Reflections from the 2025 EcoHack: AI & LLM Hackathon for Applications in Evidence-based Ecological Research & Practice

Jennifer D'Souza*[1], Tarek Al Mustafa†[2], Daphne Frederike Auer[2], Sarah T. Bachinger[2], Marc Brinner[3], Caren Daniel[x], Nayanika Das[4], Alexander Espig[5], Nadeen Fathallah[6], Edward Gow-Smith[7], Lorenz Gunreben[8], Nico Heider[8], Hrishikesh Jadhav[9], Basma Jalloul[9,10], Vamsi Krishna Kommineni[2], Samira Korani[11], Andrii Krutsylo[12], Zijian Ling[22], Vaishnavi Mendu[13], Shuhan Miao[14], Bartolome Ortiz-Viso[15], Anne Peter[16], Moritz Plenz[17], Javad Razavian[18], Moiz Khan Sherwani[19], Mir Nafis Sharear Shopnil[20], Will Woof[7], Birgitta König-Ries[2], and Tina Heger*[21]

[1]TIB Leibniz Information Centre for Science and Technology
[2]Friedrich-Schiller-Universität Jena
[3]University of Bielefeld
[4]Universitat de Vic-Universitat Central de Catalunya
[5]Leibniz University Hannover
[6]University of Stuttgart
[7]University of Sheffield
[8]Leipzig University
[9]University of Passau
[10]University of Sfax
[11]University of Galway
[12]Polish Academy of Sciences
[13]University of Hildesheim
[14]Technical University Berlin
[15]University of Granada
[16]Johann Heinrich von Thünen Institut
[17]Heidelberg University
[18]University of Qom
[19]University of Copenhagen
[20]Queen Mary University of London
[21]Technical University of Munich
[22]Apply U

*Corresponding authors: jennifer.dsouza@tib.eu, t.heger@tum.de. †All authors, except for the first author and the last two authors, are listed in alphabetical order by last name as members of the participating teams of the EcoHack Hackathon.

Abstract

This paper presents outcomes from the inaugural "EcoHack: AI & LLM Hackathon for Applications in Evidence-based Ecological Research & Practice," which convened participants from across Europe and beyond, culminating in 11 team submissions. These submissions highlighted six broad application areas of AI for ecology: (1) AI-enhanced decision support and automation, (2) scientific search and communication, (3) knowledge extraction and reasoning, (4) AI for ecological modeling, forecasting, and simulation, (5) causal inference and ecological reasoning, and (6) AI for biodiversity monitoring and conservation. Each team's project is summarized in a consolidated table—complete with links to source code—and described in brief papers in the appendix.

Beyond summarizing technical results, this paper offers insights into the hackathon's hybrid structure, featuring an in-person gathering in Bielefeld, Germany, alongside a global online hub that facilitated both local and virtual engagement. Throughout the event, participants showcased how large language models (LLMs) can serve as both robust tools for diverse machine learning tasks and flexible platforms for rapidly prototyping novel research applications. These efforts underscore the importance of stronger technological bridges among stakeholders in ecology, including practitioners, local farmers, and policymakers. Overall, the EcoHack outcomes highlight the transformative potential of AI in driving scientific discovery and fostering interdisciplinary collaboration in ecology.

Introduction

Science hackathons have emerged as a powerful tool for fostering collaboration, innovation, and rapid problem-solving in the scientific community [1, 2, 3]. By leveraging social media, virtual platforms, and hybrid event structures, such hackathons can be organized in a cost-effective manner while maximizing their impact and reach. In this article, we introduce the project submissions to the **EcoHack: AI & LLM Hackathon for Applications in Evidence-based Ecological Research & Practice**, detailing the broad classes of ecological challenges addressed by teams and analyzing trends in their approaches. We then present each team submission, along with a summary table listing team members and links to code repositories where available. Finally, we include the detailed project descriptions submitted by each team, showcasing the depth and breadth of innovation demonstrated during the hackathon.

EcoHack Hackathon Event Overview

From January 20th to 22nd, 2025, we hosted the inaugural edition of EcoHack 2025, an AI & LLM hackathon dedicated to advancing evidence-based ecological research and practice. The event convened students and researchers from computer science and ecology, fostering interdisciplinary collaboration across virtual and in-person formats. Hosted at the ZiF Campus at Bielefeld University, EcoHack was organized as part of the ZiF resident group Mapping Evidence to Theory in Ecology: Addressing the Challenges of Generalization and Causality, convened by Tina Heger. The event attracted 30 participants, primarily from Germany, alongside attendees from Spain, the UK, and Poland.

The event commenced with a series of welcome talks providing participants with the broader research context. Tina Heger introduced the objectives of the ZiF resident group (video link), Birgitta König-Ries presented the planned EcoWeaver Toolkit (video link), and Jennifer D'Souza, the EcoHack Convener, delivered an introductory talk on LLM-based search systems, *Long-Term Solutions vs. Hackathon Prototypes* (video link). Following these presentations, participants self-organized into teams or chose to work individually, dedicating two days to prototyping solutions.

To support participants, we provided pre-event resources via the EcoHack mailing list. One of which were Miro boards for collaborative brainstorming and idea development. Two Miro boards were made available: the Hackathon Starter Kit, offering new participants guidance on hackathon structure, team formation, and ideation strategies, and the Hackathon Collaboration and Ideation board, where participants could propose ideas and form teams. To facilitate communication, we established a dedicated Element channel featuring spaces for Introductions, Resources, Announcements, and general discussions. Participants could also directly message others to form teams. Additionally, we released a resource guide with potential project ideas. The challenge was intentionally open-ended, encouraging participants to explore a wide range of ecological applications using open-source, state-of-the-art LLMs to develop innovative, impactful, and scalable solutions.

The event culminated on January 22nd with a lightning talk session, where teams presented their prototypes in concise two-minute pitches. This was followed by an interactive demonstration session, allowing attendees to engage directly with project teams and explore their solutions. The prize evaluation process

Figure 1: Feedback collected on Mentimeter from the EcoHack participants to the question "Tell us about your hackathon experience".

spanned two weeks, during which organizers assessed submissions. On February 7th, 2025, the awards ceremony was conducted virtually via Zoom, accompanied by an interactive feedback session using Mentimeter. Figure 1 presents a word cloud summarizing participants' reflections, with full feedback available here.

EcoHack 2025 successfully integrated in-person engagement with remote participation, fostering an inclusive, interdisciplinary event that transcended geographical boundaries. The hackathon's dedicated repository is accessible on GitHub.

Overview of Submissions

The hackathon resulted in 11 team submissions, categorized as shown in Table 1. From these submissions, we identified six key application areas:

1. **AI-Enhanced Decision Support and Automation**: AI tools that facilitate decision-making and automate processes in ecological research and conservation. These systems help researchers, practitioners, and governmental organizations monitor, manage, and analyze ecological data more efficiently through dashboards, workflow automation, and intelligent interfaces.

2. **Scientific Search and Communication**: AI-powered systems that enhance the accessibility and communication of scientific knowledge, particularly in ecological and agricultural contexts. These tools bridge the gap between farmers, researchers, policymakers, and practitioners by making ecological insights more understandable and actionable.

3. **Knowledge Extraction and Reasoning**: AI tools designed to structure, extract, and recommend ecological knowledge, supporting scientific discovery and decision-making. These systems enable researchers to navigate complex information landscapes by identifying relationships, concepts, and relevant research.

