority and scope of any legislation. Additionally, there's a lack of uniformity in AI regulations worldwide, which affects international cooperation and the synchronization of regulations and development (for example, while the European Union enacted an AI act in 2024,⁷⁰ in the USA, to date, there are *no comprehensive federal legislation or regulations in the USA that regulate the development of AI or specifically prohibit or restrict their use—no overarching AI act*).⁷¹ Lastly, legislative bodies responsible for creating these laws, along with their scope and

operational intensity, need to be established at national levels and globally.^{69,71} While debates have begun, they often struggle to keep pace with the rapid development and implementation of AI.

Conclusion

We are witnessing a technological revolution, and a big part of it is the development of AI. A big leap happened recently, and since then, the development has been in steady growth. However, we need to acknowledge this process in its beginnings, still relying heavily on human manual input. The same goes for bioacoustics, and, although it will not be able to fully automate the entire research process, what bioacousticians are focused on is the amount of data the machines can process incomparably quicker. That is a big burden taken off our backs. One big chunk of work remains—data labeling, but we can see how AI engineers are going forward with this approach in the shape of self-supervised models, for example.¹³

This area evolves at a breakneck pace, and in some concepts, we are not able to keep up. This applies mostly to legislation on the use of AI, and further, to data collected in this way. This is a very important issue, and I believe the use should be cautious and restricted, until the general code of conduct is assembled and applied. With the development of the technology, the recommendations of use should be constantly discussed and periodically revised, as technology changes and our use of it.

Finally, the most exciting part is bridging the gap between the scientific and civil communities, where each can contribute to the development of this method, which has the potential to influence life on this planet, on a scale parallel to the invention of electricity.⁷²

Acknowledgments

Divna would like to thank the external reviewers for insightful comments, Mira Lukanovic for a supportive feedback on the article and the publisher for commissioning this work.

References

- 1 Paré G, Kitsiou S. Chapter 9: Methods for literature reviews. In: Lau F, Kuziemyk C, editors. Handbook of eHealth evaluation: an evidence-based approach. Victoria (BC): University of Victoria; 2017. Available from: <https://www.ncbi.nlm.nih.gov/books/NBK481583/>
- 2 Rutz C, Bronstein M, Vernes CS, Zaccarian K, Blasi ED. Using machine learning to decode animal communication. *Science* (1979). 2023;381(6654):152–5.
- 3 Jeantet L, Dufourq E. Improving deep learning acoustic classifiers with contextual information for wildlife monitoring. *Ecol Inform*. 2023;77:102256.
- 4 Simpson AJR, Roma G, Plumbley MD. Deep karaoke: extracting vocals from musical mixtures using a convolutional deep neural network. In: Emmanuel V, Yeredor A, Koldovsky Z, Tichavský P, editors. Latent variable analysis and signal separation. Cham: Springer International Publishing; 2015. p. 429–36.
- 5 Stowell D. Computational bioacoustics with deep learning: A review and roadmap. *PeerJ*. 2022;10:e13152. <https://doi.org/10.7717/peerj.13152>
- 6 Kershenbaum A, Blumstein DT, Roch MA, Akçay Ç, Backus G, Bee MA, et al. Acoustic sequences in non-human animals: a tutorial review and prospectus. *Biol Rev*. 2016;91(1):13–52.
- 7 Elemans CPH, Jiang W, Jensen MH, Pichler H, Mussman BR, Nattestad J, et al. Evolutionary novelties underlie sound production

