Interpreting ML in Ecology: A RAG-Based Approach to Explainability in Species Distribution Modeling

Renato O. Miyaji[1], **Pedro L. P. Corrêa**[1]

[1]Escola Politécnica – Universidade de São Paulo (USP)

{re.miyaji,pedro.correa}@usp.br

Abstract. *Species Distribution Modeling (SDM) relies increasingly on Machine Learning (ML), but many models remain opaque, limiting their usability for non-experts. While SHAP and LIME improve interpretability, they still require technical expertise. This study proposes an agentic Retrieval-Augmented Generation (RAG) framework, integrating ML models (Logistic Regression, Random Forests, MLP), XAI techniques (SHAP, LIME), and a LLM-powered explanation system to enhance explainability. Using GoAmazon 2014/15 environmental data and GBIF species occurrences, we evaluated explanations based on completeness, and context-awareness, achieving a significant improvement in this study case. Results indicate that LLMs combined with XAI can significantly enhance explainability in SDM.*

1. Introduction

Species Distribution Modeling (SDM) is a crucial approach in Computational Ecology that seeks to understand the relationships between species occurrences and environmental variables [Elith and Leathwick 2009]. By identifying the factors that influence species distributions, SDM plays a fundamental role in ecological research, conservation planning, and biodiversity management. These models provide insights into ecological niches and habitat suitability, supporting decision-making processes in environmental policy and species conservation strategies [Hutchinson 1991].

Machine Learning (ML) techniques have increasingly been used in SDM due to their high predictive power and ability to model complex ecological patterns [Beery et al. 2021]. Unlike traditional statistical approaches, ML models can capture non-linear relationships and interactions among environmental variables, leading to more accurate and robust predictions. However, this increased predictive performance often comes at the cost of interpretability, as many ML models are considered ”black boxes” that provide little insight into their decision-making processes [Ryo et al. 2021].

In the literature on explainability, [Doran et al. 2017] classify ML models into three categories: comprehensible, interpretable, and opaque. Many of the ML models commonly used in SDM fall into opaque or interpretable categories due to their complexity and lack of transparency [Ryo et al. 2021]. This opacity poses a significant challenge in ecological applications, where the end users—such as conservation biologists, policymakers, and land managers—often lack expertise in ML and Data Science, making it difficult for them to understand and trust the model outputs [Ryo et al. 2021].

To address this issue, there is a growing interest in enhancing the explainability of ML models for non-expert users. Recent advancements in Large Language Models (LLMs) have demonstrated remarkable capabilities in communication, domain knowledge integration, and contextual reasoning. These models can process and generate human-like text, making them valuable tools for bridging the gap between complex computational techniques and user-friendly explanations [Spitzer et al. 2024] [Zytek et al. 2024]. By leveraging LLMs, it is possible to translate intricate ML model outputs into accessible, domain-relevant insights, improving transparency and trust in ecological modeling.

In response to these developments, researchers have begun exploring methods that integrate LLMs to improve ML explainability [Spitzer et al. 2024] [Zytek et al. 2024]. This emerging field seeks to utilize LLMs' ability to process vast amounts of knowledge and generate tailored explanations, thereby facilitating better understanding and decision-making among users without ML expertise. Such approaches hold promise for a wide range of applications, particularly in fields where technical complexity often hinders accessibility and adoption [Spitzer et al. 2024] [Zytek et al. 2024].

Following this trend, this study aims to explore the integration of LLMs into ML-based SDM to enhance model interpretability for non-expert users. By developing and evaluating techniques that leverage LLMs for explanation generation, we seek to make ML-based SDM more transparent and comprehensible to ecologists, conservationists, and other stakeholders who rely on these models for critical decision-making.

This study makes several key contributions. To the best of our knowledge, it is the first to explore the integration of LLMs with Explainable AI (XAI) techniques to enhance the interpretability of ML-based SDM models. This approach not only improves the transparency of species distribution predictions but also has the potential to be extended to other Computational Ecology applications that rely on ML. Furthermore, we propose a more robust implementation of Retrieval-Augmented Generation (RAG) with agentic capabilities, incorporating domain-specific knowledge and Web Search to generate context-aware explanations. To evaluate the effectiveness of our approach, we develop case studies focused on the Amazon Basin, a region of critical ecological importance. These case studies provide empirical insights into how LLM-enhanced explainability can support conservation efforts, ecological research, and environmental decision-making.