4. **AI for Ecological Modeling, Forecasting, and Simulation**: AI-driven models that simulate ecological processes, predict environmental changes, and support ecosystem recovery planning. These tools assist researchers and other relevant stakeholders in understanding future ecological trends, assessing risks, and optimizing restoration strategies.

5. **Causal Inference and Ecological Reasoning**: AI models that support causal reasoning in ecological studies by distinguishing correlation from causation. These tools improve our ability to understand complex ecological interactions and validate hypotheses about environmental systems.

6. **AI for Biodiversity Monitoring and Conservation**: AI applications that aid in wildlife monitoring, species conservation, and environmental protection. These tools use machine learning, bioacoustics, and computer vision to analyze ecological data and support conservation efforts.

We next discuss each application area in more detail and highlight exemplar projects in each.

1. AI-Enhanced Decision Support and Automation

The central theme here is the application of AI to create **prescriptive analysis** tools that not only interpret past and present data (as in descriptive and predictive analytics) but also suggest specific actions or policies to achieve desired outcomes. In terms of the nuts and bolts of such a system, **AI-enhanced decision support and automation in spatial planning** could rely on integrating heterogeneous datasets, including global biodiversity databases (e.g., GBIF), remote sensing and land-use data, socio-economic datasets, and spatial metrics to model ecosystem dynamics and human-environment interactions. Various AI methodologies—such as greedy algorithms, metaheuristics, mixed-integer linear programming (MILP), constraint programming (CP), and reinforcement learning—can be applied to process and optimize data-driven decisions, each balancing efficiency, interpretability, and scalability. While **black-box machine learning models** automate predictive insights, **symbolic AI approaches** like CP and MILP enable transparent, constraint-driven optimization, making them particularly relevant for policy-driven and stakeholder-engaged decision processes [4].

Beyond spatial planning, AI-driven innovations leveraging machine learning (ML), deep learning (DL), computer vision (CV), and natural language processing (NLP) enable real-time environmental data collection, analysis, and action. These tools enhance forecasting models, improve data collection via IoT sensor networks, and facilitate decision-making in areas such as **precision agriculture, climate change mitigation, and wildlife conservation** [5]. By identifying critical areas requiring intervention, AI enhances **precision conservation**, shifting strategies from reactive to proactive management [6]. However, while AI presents numerous advantages in ecological contexts, challenges related to **data privacy, ethical implications, and over-reliance on technology** must be addressed to ensure its sustainable and responsible use in environmental monitoring.

Exemplar project: One submission could be categorized here. The **EcoSmile** one-person team (Jalloul, section 1) developed the EcoGuard Insights Dashboard as a web-based decision-support system designed to integrate and analyze decentralized environmental datasets, addressing deforestation, carbon emissions, and biodiversity loss. Built using Python and Streamlit, the system processes heterogeneous data sources (e.g., Global Forest Watch, Map of Life) through data aggregation and statistical analysis with pandas and NumPy. Geospatial mapping, implemented via folium, enables interactive visualization of forest loss and biodiversity hotspots, while time-series forecasting models, leveraging scikit-learn regression techniques, predict deforestation trends up to 2046 based on historical data (2001–2023). Dynamic visualizations powered by matplotlib and Plotly provide real-time analytical updates, equipping policymakers and researchers with actionable insights. With its modular and scalable architecture, EcoGuard facilitates evidence-based decision-making, bridging the gap between complex environmental data and practical conservation strategies.

2. Scientific Search and Communication

AI-powered systems are transforming scientific search and communication, particularly in ecological and agricultural contexts. Broadly, LLM-based techniques make it possible to obtain highly **condensed scientific knowledge summaries** of research literature, improving knowledge accessibility [7]. Automated AI-driven knowledge synthesis extends beyond summarization to compiling **structured overviews of key ecological research aspects**, ensuring insights remain contextually rich [8, 9]. Enhancing search efficiency, AI-driven scientific search engines use deep learning-based ranking methods, such as bi-encoders and retrieval-based architectures, to improve literature discovery while mitigating biases in traditional search systems [10, 11, 12]. Beyond retrieval, AI also facilitates clearer and more **inclusive communication of scientific knowledge**. GenAI tools assist non-native English speakers in producing coherent academic texts and generating research summaries, abstracts, and social media posts to improve accessibility and policy engagement [13]. AI-powered real-time interpretation and subtitles further promote inclusivity in scientific discussions by enabling broader participation [14]. Additionally, automated slide-generation techniques like DOC2PPT [15] streamline research dissemination by converting scientific papers into structured presentations, making **complex insights more accessible to diverse audiences**.

Despite these advancements, ensuring the accuracy, transparency, and ethical use of AI-generated content remains a challenge. Risks such as **misinformation and fraud** must be mitigated, for instance in use of AI to generate multimedia from biodiversity surveys and citizen science projects [16]. AI interpretation technologies require further refinement to **support diverse languages and dialects** [14]. Algorithmic biases in AI-driven search and summarization tools necessitate continuous evaluation and improvements [17, 18]. Additionally, concerns around data privacy and technological readiness must be addressed to build trust in AI applications for scientific communication [19, 20]. Collaboration between AI developers and researchers is crucial to overcoming these challenges and ensuring AI-powered tools equitably enhance scientific discourse, knowledge dissemination, and decision-making across ecological and agricultural domains.

Exemplar projects: In practice, researchers and developers are actively exploring ways to refine and apply AI technologies, as seen in EcoHack's four hackathon projects. The team behind **AutoDeck-AI** (Jadhav et al., section 2) developed an AI-powered slide generator designed to streamline scientific communication in ecology by automatically extracting figures, tables, and key insights from PDFs using PyPDF2, PyMuPDF, and GPT-4o to generate structured presentations tailored for researchers, practitioners, and funding bodies. Meanwhile, the team working on **Agri-Chatbot** (Kommineni et al., section 3) tackled the challenge of making agricultural research more accessible to farmers by integrating retrieval-augmented generation (RAG) [10] with a FAISS vector store containing over 500,000 structured data entries, leveraging LLaMA 3.3-70B to generate precise, science-backed responses. At the same time, **FarmGuide** (Viso et al., section 4) sought to bridge the gap between ecological research and practical farming by using LLMs (GPT-4, Llama3.1 via Ollama) to extract relevant insights from scientific literature, ranking articles based on all-MiniLM-v6 embeddings, and providing an interactive chatbot interface for personalized guidance on crop protection and pollination strategies. In a different vein, **DiversiTeam** (Bachinger et al., section 5) introduced EcoSearch, a tool that enhances literature discovery in ecology by re-ranking search results based on the geographical distribution of authors, fostering a more inclusive and representative scientific discourse. Collectively, these projects demonstrate the power of AI to enhance scientific communication, facilitate knowledge transfer, and support data-driven decision-making across ecological and agricultural domains.