- in baleen whales. *Nature*. 2024;627(8002):123–9. <https://doi.org/10.1038/s41586-024-07080-1>
- 8 Benko A, Sik Lányi C. History of artificial intelligence. In: *Encyclopedia of information science and technology*. 2nd ed. IGI Global; 2009. p. 1759–62.
 - 9 Flasiński M. History of artificial intelligence. In: *Introduction to artificial intelligence*. Cham: Springer International Publishing; 2016. p. 3–13.
 - 10 Haenlein M, Kaplan A. A brief history of artificial intelligence: on the past, present, and future of artificial intelligence. *Calif Manage Rev*. 2019;61(4):5–14.
 - 11 Bermant PC, Bronstein MM, Wood RJ, Gero S, Gruber DF. Deep machine learning techniques for the detection and classification of sperm whale bioacoustics. *Sci Rep*. 2019;9(1):12588.
 - 12 Raschka S. *Python machine learning*; 2016. Available from: <https://github.com/rasbt/python-machine-learning-book>
 - 13 China CR. Machine learning types. IBM; 2023 [Accessed 27 October 2024]. Types of Machine Learning | IBM. Available from: <https://www.ibm.com/think/topics/machine-learning-types>
 - 14 Sutton RS, Barto AG. *Reinforcement learning: an introduction*. Cambridge, Massachusetts; London, England: The MIT Press; 2015.
 - 15 Arulkumaran K, Deisenroth MP, Brundage M, Bharath AA. A brief survey of deep reinforcement learning. *IEEE signal processing magazine*, special issue on deep learning for image understanding; 2017. Available from: <http://arxiv.org/abs/1708.05866>
 - 16 Lecun Y, Bengio Y, Hinton G. Deep learning. *Nature*. 2015;521(7553):436–44. Available from: <http://colah.github.io/>
 - 17 Kanelis D, Liolios V, Papadopoulou F, Rodopoulou MA, Kampelopoulos D, Siozios K, et al. Decoding the behavior of a queenless colony using sound signals. *Biology (Basel)*. 2023;12(11):1392. <https://doi.org/10.3390/biology12111392>
 - 18 Licciardi A, Carbone D. WhaleNet: A novel deep learning architecture for marine mammals vocalizations on watkins marine mammal sound database. *IEEE Access*. 2024;12:154182–94. Available from: <https://ieeexplore.ieee.org/document/10720021/>
 - 19 Lokhandwala S, Sinha R, Ganji S, Pailla B. Decoding Asian elephant vocalisations: unravelling call types, context-specific behaviors, and individual identities. In: *Lecture notes in computer science (including subseries lecture notes in artificial intelligence and lecture notes in bioinformatics)*. Springer Science and Business Media Deutschland GmbH; 2023. p. 367–79.
 - 20 Çiğ A, Koçak MA, Mikail N. A different factor in the use of plants in landscape architecture: sound (type, intensity and duration) in the example of *Hyacinthus orientalis* L. *Not Bot Horti Agrobo*. 2023;51(3):13271.
 - 21 Nolasco I, Singh S, Morfi V, Lontanen V, Strandburg-Peshkin A, Vidaña-Vila E, et al. Learning to detect an animal sound from five examples. *Ecol Inform*. 2023;77:102258. Available from: <http://arxiv.org/abs/2305.13210>
 - 22 Pala A, Oleynik A, Malde K, Handegard NO. Self-supervised feature learning for acoustic data analysis. *Ecol Inform*. 2024;84:102878.
 - 23 Cauzinille J, Favre B, Marxer R, Clink D, Ahmad AH, Rey A. Investigating self-supervised speech models' ability to classify animal vocalizations: the case of gibbon's vocal identity. Available from: <https://github.com/jcauzi/>
 - 24 Rodríguez Ballesteros A, Desjonquères C, Hevia V, García Llorente M, Ulloa JS, Llusia D. Towards acoustic monitoring of bees: wingbeat sounds are related to species and individual traits. *Philos Trans R Soc Lond B Biol Sci*. 2024;379(1904):20230111. <https://doi.org/10.1098/rstb.2023.0111>
 - 25 Freeman BG. Shazam for birds. *Proc Natl Acad Sci U S A*. 2024;121(36):e2414224121. <https://doi.org/10.1073/pnas.2414224121>
 - 26 Best P, Paris S, Glotin H, Marxer R. Deep audio embeddings for vocalisation clustering. *PLoS One*. 2025;18(7):0283396. Available from: <https://pubmed.ncbi.nlm.nih.gov/37428759/>
 - 27 Knight E, Rhinehart T, de Zwaan DR, Weldy MJ, Cartwright M, Hawley SH, et al. Individual identification in acoustic recordings. *Trends Ecol Evol*. 2024;39(10):947–960. Available from: <https://pubmed.ncbi.nlm.nih.gov/38862357/>
 - 28 Sarkar E, Doss MM. Can self-supervised neural representations pre-trained on human speech distinguish animal callers? <https://doi.org/10.48550/arXiv.2305.14035>
 - 29 Sharma S, Sato K, Gautam BP. A methodological literature review of acoustic wildlife monitoring using artificial intelligence tools and techniques. *Sustainability*. 2023;15(9):7128.