2. Related Works

The quest for explainability in ML has been a focal point of research, especially concerning the transparency of models used in critical applications. Opaque systems, often referred to as "black boxes," provide no insight into their internal mechanisms. Interpretable systems allow users to mathematically analyze their algorithmic processes, while comprehensible systems emit symbols enabling user-driven explanations of how conclusions are reached [Doran et al. 2017].

In SDM, the integration of ML has been hindered by challenges related to model interpretability. Previous studies, such as those by [Ryo et al. 2021] and [Miyaji et al. 2023], have demonstrated the use of explainability techniques, including SHapley Additive exPlanations (SHAP) [Lundberg and Lee 2017] and Local Interpretable Model-agnostic Explanations (LIME) [Ribeiro et al. 2016], to enhance the understanding of ML-based predictions in ecological applications. However, these methods often demand a high level of technical expertise in both ML and XAI, posing a barrier to their widespread adoption by non-expert users [Ryo et al. 2021].

The emergence of LLMs has opened new avenues for enhancing the comprehensibility of ML models [Spitzer et al. 2024] [Zytek et al. 2024]. LLMs possess significant domain knowledge and advanced communication capabilities, making them suitable for generating human-understandable explanations. Recent studies have explored the integration of LLMs with XAI techniques to transform opaque models into comprehensible systems. For instance, [Spitzer et al. 2024] discussed the potential of LLMs to provide natural language explanations, thereby making complex models more accessible to non-specialists. Initial explorations in this domain have utilized various strategies, including In-Context Learning (ICL) [Zytek et al. 2024], and RAG [Spitzer et al. 2024]. ICL involve crafting specific prompts to guide LLMs in generating relevant explanations without additional training. RAG combines LLMs with external knowledge sources, enabling the generation of contextually enriched explanations.

Despite these advancements, the application of LLMs for enhancing explainability in SDM remains unexplored. This gap presents an opportunity to leverage LLMs' capabilities to make SDM models more comprehensible.

2.1. Machine Learning Models for SDM

SDM relies on ML models to predict species occurrences based on environmental variables [Beery et al. 2021]. Among the commonly used models, Logistic Regression is favored for its interpretability. The model assigns coefficients to each predictor variable [James et al. 2013], indicating their impact on the probability of species occurrence. This transparency makes it a useful tool for understanding ecological patterns and decision-making.

Random Forests (RF) [Breiman 2001] are also widely employed in SDM due to their high predictive performance and robustness. RF is an ensemble learning method that constructs multiple decision trees using randomly selected subsets of data and features. The model's interpretability is mainly derived from the structure of decision trees and feature importance metrics [James et al. 2013]. In contrast, Multilayer Perceptrons (MLPs), a type of artificial neural network, are considered opaque or "black-box" models due to their complex internal structure. Unlike decision trees or linear models, MLPs consist of multiple hidden layers with numerous interconnected neurons, making it difficult to directly interpret how input variables influence the model's predictions [James et al. 2013]. This lack of transparency presents a significant challenge [Doran et al. 2017] in ecological applications.

2.2. XAI Models

To improve the interpretability of ML models, Explainable AI (XAI) techniques have been developed. One of the most widely used techniques is SHAP [Lundberg and Lee 2017], which is based on cooperative game theory. SHAP assigns importance values to each feature by estimating their marginal contributions to the prediction [Lundberg and Lee 2017]. This approach allows for a detailed analysis of how each environmental variable influences the model's output, making it particularly useful for understanding SDM models [Ryo et al. 2021].

Another popular XAI method is LIME [Ribeiro et al. 2016]. LIME creates locally interpretable surrogate models by perturbing the input data and training simple linear models to approximate the predictions of the complex ML model. This enables users to understand how specific predictions are made by analyzing the contributions of individual features [Ribeiro et al. 2016].

Despite their effectiveness in enhancing model interpretability, both SHAP and LIME require a solid understanding of ML and XAI principles. As a result, these methods, while valuable, may not be fully accessible to ecologists, conservationists, and other stakeholders without ML expertise [Ryo et al. 2021].