3. Knowledge Extraction and Reasoning

Tapping into LLMs' potential for **expert-level scientific creativity** requires proficiency in specialized knowledge and deductive reasoning [21]. One approach to enhancing this capability is the construction of ontological knowledge graphs from scientific literature, which uncover interdisciplinary relationships, aid researchers in navigating complex information landscapes, and advance discovery. These graphs leverage transitive and isomorphic properties to reveal novel connections, enabling tasks like query answering, identifying knowledge gaps, proposing new material designs, and predicting material behaviors. **Deep node embeddings** facilitate combinatorial node similarity ranking, using path sampling strategies to link previously unrelated concepts [22]. Ontologies serve as interdisciplinary bridges, employing cognitive modeling to analyze causal relationships and visualize structured knowledge, integrating research across energy, ecology, and quality of life [23]. Ontologies and Bayesian networks also support the integration and interpretation of heterogeneous ecological data, aiding in the discovery of **ecological interactions**, such as behavioral relationships between individual plants and insects and their population-level consequences [24]. While AI tools enhance ecological knowledge extraction, their anthropocentric design risks overlooking the interconnectedness of human and ecological systems. Aligning AI development with **ecological thinking** is recommended for fostering a more sustainable human-environment relationship and eco-centric reasoning [25].

Exemplar project: There was one EcoHack project in this category with a unique search application to help reduce manual effort in systematic reviews. The team behind **MatchBox** (Brinner et al., section 6), focused on improving concept extraction and ontology mapping in ecological research using LLMs and transformer-based models. It employed Llama-3-8B-Instruct [26] for concept extraction from abstracts, DeBERTa [27] for token-level classification and semantic similarity scoring, and a reranker model for refining ontology alignment. The system integrates ENVO and INBIO ontologies, providing an interactive tool for highlighting key terms, linking them to ontology definitions, and enabling concept-based literature search via similarity embeddings rather than keyword matching.

4. AI for Ecological Modeling, Forecasting, and Simulation

Multiagent systems (MAS) with AI and model-driven engineering offer a powerful framework for modeling and verifying ecosystem resilience dynamics. These systems synthesize heterogeneous ecological data and employ adaptable neural network architectures to simulate complex interactions between species, environmental factors, and stakeholders, enhancing predictive modeling for ecosystem management [28]. However, a key challenge lies in addressing AI's lack of robustness and generalization across diverse ecological contexts, necessitating a more purposeful synergy between AI and ecological resilience research [29]. Advances in MAS have also led to the development of open-source ecosystem simulators, such as Ecotwin [30], which leverage game engines like Unity to model ecosystems containing inanimate objects, flora, and fauna. These simulators serve as valuable tools for predicting the consequences of human interventions, training AI agents in dynamic environments, and improving our understanding of ecological systems. Beyond modeling ecological processes, ecological principles themselves are influencing AI development, as seen in collaborative multi-agent systems inspired by mutualistic interactions in nature [31]. Techniques like multi-agent reinforcement learning (MARL) are also being applied to simulate complex socio-ecological interactions, contributing to sustainability research and insights into human decision-making [32].

Beyond MAS, vision-based deep learning models enhance ecological forecasting. Deepbiosphere integrates remote sensing, citizen science, and convolutional neural networks to track plant community shifts [33]. Satellite platforms like Landsat and Sentinel-2, combined with Vision Transformers and Generative Adversarial Networks, improve biodiversity and vegetation monitoring [34, 35]. These AI-driven tools support applications from post-fire recovery to species distribution modeling [36, 37], though challenges remain in data quality, model interpretability, and ecosystem complexity [38].

Exemplar projects: There were two EcoHack projects in this category. **BioSim** (Ling & Miao, section 8) explores the challenge of biological invasions through a multi-agent simulation framework that integrates LLMs with agent-based modeling. By enabling species agents to interact dynamically based on ecological principles, real-time literature integration, and structured input data, BioSim enhances scalability and adaptability in predicting invasion dynamics. The framework employs visualization tools and multi-agent interactions to model species competition and conservation scenarios. On the other hand, **HealingFactor** (Krutsylo) addresses a different ecological issue—forecasting ecosystem recovery in war-affected regions of Ukraine using Sentinel-2 satellite imagery accessed via Microsoft Planetary Computer. The system applies a convolutional long short-term memory model, utilizing spectral bands to capture spatial and temporal vegetation patterns and predict health trends. While BioSim simulates species interactions through AI-driven multi-agent modeling, HealingFactor leverages remote sensing and deep learning to forecast ecological recovery. Together, they highlight the versatility of AI in addressing ecological challenges, from species-level dynamics to landscape-scale restoration.

5. Causal Inference and Ecological Reasoning

In ecological studies, identifying causal relationships is crucial for understanding how **environmental factors influence ecosystem structure and function**. AI-driven causal inference models, such as **Structural Causal Models (SCMs)** and **Bayesian Networks (BNs)**, have been applied to infer causal dependencies in biodiversity and ecosystem productivity, helping researchers understand how **species richness affects ecosystem resilience** [39]. SCMs have also identified temperature, salinity, and pH as key causal drivers of chlorophyll concentration in marine ecosystems [40] and soil temperature as the dominant factor in wetland CH_4 emissions [41]. BNs, in particular, have been used to explore causal relationships between **forest aboveground biomass** and its potential driving factors, offering a probabilistic framework to handle uncertainty and complexity in ecological systems [42]. They are also applied to model uncertainty, uncovering drivers of cyanobacteria blooms [43] and linking land-cover changes to bird abundance [44]. Additionally, machine learning-based causal discovery, such as Granger causality and EcohNet, enhances **forecasting of ecosystem components** [43], while PCMCI and Optimal Information Flow (OIF) models infer long-term causal dependencies in **climate-ecosystem interactions** [45, 46].

A new avenue of exploration is training LLMs to engage in causal reasoning about ecological systems. Current LLMs primarily capture statistical associations, but benchmarks designed to evaluate their ability to distinguish causal from correlational reasoning would advance their applicability in ecological research. By integrating AI-driven causal inference with **LLM-based hypothesis generation and validation**, researchers can develop more robust, scalable tools for analyzing ecological interactions and predicting environmental change.

Exemplar project: In an individual contribution, the **EcoLogic** (Heider, section 10) project introduced an algorithm to build a benchmark to assess LLMs' ability to distinguish causal from correlational reasoning in ecological systems. Using food-web-based tasks derived from the Global Biotic Interactions (GloBi) dataset, it evaluated models on causal, correlational, and mixed reasoning. An automated solver, based on discrete simulations of predator-prey dynamics, generated ground truth data, ensuring scalable assessment. Preliminary results showed that frontier models excel, while smaller models struggled with tasks like interpreting Latin animal names.

6. AI for Biodiversity Monitoring and Conservation

AI is transforming biodiversity monitoring and conservation through automated species identification, real-time monitoring, acoustic detection, and ecological forecasting. Machine learning, computer vision, and bioacoustics enable efficient analysis of ecological datasets, improving conservation strategies [47]. Large datasets like BioTrove (161 million images, 366,600 species) and the iNaturalist Species Classification and Detection Dataset (859,000 images, 5,000 species) support AI-driven biodiversity research [48, 49]. Deep learning models such as YOLO [50] and Phi-3.5-vision-instruct achieve high accuracy in species detection [51, 52], while AI-powered alert systems like TrailGuard detect poachers and habitat disturbances [53]. AI-driven bioacoustic monitoring enables non-invasive species tracking, with tools like Haikubox identifying bird species through their songs [54]. CNN-based acoustic monitoring detects species like the marbled murrelet, and AI-integrated UAVs track endangered species such as black rhinos in remote environments [55, 56]. AI also supports habitat mapping and restoration, analyzing satellite imagery to assess environmental changes and predict biodiversity threats, such as species population declines and habitat loss [57].