 - 30 Gonçalves MIC, Djokic D, Baumgarten JE, Marcondes MCC, Padovese LR, Eugenio LDS, et al. Abrupt change in humpback whale song from Brazil suggests cultural revolutions may occur in the South Atlantic. Vol. 40. *Marine Mammal Science*. John Wiley and Sons Inc; 2023.
 - 31 Schall E, Djokic D, Ross-Marsh EC, Oña J, Denkinger J, Ernesto Baumgarten J, et al. Song recordings suggest feeding ground sharing in Southern Hemisphere humpback whales. *Sci Rep*. 2022;12(1):13924. <https://doi.org/10.1038/s41598-022-17999-y>
 - 32 Kather V, Seipel F, Berges B, Davis G, Gibson C, Harvey M, et al. Development of a machine learning detector for North Atlantic humpback whale song. *J Acoust Soc Am*. 2024;155(3):2050–64. <https://doi.org/10.1121/10.0025275>
 - 33 Best P, Paris S, Glotin H, Marxer R. Deep audio embeddings for vocalisation clustering. *PLoS One*. 2023;18(7):0283396.
 - 34 Malige F, Djokic D, Patris J, Sousa-Lima R, Glotin H. Use of recurrence plots for identification and extraction of patterns in humpback whale song recordings. *Bioacoustics*. 2021;30(6):680–95.
 - 35 Ojeda SAA, Peramo EC, Solano GA. Application of deep learning in recurrence plots for multivariate nonlinear time series forecasting. In: *Tsibrintzis GA, Virvou M, Jain LC, editors. Advances in machine learning/deep learning-based technologies: selected papers in honour of professor Nikolaos G Bourbakis*. Vol. 2. Cham: Springer International Publishing; 2022. p. 169–85. https://doi.org/10.1007/978-3-030-76794-5_9
 - 36 Walsh I, Titma T, Psomopoulos FE, Tosatto S. Recommendations for machine learning validation in biology. 2020. Available from: <https://www.researchgate.net/publication/342547853>
 - 37 Chalmers C, Fergus P, Wich S, Longmore SN. Modelling animal biodiversity using acoustic monitoring and deep learning. In: *International joint conference on neural networks (IJCNN)*; 2021. Available from: <https://www.xeno-canto.org/>
 - 38 Bossert L, Hagendorff T. Animals and AI. The role of animals in AI research and application – an overview and ethical evaluation. *Technol Soc*. 2021;67:101678.
 - 39 Sainburg T, Thielk M, Gentner TQ. Finding, visualizing, and quantifying latent structure across diverse animal vocal repertoires. *PLoS Comput Biol*. 2020;16(10):e1008228.
 - 40 Kershenbaum A, Akçay Ç, Babu-Saheer L, Barnhill A, Best P, Cauzinille J, et al. Automatic detection for bioacoustic research: a practical guide from and for biologists and computer scientists. *Biol Rev Camb Philos Soc*. 2024;100(2):620–46. <https://doi.org/10.1111/brv.13155>
 - 41 Mutanu L, Gohil J, Gupta K, Wagio P, Kotonya G. A review of automated bioacoustics and general acoustics classification research. *Sensors*. 2022;22(21):8361.
 - 42 Snell J, Swersky K, Zemel TR. Prototypical networks for few-shot learning. In: *31st conference on neural information processing systems (NIPS 2017)*. Long Beach; 2017.
 - 43 Wang Y, Yao Q, Kwok JT, Ni LM. Generalizing from a few examples: a survey on few-shot learning. *ACM Comput Surv*. 2020;53(3):1–34.
 - 44 Beery S, Horn G, Perona P. Recognition in terra incognita. In: *Computer vision – ECCV 2018*. Springer; 2018. p. 472–89. Available from: <https://beerys.github.io/CaltechCameraTraps/>
 - 45 Khalighifar A, Jiménez-García D, Campbell LP, Ahadi-Dabla KM, Aboagye-Antwi F, Ibarra-Juárez LA, et al. Application of deep learning to community-science-based mosquito monitoring and detection of novel species. *J Med Entomol*. 2022;59(1):355–62.
 - 46 Juliano SA, Lounibos LP. Ecology of invasive mosquitoes: effects on resident species and on human health. *Ecol Lett*. 2005;8(5):558–74.
 - 47 Allen AN, Harvey M, Harrell L, Jansen A, Merken KP, Wall CC, et al. A convolutional neural network for automated detection of humpback whale song in a diverse, long-term passive acoustic dataset. *Front Mar Sci*. 2021;8:607321.
 - 48 Chen A. Pattern radio: whale songs; 2019. Available from: <https://medium.com/@alexanderchen/pattern-radio-whale-songs-242c692ff60>
 - 49 Pardo MA, Fristrup K, Lolchuragi DS, Poole JH, Granli P, Moss C, et al. African elephants address one another with individually specific name-like calls. *Nat Ecol Evol*. 2024;8(7):1353–64.
 - 50 Caldwell MC, Caldwell DK. Individualized whistle contours in bottle-nosed dolphins (*Tursiops truncatus*). *Nature*. 1965;207(4995):434–5. Available from: <https://www.nature.com/articles/207434a0>