2.3. LLMs and RAG

LLMs have emerged as powerful tools in Natural Language Processing (NLP), capable of generating human-like text and reasoning across various domains [Caseli and Nunes 2023]. However, LLMs often have limitations in specialized knowledge areas. To address these limitations, specific techniques are required to enhance their domain specialization and ensure they provide accurate and relevant explanations [Zhu et al. 2024].

One approach to improving LLMs' domain knowledge is Fine-Tuning, where the model is trained on specialized datasets to adapt its parameters to a particular field. However, it comes with several challenges, including high computational costs, the need for large labeled datasets, and difficulties in keeping the model updated with new scientific findings [Caseli and Nunes 2023].

An alternative approach is RAG, which offers a more dynamic and cost-effective solution. RAG combines the generative capabilities of LLMs with an external retrieval mechanism, allowing the model to query external knowledge sources before generating responses. This enables the model to provide up-to-date and contextually rich explanations, even in highly specialized domains [Zhu et al. 2024]. A key component of RAG is the vector database, which stores knowledge in the form of high-dimensional embeddings. These vector representations enable efficient similarity searches, allowing the model to retrieve relevant documents or information before generating explanations. This approach enhances the LLM's ability to provide informed responses without requiring extensive fine-tuning [Caseli and Nunes 2023] [Zhu et al. 2024].

With the increasing adoption of agentic LLM architectures, additional tools can be integrated to further enhance RAG. For instance, Web Search tools can be embedded into the model pipeline, enabling real-time access to scientific literature, environmental reports, and biodiversity databases [Wang et al. 2023]. This capability ensures that the model remains updated with the latest research and provides accurate explanations in SDM applications.

By leveraging LLMs and RAG, it becomes possible to transform ML models from merely interpretable to comprehensible systems [Spitzer et al. 2024] [Zytek et al. 2024]. Instead of requiring users to understand technical aspects of SHAP or LIME, LLMs can generate natural language explanations that are tailored to the needs of non-expert users. This shift represents a major advancement in making ecological modeling more accessible and actionable.

3. Method

3.1. Proposed Method

The proposed method consists of five key components, each designed to enhance explainability in SDM by integrating ML, XAI, and LLMs with a RAG system. Figure 1 presents the proposed method.

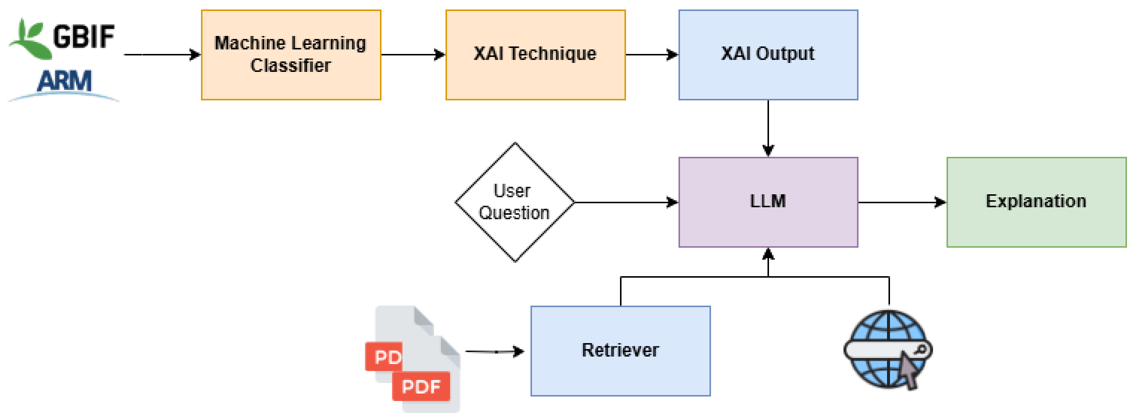

Figure 1. Proposed Method for Explainability: an agentic system powered by a LLM is used to generate natural language explanations

A vector database is used to store and retrieve relevant documents, including scientific articles on the analyzed species, SDM methodologies, XAI techniques, and ML models. These documents serve as a knowledge base to support the RAG system in generating accurate and contextually informed explanations.