Despite advancements, challenges persist in data quality, model biases, and ethical concerns. Detecting smaller species with drone-based AI requires advanced image processing [56], while privacy and security in real-time monitoring demand careful consideration [58]. Nonetheless, AI continues to improve biodiversity research, offering scalable and efficient solutions for species conservation and ecosystem management.

Exemplar project: One EcoHack team belongs to this category. The BirdTeam (Mendu et al., section 11) developed a machine learning model to classify bird alarm calls as indicators of stress from anthropogenic disturbances. Using the BirdCLEF 2022 dataset, they fine-tuned the BirdNet residual neural network, incorporating pre-processing, four residual stacks, and a classification block. Their classifier was claimed to achieve over 90% AUPRC and AUROC on the training set.

Reflections and Lessons Learned

For most of the organizers, EcoHack was their first experience hosting an event of this kind. Reflecting on our journey, we summarize below the key lessons learned, with the hope that these insights will support and inspire other first-time hackathon organizers:

- **Pre-event preparation was valuable, but early group formation proved challenging.** While participants made extensive use of the resources we provided in advance—and found them highly useful—our expectation that teams would form prior to the event proved overly optimistic. Many individuals did come with concrete project ideas, which served as natural starting points on the first day. In hindsight, hosting a virtual meet-and-greet or ideation session one to two weeks before the event might have encouraged earlier team formation and topic alignment.

- **The hybrid format worked well, but time zone differences posed challenges.** EcoHack functioned effectively as a hybrid event, enabling both in-person and remote participation. However, coordinating across time zones proved more difficult than expected and may have contributed to participant attrition. To enable more inclusive participation, future events could consider running parallel sessions by time zone, scheduling staggered working days, or focusing on a specific set of time zones while communicating this clearly in advance.
- **Access to domain experts was highly beneficial—but underutilized.** Co-location with a workshop by the ZiF resident group provided access to several ecologists who engaged with participants. These interactions were especially valuable and helped shape several solutions. Nevertheless, not all teams took advantage of this opportunity. To encourage such engagement, organizers might consider integrating structured formats—such as expert Q&A sessions, mentoring hours, or facilitated discussion slots—into the event agenda.
- **There is room to broaden the event's thematic scope.** This edition of EcoHack was strongly focused on coding and technical development. Future editions could explore a broader thematic range, including conceptual, theoretical, or policy-oriented work. Expanding the scope could help attract participants from diverse backgrounds and skill sets, enriching the event's interdisciplinary character.
- **The effort was significant—but overwhelmingly worthwhile.** Perhaps the most important lesson was that organizing EcoHack was genuinely worth the effort. While the workload—before, during, and after the event—should not be underestimated, the experience, participant engagement, and quality of outcomes far exceeded our expectations.

Discussion: Hackathons as a Catalyst for SDG Innovation

The United Nations Sustainable Development Goals (SDGs) provide a global framework for addressing critical challenges, including climate change, biodiversity loss, and environmental sustainability [59]. Events like the EcoHack serve as incubators for translating these ambitious objectives into actionable solutions, leveraging AI-driven innovations to enhance decision-making, knowledge extraction, and ecological monitoring. In particular, SDG 13 (Climate Action) [60], SDG 14 (Life Below Water) [61], and SDG 15 (Life on Land) [62] underscore the urgency of protecting ecosystems and mitigating environmental risks. AI-powered modeling and forecasting are already transforming climate adaptation strategies, sustainable marine management, and terrestrial conservation by enabling more precise predictions, automation, and large-scale data analysis. Beyond ecological goals, AI applications also intersect with SDG 2 (Zero Hunger) [63], through advancements in sustainable agriculture, and SDG 4 (Quality Education) [64], by enhancing accessibility to scientific knowledge. Hackathons like EcoHack create a collaborative space where interdisciplinary teams can prototype AI solutions that bridge research and real-world implementation, accelerating progress toward sustainability targets. By fostering innovation and cross-sectoral collaboration, such events not only generate novel technological approaches but also cultivate a mindset of applied problem-solving, positioning AI as a key enabler of sustainable development.

Conclusion

The **2025 EcoHack** highlighted the transformative role of AI and LLMs in ecological research, with participants developing innovative solutions across six key areas, including decision support, scientific search, biodiversity monitoring, and causal inference. Projects leveraged diverse AI methodologies, from retrieval-augmented generation and knowledge graph-based reasoning to multi-agent simulations and deep learning models for ecological forecasting. Approaches such as natural language processing for literature synthesis, Bayesian networks for causal inference, and computer vision for species monitoring showcased AI's versatility in addressing complex ecological questions.

The hackathon's hybrid format fostered interdisciplinary collaboration, uniting researchers across ecology and AI to explore new ways of integrating machine learning into ecological research and practice. By applying AI-driven techniques to ecological modeling, forecasting, and knowledge extraction, teams demonstrated how computational methods can bridge scientific insights with real-world conservation and management efforts. As AI continues to integrate into ecology, **EcoHack** reinforces the value of collaborative hackathons in driving innovation, fostering scientific partnerships, and accelerating the development of AI-powered solutions for evidence-based ecological research and decision-making.

Table 1: Overview of the tools developed by the various teams, and links to source code repositories. Full descriptions of the projects can be found in the appendix.

Project	Authors	Links
AI-Enhanced Decision Support and Automation		
EcoSmile at EcoHack-2025: EcoGuard Insights Dashboard for Planetary Preservation	Basma Jalloul	GitHub
Scientific Search and Communication		
AutoDeck-AI at EcoHack-2025: Eco-Centric Slide Generator	Hrishikesh Jadhav, Javad Razavian, Moiz Khan Sherwani	GitHub
Agri Chatbot: From Science to Soil	Vamsi Krishna Kommineni, Anne Peter, Caren Daniel, Alexander Espig	GitHub
FarmGuide: A bridge between scientists and farmers for natural agriculture practices	Bartolome Ortiz Viso, Lorenz Gunreben, Mir Nafis Sharear Shopnil, Nayanika Das	GitHub
DiversiTeam at EcoHack-2025: EcoSearch	Sarah T. Bachinger, Daphne Frederike Auer, Edward Gow-Smith	GitHub
Knowledge Extraction and Reasoning		
Matchbox at EcoHack-2025: Empowering Ecology Research with Efficient Concept Mapping	Marc Brinner, Nadeen Fathallah, Tarek Al Mustafa	GitHub
EcoSci Recommender	Samira Korani	GitHub
AI for Ecological Modeling, Forecasting, and Simulation		
BioSim: A Multi-Agent Framework for Biological Invasion Simulation	Zijian Ling, Shuhan Miao	GitHub
Healing Factor at EcoHack-2025: Forecasting Ecosystem Recovery Efforts in Ukraine	Andrii Krutsylo	GitHub
Causal Inference and Ecological Reasoning		
EcoLogic: A Benchmark for Causal and Correlational Reasoning of LLMs Based on Ecological Interactions	Nico Heider	GitHub
AI for Biodiversity Monitoring and Conservation		
BirdTeam: Bird Alarm Call Classifier	Vaishnavi Mendu, Moritz Plenz, Will Woof	Hugging-Face

Acknowledgments

This event was sponsored by the Center for Interdisciplinary Research (ZiF: Zentrum für interdisziplinäre Forschung), Bielefeld University's Institute for Advanced Study. More details are available at https://www.uni-bielefeld.de/einrichtungen/zif/. The event planning was supported by the ZiF research group Mapping Evidence to Theory in Ecology, led by PI Tina Heger (https://www.uni-bielefeld.de/einrichtungen/zif/groups/ongoing/mapping-evidence/).