- 51 Lyon RF. On logarithmic and power-law hearing. In: Lyon RF, editor. Human and machine hearing: extracting meaning from sound. Cambridge: Cambridge University Press; 2017. p. 33–45. Available from: <https://www.cambridge.org/core/product/5D301ACA2521CB56CD0AE7F3C8FD10B9>
- 52 Coghlan S, Parker C. Helping and not harming animals with AI. Vol. 37. Philosophy and Technology. Springer Science and Business Media B.V.; 2024.
- 53 Coghlan S, Parker C. Harm to nonhuman animals from AI: a systematic account and framework. Philos Technol. 2023;36:25. <https://doi.org/10.1007/s13347-023-00627-6>
- 54 Bossert LN. Benefitting nonhuman animals with AI: why going beyond “do no harm” is important. Vol. 36. Philosophy and Technology. Springer Science and Business Media B.V.; 2023.
- 55 Resnik DB, Hosseini M. The ethics of using artificial intelligence in scientific research: new guidance needed for a new tool. AI Ethics. 2024;5(2):1499–521. <https://doi.org/10.1007/s43681-024-00493-8>
- 56 Müller U, Ten Eycke K, Baker L. Piaget’s theory of intelligence. In: Handbook of intelligence. New York, NY: Springer New York; 2015. p. 137–51.
- 57 Crosby M, Beyret B, Halina M. The animal-AI olympics. Nat Mach Intell. 2019;1(5):257. <https://doi.org/10.1038/s42256-019-0050-3>
- 58 Mitchell M. Artificial intelligence: a guide for thinking humans. New York: Picador, Farrar, Straus and Giroux; 2019.
- 59 Mikolov T, Corrado GS, Chen K, Dean J. Efficient Estimation of word representations in vector space; 2013. p. 1–12. <https://arxiv.org/abs/1301.3781>
- 60 Arnon I, Kirby S, Allen JA, Garrigue C, Carroll EL, Garland EC. Whale song shows language-like statistical structure. Science. 2025;387(6734):649–53.
- 61 Oswald JN, Van CAM, Dassow A, Elliott T, Johnson MT, Ravignani A, et al. A collection of best practices for the collection and analysis of bioacoustic data. Appl Sci. 2022;12(23):649–53.
- 62 Ford AT, Ali AH, Colla SR, Cooke SJ, Lamb CT, Pittman J, et al. Understanding and avoiding misplaced efforts in conservation. FACETS. 2021;6(1):252–71. Available from: www.facetsjournal.com
- 63 Rekdahl ML, Garland EC, Carvajal GA, King CD, Collins T, Razafindrakoto Y, et al. Culturally transmitted song exchange between humpback whales (*Megaptera novaeangliae*) in the southeast Atlantic and southwest Indian Ocean basins. R Soc Open Sci. 2018;5:172305. Available from: <https://royalsocietypublishing.org/doi/10.1098/rsos.172305>
- 64 Bloomberg Originals. Could AI unlock the secrets of animal communication? The Future With Hannah Fry; 2023. Available from: <https://youtu.be/ka894z9pNls?feature=shared>. 2024
- 65 Vuillimienet A, Balvanera SM, Mac Aodha O, Jones KE, Wilson D. acoupi: an open-source python framework for deploying bioacoustic AI models on edge devices. HardwareX. 2025;12:e00337. Available from: <http://arxiv.org/abs/2501.17841>
- 66 Kiarie G, wa Maina C. Raspberry Pi based recording system for acoustic monitoring of bird species. In: 2021 IST-Africa conference (IST-Africa). South Africa; 2021. p. 1–8. Available from: <https://ieeexplore.ieee.org/document/9576984>
- 67 Florentin J, Verlinden O. Autonomous wildlife soundscape recording station using raspberry Pi. In: 24th international congress on sound and vibration. London: ICSV24; 2017. Available from: <https://www.ioa.org.uk/system/files/publications/j%20florentin%20o%20verlinden%20autonomous%20wildlife%20soundscape%20recording%20station%20using%20a%20raspberry%20pi.pdf>
- 68 Kiarie G, Kabi J, Wa Maina C. DSAIL power management board: powering the raspberry Pi autonomously off the grid. HardwareX. 2022;12:e00337. <https://doi.org/10.1016/j.ohx.2022.e00337>
- 69 Hoffman RR, Mueller ST, Klein G, Jalaeian M, Tate C. Explainable AI: roles and stakeholders, desiderata and challenges. Front Comput Sci. 2023;5:1117848.
- 70 Regulation - EU - 2024/1689 - EN - EUR-Lex. Regulation (EU) 2024/1689 of the European Parliament and of the Council. Off J Eur Union. 2024;2024/1689. Available from: <https://eur-lex.europa.eu/eli/reg/2024/1689/oj>
- 71 IAPP Research and Insight. Global AI law and policy tracker. 2024. Available from: https://iapp.org/media/pdf/resource_center/global_ai_law_policy_tracker.pdf
- 72 Jewell C. Artificial intelligence: the new electricity. WIPO Magazine. 2019. Available from: https://www.wipo.int/wipo_magazine/en/2019/03/article_0001.html