The method relies on environmental variables and species occurrence data obtained from the Global Biodiversity Information Facility (GBIF) [Global Biodiversity Information Facility 2025] and the GoAmazon 2014/15 Atmospheric Radiation Measurement (ARM) campaign [ARM 2025]. These datasets provide the necessary information to train and validate SDM models.

Three ML models with different levels of interpretability are used to model species distributions. Logistic Regression, which is inherently interpretable [James et al. 2013]. Random Forests, which offer interpretability through feature importance metrics [Breiman 2001]. MLP, a neural network model considered opaque [James et al. 2013]. To enhance the interpretability of the ML classifiers, two widely used XAI techniques are applied. SHAP [Lundberg and Lee 2017] and LIME [Ribeiro et al. 2016].

A RAG system is implemented to improve model explainability [Zhu et al. 2024]. This system consists of three components. The vector database for Information Retrieval enables efficient storage and retrieval of domain-specific documents [Caseli and Nunes 2023]. The Web Search Integration allows real-time retrieval of up-to-date scientific information [Wang et al. 2023]. The Agentic Framework for Explanation Generation uses an LLM to process retrieved documents and generate natural language explanations tailored to non-expert users [Wang et al. 2023].

3.2. Case Study and Evaluation

To assess the proposed method, case studies were conducted using environmental data collected by the GoAmazon 2014/15 project [ARM 2025] and interpolated by [Miyaji et al. 2021], combined with species occurrence records from GBIF [Global Biodiversity Information Facility 2025]. The complete dataset was made available by [Miyaji et al. 2021] and includes high-resolution environmental variables and species presence data for the Amazon Basin.

The environmental predictor variables were selected based on their ecological relevance and availability in the GoAmazon dataset [Martin et al. 2017]. These variables are presented in Table 1.

Table 1. Description of the Environmental Predictor Dataset. [Martin et al. 2017]

Atmospheric Variables	Temperature, CO, O_3, NO_X, CO_2, CH_4, Isoprene, Acetonitrile CPC and H_2O
Data Acquisition	G-1 Aircraft (Altitude: 700 m to 2000 m)
Spatial Coverage	Latitude: -3.632° to -2.813° Longitude: -60.831° to -59.937°
Spatial Resolution (after Interpolation)	Latitude: 0.001° Longitude: 0.001°
Date	Wet Season: 02/01/2014 to 03/31/2014 Dry Season: 08/15/2014 to 10/15/2014 Flight Time: 11:00 AM to 1:00 PM, local time

The SDM problem was framed as a binary classification task in ML [Beery et al. 2021], where the positive class consists of observed occurrences of the analyzed species. The negative class consists of pseudo-absence samples, which are generated by selecting locations where the species is not reported but with similar environmental conditions to the presence locations [Beery et al. 2021].

For the case study, the species *Thraupis episcopus* (Blue-gray Tanager) and *Pitangus sulphuratus* (Great Kiskadee) were selected for analysis using SDM [Amâncio et al. 2008] [Cueva et al. 2022]. These species were chosen due to their high occurrence frequencies within the study area, ensuring a robust dataset for model training and validation. Additionally, both species have been the subject of previous scientific studies analyzing their distribution patterns, providing a valuable basis for comparison. *Thraupis episcopus* is a widely distributed neotropical passerine commonly found in urban, agricultural, and forested environments, known for its adaptability to different habitats [Amâncio et al. 2008]. *Pitangus sulphuratus*, a conspicuous and vocal tyrant flycatcher, is highly adaptable and frequently observed in open areas, including urban parks, wetlands, and forest edges [Cueva et al. 2022]. Their ecological flexibility and well-documented occurrences make them ideal candidates for evaluating SDM performance in this case study.

To generate human-readable explanations, the OpenAI's GPT-4o-mini [OpenAI 2025] model was used within an agentic RAG framework, implemented using LangChain (0.3.21) for orchestrating the RAG pipeline, Chroma (k=10) as a vector database for storing and retrieving relevant documents, and Tavily Web Search for retrieving real-time external knowledge when necessary.