References

[1] Alexander Nolte, Linda Bailey Hayden, and James D. Herbsleb. "How to Support Newcomers in Scientific Hackathons - An Action Research Study on Expert Mentoring". In: *Proc. ACM Hum.-Comput. Interact.* 4.CSCW1 (May 2020). DOI: 10.1145/3392830. URL: https://doi.org/10.1145/3392830.

[2] Ei Pa Pa Pe-Than and James D. Herbsleb. "Understanding Hackathons for Science: Collaboration, Affordances, and Outcomes". In: *Information in Contemporary Society*. Ed. by Natalie Greene Taylor et al. Cham: Springer International Publishing, 2019, pp. 27–37. ISBN: 978-3-030-15742-5.

[3] Ben Heller et al. "Hack your organizational innovation: literature review and integrative model for running hackathons". In: *Journal of Innovation and Entrepreneurship* 12.1 (2023), p. 6.

[4] Dimitri Justeau-Allaire. "AI and spatial planning for sustainable socio-ecosystems". In: *Thirty-Second International Joint Conference on Artificial Intelligence {IJCAI-23}*. International Joint Conferences on Artificial Intelligence Organization. 2023, pp. 6370–6377.

[5] Froilan Delute Mobo, Ana Liza Reclozado Garcia, and Katarzyna Miłek. "Leveraging AI for Real-Time Environmental Monitoring: Innovations and Impacts". In: *Harnessing AI in Geospatial Technology for Environmental Monitoring and Management*. IGI Global Scientific Publishing, 2025, pp. 201–212.

[6] Dwijendra Nath Dwivedi, Ghanashyama Mahanty, and Varunendra nath Dwivedi. "Intelligent conservation: a comprehensive study on AI-enhanced environmental monitoring and preservation". In: *The Convergence of Self-Sustaining Systems With AI and IoT*. IGI Global, 2024, pp. 215–226.

[7] Isabel Cachola et al. "TLDR: Extreme Summarization of Scientific Documents". In: *Findings of the Association for Computational Linguistics: EMNLP 2020*. 2020, pp. 4766–4777.

[8] Sruthi M Krishna Moorthy et al. "Harnessing Large Language Models for Ecological Literature Reviews: A Practical Pipeline". In: (2025).

[9] Jennifer D'Souza et al. "Mining for Species, Locations, Habitats, and Ecosystems from Scientific Papers in Invasion Biology: A Large-Scale Exploratory Study with Large Language Models". In: *arXiv preprint arXiv:2501.18287* (2025).

[10] Patrick Lewis et al. "Retrieval-augmented generation for knowledge-intensive nlp tasks". In: *Advances in neural information processing systems* 33 (2020), pp. 9459–9474.

[11] Taylor Shin et al. "AutoPrompt: Eliciting Knowledge from Language Models with Automatically Generated Prompts". In: *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. 2020, pp. 4222–4235.

[12] Wang-Chiew Tan et al. "Reimagining Retrieval Augmented Language Models for Answering Queries". In: *Findings of the Association for Computational Linguistics: ACL 2023*. Ed. by Anna Rogers, Jordan Boyd-Graber, and Naoaki Okazaki. Toronto, Canada: Association for Computational Linguistics, July 2023, pp. 6131–6146. DOI: 10.18653/v1/2023.findings-acl.382. URL: https://aclanthology.org/2023.findings-acl.382/.

[13] Rafael D Zenni and Nigel R Andrew. "Artificial Intelligence text generators for overcoming language barriers in ecological research communication". In: *Austral Ecology* 48.7 (2023), pp. 1225–1229.

[14] Heather Fair et al. "Can AI interpretation increase inclusivity?" In: *Frontiers in Ecology and the Environment* 22.10 (2024), e2821. DOI: https://doi.org/10.1002/fee.2821. eprint: https://esajournals.onlinelibrary.wiley.com/doi/pdf/10.1002/fee.2821. URL: https://esajournals.onlinelibrary.wiley.com/doi/abs/10.1002/fee.2821.

[15] Tsu-Jui Fu et al. “DOC2PPT: Automatic Presentation Slides Generation from Scientific Documents”. In: *Proceedings of the AAAI Conference on Artificial Intelligence* 36.1 (June 2022), pp. 634–642. DOI: 10.1609/aaai.v36i1.19943. URL: https://ojs.aaai.org/index.php/AAAI/article/view/19943.

[16] Matthias C Rillig et al. “How widespread use of generative AI for images and video can affect the environment and the science of ecology”. In: *Ecology Letters* 27.3 (2024), e14397.

[17] Shalece Kohnke and Tiffanie Zaugg. “Artificial Intelligence: An Untapped Opportunity for Equity and Access in STEM Education”. In: *Education Sciences* 15.1 (2025). ISSN: 2227-7102. DOI: 10.3390/educsci15010068. URL: https://www.mdpi.com/2227-7102/15/1/68.

[18] Juan Sebastián Laverde Gonzalez and Lay Aracely Rodríguez Hernández. “AI and education: combination to enhance knowledge”. In: *Seminars in Medical Writing and Education*. Vol. 4. AG Editor. 2025, p. 4.

[19] Enkelejda Kasneci et al. “ChatGPT for good? On opportunities and challenges of large language models for education”. In: *Learning and Individual Differences* 103 (2023), p. 102274. ISSN: 1041-6080. DOI: https://doi.org/10.1016/j.lindif.2023.102274. URL: https://www.sciencedirect.com/science/article/pii/S1041608023000195.

[20] Lixiang Yan et al. “Practical and ethical challenges of large language models in education: A systematic scoping review”. In: *British Journal of Educational Technology* 55.1 (2024), pp. 90–112.

[21] Anirban Mukherjee and Hannah Hanwen Chang. “AI Knowledge and Reasoning: Emulating Expert Creativity in Scientific Research”. In: *arXiv preprint arXiv:2404.04436* (2024).

[22] Markus J Buehler. “Accelerating scientific discovery with generative knowledge extraction, graph-based representation, and multimodal intelligent graph reasoning”. In: *Machine Learning: Science and Technology* 5.3 (2024), p. 035083.

[23] Tatyana Vorozhtsova, Dmitry Pesterev, and Gleb An. “Using intelligent technologies for knowledge formation in research on the impact of power industry on ecology and quality of life”. In: *E3S Web of Conferences*. Vol. 209. EDP Sciences. 2020, p. 02032.

[24] Willem Coetzer, Deshendran Moodley, and Aurona Gerber. “A knowledge-based system for generating interaction networks from ecological data”. In: *Data & Knowledge Engineering* 112 (2017), pp. 55–78. ISSN: 0169-023X. DOI: https://doi.org/10.1016/j.datak.2017.09.005. URL: https://www.sciencedirect.com/science/article/pii/S0169023X17300459.

[25] Chunchen Xu and Xiao Ge. “AI as a Child of Mother Earth: Regrounding Human-AI Interaction in Ecological Thinking”. In: *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*. CHI EA ’24. Honolulu, HI, USA: Association for Computing Machinery, 2024. ISBN: 9798400703317. DOI: 10.1145/3613905.3644065. URL: https://doi.org/10.1145/3613905.3644065.

[26] Aaron Grattafiori et al. “The llama 3 herd of models”. In: *arXiv preprint arXiv:2407.21783* (2024).

[27] Pengcheng He et al. “Deberta: Decoding-enhanced bert with disentangled attention”. In: *arXiv preprint arXiv:2006.03654* (2020).

[28] Tiago Sousa. “Towards Modeling and Predicting the Resilience of Ecosystems”. In: *2023 ACM/IEEE International Conference on Model Driven Engineering Languages and Systems Companion (MODELS-C)*. Västerås, Sweden: IEEE Press, 2023, pp. 159–165. DOI: 10.1109/MODELS-C59198.2023.00042. URL: https://doi.org/10.1109/MODELS-C59198.2023.00042.

[29] Barbara A. Han et al. “A synergistic future for AI and ecology”. In: *Proceedings of the National Academy of Sciences* 120.38 (2023), e2220283120. DOI: 10.1073/pnas.2220283120. eprint: https://www.pnas.org/doi/pdf/10.1073/pnas.2220283120. URL: https://www.pnas.org/doi/abs/10.1073/pnas.2220283120.

[30] Claes Strannegård et al. “The Ecosystem Path to General AI”. In: *arXiv preprint arXiv:2108.07578* (2021).

[31] Alexander A Nguyen, Faryar Jabbari, and Magnus Egerstedt. “Mutualistic interactions in heterogeneous multi-agent systems”. In: *2023 62nd IEEE Conference on Decision and Control (CDC)*. IEEE. 2023, pp. 411–418.

[32] Uri Hertz et al. "Beyond the matrix: Experimental approaches to studying cognitive agents in social-ecological systems". In: *Cognition* 254 (2025), p. 105993.

[33] Lauren Gillespie, Megan Ruffley, and Moisés Expósito-Alonso. "An image is worth a thousand species: mapping plant biodiversity with citizen science, remote sensing, and deep learning". In: *bioRxiv* (2022). DOI: 10.1101/2022.08.16.504150. URL: https://doi.org/10.1101/2022.08.16.504150.

[34] Stephanie A Insalaco et al. "Monitoring an Ecosystem in Crisis: Measuring Seagrass Meadow Loss Using Deep Learning in Mosquito Lagoon, Florida". In: *Photogrammetric Engineering & Remote Sensing* 90.6 (2024), pp. 363–370.

[35] Jorge Rodríguez et al. "Biodiversity Assessment in Drylands Using Augmented Satellite Imagery Through Deep Learning Models". In: *EGU General Assembly Conference Abstracts*. 2024, p. 20836.

[36] Joaquim Estopinan et al. "Modelling species distributions with deep learning to predict plant extinction risk and assess climate change impacts". In: *arXiv preprint arXiv:2401.05470* (2024).

[37] R Shanmuga Priya and K Vani. "Vegetation change detection and recovery assessment based on post-fire satellite imagery using deep learning". In: *Scientific Reports* 14.1 (2024), p. 12611.

[38] Nathalie Pettorelli et al. "Deep learning and satellite remote sensing for biodiversity monitoring and conservation". In: *Remote Sensing in Ecology and Conservation* (2024).

[39] Jakob Runge. "Modern causal inference approaches to investigate biodiversity-ecosystem functioning relationships". In: *nature communications* 14.1 (2023), p. 1917.

[40] Želimir Kurtanjek. "Kauzalni ekološki model sjevernog Jadrana temeljem podataka EU projekta "LTER Northern Adriatic Sea"". In: *Kemija u industriji: Časopis kemičara i kemijskih inženjera Hrvatske* 71.11-12 (2022), pp. 711–718.

[41] Kunxiaojia Yuan et al. "Causality guided machine learning model on wetland CH4 emissions across global wetlands". In: *Agricultural and Forest Meteorology* 324 (2022), p. 109115.

[42] Yubo Tao et al. "Application of Bayesian Causal Inference in the Study of the Relationship Between Biodiversity and Aboveground Biomass of Subtropical Forest in Eastern China". In: *Forests* 15.11 (2024), p. 1841.

[43] Kenta Suzuki, Shin-ichiro S Matsuzaki, and Hiroshi Masuya. "Decomposing predictability to identify dominant causal drivers in complex ecosystems". In: *Proceedings of the National Academy of Sciences* 119.42 (2022), e2204405119.

[44] Daria Bystrova et al. "Causal discovery from ecological time-series with one timestamp and multiple observations". In: *bioRxiv* (2024), pp. 2024–10.

[45] C. Krich et al. "Estimating causal networks in biosphere–atmosphere interaction with the PCMCI approach". In: *Biogeosciences* 17.4 (2020), pp. 1033–1061. DOI: 10.5194/bg-17-1033-2020. URL: https://bg.copernicus.org/articles/17/1033/2020/.

[46] Jie Li and Matteo Convertino. "Inferring ecosystem networks as information flows". In: *Scientific reports* 11.1 (2021), p. 7094.

[47] Victoria Bukky Ayoola et al. "The role of big data and AI in enhancing biodiversity conservation and resource management in the USA". In: *World Journal of Advanced Research and Reviews* 23.2 (2024), pp. 1851–1873.

[48] Chih-Hsuan Yang et al. "BioTrove: A Large Curated Image Dataset Enabling AI for Biodiversity". In: *Advances in Neural Information Processing Systems* 37 (2024), pp. 102101–102120.

[49] Grant Van Horn et al. "The INaturalist Species Classification and Detection Dataset". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. June 2018.

[50] Joseph Redmon et al. "You only look once: Unified, real-time object detection". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016, pp. 779–788.

[51] Paul Fergus et al. "Towards context-rich automated biodiversity assessments: deriving AI-powered insights from camera trap data". In: *Sensors* 24.24 (2024), p. 8122.

[52] Paul Fergus et al. "Harnessing Artificial Intelligence for Wildlife Conservation". In: *Conservation* 4.4 (2024), pp. 685–702.

[53] Jeremy S Dertien et al. "Mitigating human–wildlife conflict and monitoring endangered tigers using a real-time camera-based alert system". In: *BioScience* 73.10 (2023), pp. 748–757.

[54] Brian L Sullivan et al. "eBird: A citizen-based bird observation network in the biological sciences". In: *Biological conservation* 142.10 (2009), pp. 2282–2292.

[55] Adam Duarte et al. "Passive acoustic monitoring and convolutional neural networks facilitate high-resolution and broadscale monitoring of a threatened species". In: *Ecological Indicators* 162 (2024), p. 112016.

[56] Alice Hua et al. "Protecting endangered megafauna through AI analysis of drone images in a low-connectivity setting: a case study from Namibia". In: *PeerJ* 10 (2022), e13779.

[57] Daniele Silvestro et al. "Improving biodiversity protection through artificial intelligence". In: *Nature sustainability* 5.5 (2022), pp. 415–424.

[58] Prasenjit Mitra et al. "Info-Wild: Knowledge Extraction and Management for Wildlife Conservation". In: *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management*. 2023, pp. 5281–5284.

[59] United Nations General Assembly. *Transforming our world: the 2030 Agenda for Sustainable Development*. A/RES/70/1. 2015. URL: https://sdgs.un.org/2030agenda.

[60] United Nations. *Sustainable Development Goal 13: Climate Action*. Accessed: 2025-03-05. 2015. URL: https://sdgs.un.org/goals/goal13.

[61] United Nations. *Sustainable Development Goal 14: Life Below Water*. Accessed: 2025-03-05. 2015. URL: https://sdgs.un.org/goals/goal14.