The effectiveness of the proposed method was assessed using an LLM-as-a-judge approach with a different LLM, where the generated explanations were evaluated from 0 to 5 based on four key criteria proposed by [Zytek et al. 2024]: Soundness, which is the correctness of the information included in the narrative. Fluency, which is the extent to which the narrative sounds "natural" or like it was generated by a human peer in conversation. Completeness, which is the amount of information included in the narrative. Context-awareness, which is the degree to which the narrative "explains the explanation" by providing external context.

This evaluation framework ensures that the proposed method not only enhances interpretability but also transforms SDM models into truly comprehensible systems, making them more accessible for ecologists, conservationists, and policymakers.

4. Results

The experiments were conducted using the defined case study, focusing on the species *Thraupis episcopus* and *Pitangus sulphuratus*. The machine learning classifiers employed were Logistic Regression, Random Forests, and MLP, while SHAP and LIME were used as explainability techniques. All classifiers demonstrated strong predictive performance, attaining ROC-AUC values above 80% on the test set. For each classifier and XAI technique, we evaluated three different system configurations for generating explanations. Zero-shot prompting, in which the LLM received only the textual outputs of the applied XAI techniques, using a prompt adapted from [Zytek et al. 2024]. RAG-based system, in which the same prompt was used, but the LLM was also provided with the top-10 most relevant documents retrieved by vector similarity search from a domain-specific vector database. This database contained scientific articles referenced in this study, covering SDM, ML, XAI, and Ecology. Agentic RAG system, in which this configuration extended the RAG-based system by incorporating Web Search capabilities (Tavily Web Search), retrieving and integrating the top-5 most relevant online sources in addition to the top-10 vector database documents.

Each configuration was evaluated using an LLM-as-a-judge framework, applying the criteria defined in [Zytek et al. 2024]. Table 2 presents the comparative results obtained for each evaluation metric. Figure 2 presents an example of Explanation.

Table 2. Average ratings with different system configurations.

XAI	System	Soundness	Fluency	Completeness	Context-awareness
SHAP	Zero-Shot	4.82	4.94	4.88	1.88
SHAP	RAG-Based	4.82	4.95	4.95	1.96
SHAP	Agentic RAG	4.94	4.91	4.94	4.87
LIME	Zero-Shot	4.96	4.86	4.82	1.83
LIME	RAG-Based	4.85	4.88	4.90	1.97
LIME	Agentic RAG	4.95	4.88	4.87	4.91

The results indicate that, regardless of the ML model and XAI technique used, the performance across the evaluated criteria remains consistent. In terms of Sound-