[62] United Nations. *Sustainable Development Goal 15: Life on Land*. Accessed: 2025-03-05. 2015. URL: https://sdgs.un.org/goals/goal15.

[63] United Nations. *Sustainable Development Goal 2: Zero Hunger*. Accessed: 2025-03-05. 2015. URL: https://sdgs.un.org/goals/goal2.

[64] United Nations. *Sustainable Development Goal 4: Quality Education*. Accessed: 2025-03-05. 2015. URL: https://sdgs.un.org/goals/goal4.

Appendix: Individual Project Reports

Table of Contents

1 EcoSmile at EcoHack-2025: EcoGuard Insights Dashboard for Planetary Preservation

Authors: Basma Jalloul

1.1 Problem Addressed

The planet is facing converging environmental crises—deforestation, biodiversity loss, and greenhouse gas emissions—yet the tools available to decision-makers remain fragmented, complex, and inaccessible. These ecological threats are often studied in isolation, with data scattered across various platforms and published in inconsistent formats. As a result, decision-makers are left without an integrated view of the situation or tools that can translate raw data into actionable insights. Additionally, the lack of interoperability between datasets and analytical models prevents the formulation of holistic, evidence-based environmental strategies. There is a critical need for a system that not only connects the dots across these domains but also simplifies complex environmental patterns for non-technical stakeholders [1, 2].

1.2 Motivation

Over 10 million hectares of forest are lost each year, and this trend shows little sign of abating. These forests are not only biodiversity reservoirs but also pivotal carbon sinks, meaning their loss has a compounding effect on both climate and species survival. Conventional policy tools and forecasting methods often fail to anticipate cascading impacts, such as the way deforestation triggers biodiversity decline and elevates atmospheric carbon levels. With global environmental targets on the line, the challenge lies not just in detecting these patterns but in making them intelligible to those who shape policy [3, 4].This project was motivated by the urgent need to democratize access to environmental forecasting tools and to equip communities and lawmakers with an intuitive, evidence-based interface to support sustainable governance.

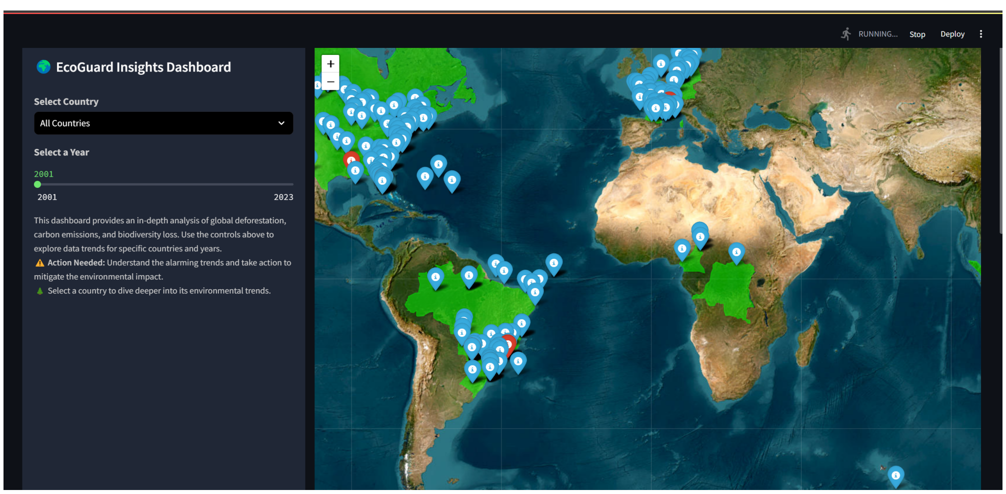

Figure 2: EcoGuard Insights Dashboard: A unified tool for visualizing deforestation, carbon emissions, and biodiversity loss

1.3 Solution

We present **EcoGuard Insights Dashboard**, a unified platform that brings together disparate datasets and transforms them into a coherent, interactive system for ecological risk analysis. The dashboard enables users to visualize country-specific environmental trends, simulate future outcomes under multiple policy scenarios, and identify areas at heightened ecological risk. By synthesizing time-series analysis, geospatial mapping, and predictive modeling, the dashboard empowers users to understand both historical patterns and projected futures. The system is designed for scalability and ease of use, targeting a broad range of stakeholders—from researchers and conservationists to policymakers and regional planners.

1.4 Non-Technical Description

The EcoGuard Insights Dashboard provides a simplified, yet powerful, interface for exploring the planet's ecological trajectory:

- **Interactive World Map:** Track forest loss, emissions, and species decline across countries and timeframes.
- **Forecasting Panel:** Explore how the environment may evolve under different policy scenarios (mitigation vs. business-as-usual).
- **Insight Cards:** Receive synthesized, easy-to-understand findings and recommendations tailored to each region's environmental profile.
- **Country Comparisons:** Benchmark performance across regions and assess progress against sustainability goals.

Designed with accessibility in mind, the platform does not require users to have a background in data science or environmental modeling.

1.5 Technical Description

EcoGuard Insights is developed as a modular, Python-based web application, using `Streamlit` for the front-end interface and a custom-built backend to manage environmental data workflows.

Core Components as show in Figure 7:

- **Data Integration:** The dashboard ingests heterogeneous datasets, including tree cover loss (Global Forest Watch), biodiversity observations (Map of Life), and carbon emissions. These datasets are normalized, cleaned, and temporally aligned to ensure interoperability.
- **Geospatial Mapping:** With `folium` and `GeoJSON`, the application delivers interactive maps highlighting deforestation intensity, biodiversity risk zones, and carbon emission hot spots.
- **Trend Analysis:** Using `pandas`, `NumPy`, and `scikit-learn`, historical data is aggregated and analyzed to identify patterns. A forecasting module extends trends into the future using regression models.
- **Dynamic Visualizations:** Time-series graphs and choropleth maps generated with `matplotlib` and `Plotly` allow users to interact with data filters, select countries or years, and view changes over time.
- **Scenario Simulation:** Scenario simulations apply multipliers based on historical policy effectiveness, a method validated in environmental modeling literature [5, 6].
- **Extensibility:** The backend is structured around modular data processing functions, enabling future extensions (new data sources or predictive models) without architectural overhaul.

1.6 Future Recommendations

1. Incorporate satellite imagery for real-time forest loss detection.
2. Implement advanced species distribution models to estimate extinction risk under future scenarios.
3. Enable multilingual output to increase accessibility for a global audience.
4. Add crowdsourced reporting features to collect local environmental data from communities and NGOs.
5. Develop a responsive mobile version for fieldwork and remote monitoring.

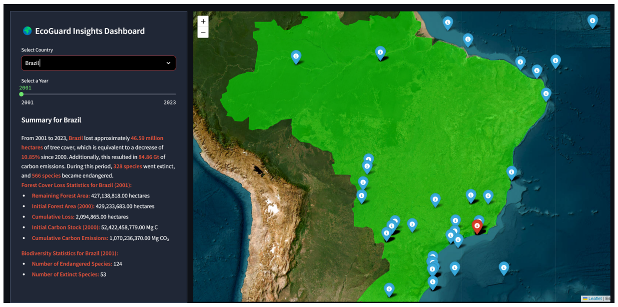

Figure 3: Interactive Map for Deforestation and Species Loss by Filtrable by Year and Country.