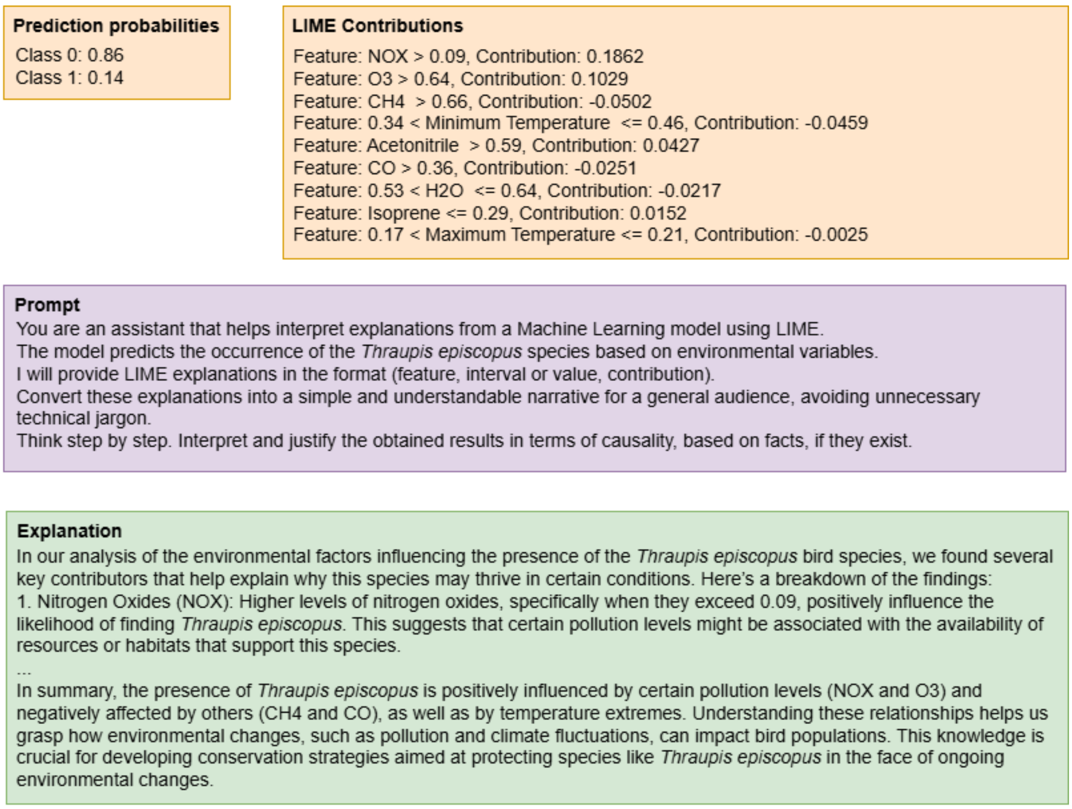

Figure 2. Example of ML Prediction, LIME Contributions and Explanation.

ness, Fluency, and Completeness, all system configurations achieve high and statistically equivalent scores, suggesting that the LLM effectively processes and communicates explanations across different settings. The most notable difference was observed in the Context-awareness metric. The comparison between Zero-Shot and RAG-Based approaches shows minimal variation, indicating that LLMs already possess strong general knowledge of broad scientific topics. However, a significant improvement occurs when incorporating an Agentic RAG system, which enhances planning and reasoning capabilities while leveraging Web Search to retrieve more targeted and contextually relevant information. As a result, the average Context-awareness score increases from 1.9 to 4.9, demonstrating a substantial and statistically significant improvement.

5. Conclusions

This study introduced an agentic RAG-based framework to enhance explainability in SDM by integrating ML, XAI, and LLMs. The results demonstrate that while traditional XAI techniques (SHAP and LIME) improve interpretability, they still require technical expertise. In contrast, LLM-powered explanations, particularly when enhanced by an Agentic RAG system, significantly improve context-awareness, achieving a substantial increase from 1.9 to 4.9 in the evaluation metric. These findings highlight the potential of LLM-based approaches to bridge the gap between opaque ML models and non-expert users, making SDM more accessible for ecologists, conservationists, and policymakers.

While this study introduces an agentic RAG-based framework for enhancing explainability in SDM, several avenues for future research remain. One key improvement is the validation of explanation relevance and comprehensibility with domain experts, replacing the LLM-as-a-judge approach with human evaluation from ecologists and conservation scientists, since LLM-as-a-Judge may introduce biases stemming from the underlying language model. Further enhancements to the vector database are also necessary, including the integration of more relevant and high-quality domain-specific documents to improve retrieval accuracy. Moreover, conducting extensive case studies in diverse ecological contexts beyond the Amazon Basin would help assess the generalizability of the proposed approach. Future work should also explore the application of LLM-driven explainability frameworks in other areas of Computational Ecology, broadening the impact of this methodology on conservation efforts and ecological research.

Acknowledgments

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Finance Code 001. It was made possible by the Thematic Projects of FAPESP ”Life cycles and aerosol clouds in the Amazon” (2017/17047-0) and ”Research Centre for Greenhouse Gas Innovation - RCG2I” (2020/15230-5).

References

Amâncio, S., Souza, V. B., and Melo, C. (2008). Columba livia e pitangus sulphuratus como indicadoras de qualidade ambiental em área urbana. *Revista Brasileira de Ornitologia*, 16(1):32–37.

ARM (2025). Arm research facility. Available on: `https://adc.arm.gov/discovery/#/results/site_code::mao`. Acessed in 23 March 2025.

Beery, S., Cole, E., Parker, J., Perona, P., and Winner, K. (2021). Species distribution modeling for machine learning practitioners: A review. In *Proceedings of ACM SIGCAS Conference on Computing and Sustainable Societies (COMPASS) 2021*.

Breiman, L. (2001). Random forests. *Machine Learning*, 45(1):5–32.

Caseli, H. and Nunes, M. (2023). *Processamento de Linguagem Natural: Conceitos, Técnicas e Aplicações em Português*. Brasileiras - Processamento de Linguagem Natural.

Cueva, D., Bravo, G., and Silveira, L. (2022). Systematics of thraupis (aves, passeriformes) reveals an extensive hybrid zone between t. episcopus (blue-gray tanager) and t. sayaca (sayaca tanager). *PLoS ONE*, 17(10).

Doran, D., Schulz, S., and Besold, T. R. (2017). What does explainable ai really mean? a new conceptualization of perspectives. In *Proceedings of the First International Workshop on Comprehensibility and Explanation in AI and ML*.

Elith, J. and Leathwick, J. R. (2009). Species distribution models: Ecological explanation and prediction across space and time. *The Annual Review of Ecology, Evolution and Systematics*, 40:677–697.

Global Biodiversity Information Facility (2025). Gbif occurrence download. `https://DOI10.15468/dl.ppwbzv`. 23 March 2025.

Hutchinson, G. E. (1991). Population studies: Animal ecology and demography. *Bulletin of Mathematical Biology*, 53(1-2):193–213.

James, G., Witten, D., Hastie, T., and Tibshirani, R. (2013). *An Introduction to Statistical Learning*. Springer, Londres.

Lundberg, S. M. and Lee, S. (2017). A unified approach to interpreting model predictions. *Proceedings of the 31st International Conference on Neural Information Processing Systems*.

Martin, S. T., Artaxo, P., Machado, L., Manzi, A. O., Souza, R. A. F. d., Schumacher, C., Wang, J., Biscaro, T., Brito, J., Calheiros, A., et al. (2017). The green ocean amazon experiment (goamazon2014/5) observes pollution affecting gases, aerosols, clouds, and rainfall over the rain forest. *Bulletin of the American Meteorological Society*, 98(5):981–997.

Miyaji, R. O., , Almeida, F. V., and Corrêa, P. L. P. (2023). Evaluating the explainability of machine learning classifiers: A case study of species distribution modeling in the amazon. In *SYMPOSIUM ON KNOWLEDGE DISCOVERY, MINING AND LEARNING (KDMILE)*, pages 49–56. SBC.

Miyaji, R. O., Almeida, F. V., Bauer, L. O., Ferrari, V., Corrêa, P. L. P., Rizzo, L. V., and Prakash, G. (2021). Spatial interpolation of air pollutant and meteorological variables in central amazonia. *Data*, 6(12):126.

OpenAI (2025). Gpt-4o mini: advancing cost-efficient intelligence. Available on: `https://openai.com/index/gpt-4o-mini-advancing-cost-efficient-intelligence/`. Acessed in 23 March 2025.

Ribeiro, M. T., Singh, S., and Guestrin, C. (2016). "why should i trust you?": Explaining the predictions of any classifier. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*.

Ryo, M., Angelov, B., Mammola, S., Kass, J. M., Benito, B. M., and F, H. (2021). Explainable artificial intelligence enhances the ecological interpretability of black-box species distribution models. *Ecography*, 44:199–205.

Spitzer, P., Celis, S., Martin, D., Kühl, N., and Satzger, G. (2024). Looking through the deep glasses: How large language models enhance explainability of deep learning models. *Proceedings of Mensch und Computer 2024*.

Wang, L., Ma, C., Feng, X., Zhang, Z., Yang, H., Zhang, J., Chen, Z., Tang, J., Chen, X., Lin, Y., Zhao, W., Wei, Z., and Wen, J. (2023). A survey on large language model based autonomous agents. *Frontiers Comput. Sci.*

Zhu, Y., Yuan, H., Wang, S., Liu, S., Liu, W., Deng, C., Chen, H., Liu, Z., Dou, Z., and Wen, J. (2024). Large language models for information retrieval: A survey. *ArXiv*.

Zytek, A., Pidò, S., and Veeramachaneni, K. (2024). Llms for xai: Future directions for explaining explanations. *ACM CHI Workshop on Human-Centered Explainable AI*.