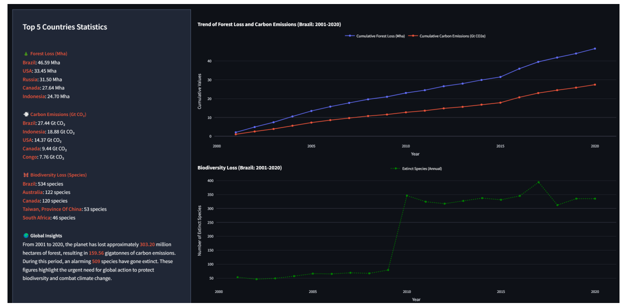

Figure 4: Country-Level Trends for Deforestation, Carbon Emissions, and Species Loss over the Years.

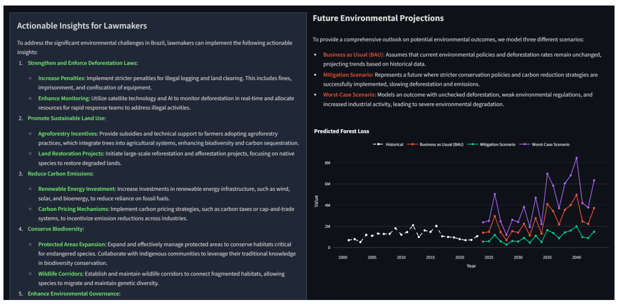

Figure 5: Actionable Insights for Law Makers and Future Predictions based on 3 Scenarios.

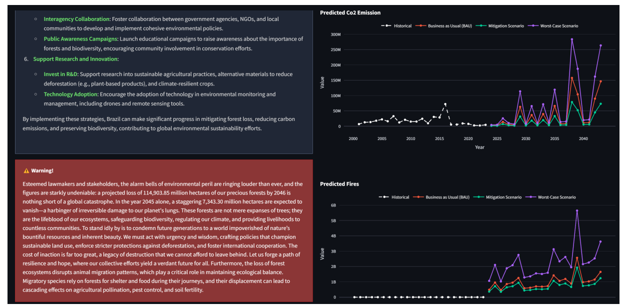

Figure 6: Country Specific Warning Based on the Future Predictions and Incentive for Policy Makers.

Figure 7: Overview of the interactive dashboard visualizing ecological forecasting, including deforestation trends, carbon emissions, and fire predictions under multiple policy scenarios.

References

[1] O. R. Wearn et al. "Reconciling the use of artificial intelligence with environmental sustainability". In: *Nature Sustainability* 5.4 (2022), pp. 301–310.

[2] M. F. Goodchild, W. Li, and N. Patel. "Spatial computing and AI: Geospatial artificial intelligence". In: *Annals of the American Association of Geographers* 110.2 (2020), pp. 285–296.

[3] D. Rolnick et al. "Tackling climate change with machine learning". In: *arXiv preprint arXiv:1906.05433* (2019).

[4] J. Wu et al. "Machine learning in ecological research: Applications and challenges". In: *Ecological Indicators* 126 (2021), p. 107691.

[5] M. Bernardini et al. "A novel missing data imputation approach based on clinical conditional generative adversarial networks applied to EHR datasets". In: *Computers in Biology and Medicine* 163 (2023), p. 107188.

[6] C. Lee et al. "A study on enhancing the visual fidelity of aviation simulators using WGAN-GP for remote sensing image color correction". In: *Applied Sciences* 14.9 (2024), p. 9227.

2 AutoDeck-AI at EcoHack-2025: Eco-Centric Slide Generator

Authors: Hrishikesh Jadhav, Javad Razavian, Moiz Khan Sherwani

2.1 Problem Statement

Creating targeted scientific presentations for ecological research is both inefficient and labor-intensive. Ecologists and related stakeholders must synthesize raw data, abstracts, and supplementary materials from diverse sources, yet current presentation tools do not effectively extract critical figures or adapt content for varied audiences. This inadequacy hampers communication with researchers, practitioners, and funding agencies, thereby impeding timely decision-making and dissemination of research findings [1, 2].

2.2 Motivation

Ecological research is essential to address environmental issues such as climate change, species extinction, and conservation efforts [3]. The ability to present complex data in a clear, audience-tailored manner can foster interdisciplinary collaboration, secure research funding, and translate scientific discoveries into actionable strategies. The motivation for our work stems from the need to simplify the generation of high-quality presentations from scientific manuscripts and supplementary materials, thereby reducing the time researchers spend on formatting and allowing them to focus on core scientific questions [4].

2.3 Solution

AutoDeck-AI integrates advanced artificial intelligence techniques to automatically generate presentations tailored to the needs of ecological research audiences. In our approach, we generate ecological research presentations tailored to specific audiences, including researchers, practitioners, and funding organizations. The solution automatically extracts visual elements such as images, charts, and tables from the uploaded manuscript, generates corresponding captions and assigns them to the appropriate slide sections; a functionality notably absent in existing tools. Moreover, the pipeline is designed to work efficiently even with incomplete drafts or abstracts accompanied by supplementary materials, rendering it uniquely adaptable for preliminary research presentations. Our system leverages a fine-tuned language model (based on GPT-4) to generate appropriate content for the targetted audience. The resulting outputs are merged using prompt engineering techniques to remove redundancies and ensure clarity, and they are subsequently formatted into a professional PowerPoint presentation via python-pptx. This integration of multiple data streams ensures that the final slide deck is both scientifically robust and practically applicable [5].

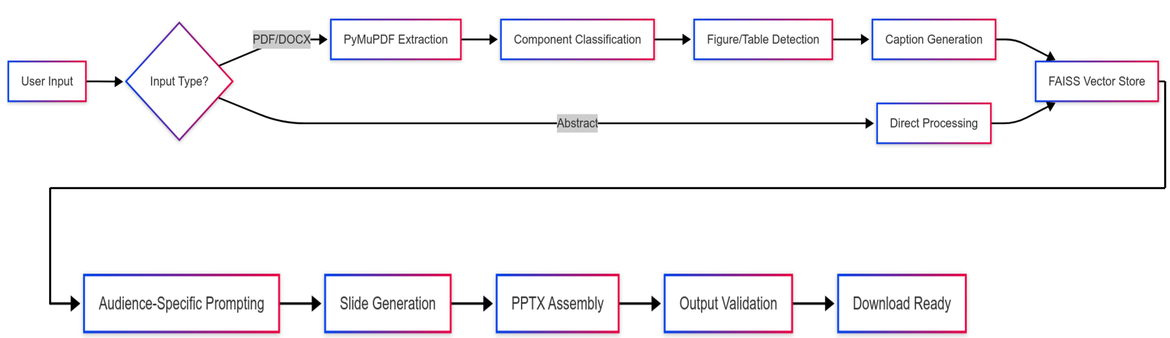

Figure 8: End-to-end workflow of AutoDeck-AI system: From user input (PDF/abstract) through content extraction, audience-specific slide generation, to validated PowerPoint output. Dashed lines indicate optional processing paths.

2.4 Non-technical Description

AutoDeck-AI offers an intuitive, user-friendly interface that enables researchers and practitioners to generate high-quality presentations without extensive manual formatting. Users simply upload their scientific manuscripts or abstracts along with supplementary materials. The system then processes the documents, extracts essential figures and tables, and automatically organizes the content into a cohesive slide deck. This streamlined approach not only saves time but also improves the accessibility of complex scientific information, making it easier for non-experts to understand and apply the findings.

2.5 Technical Description

The tool uses PyPDF2 and PyMuPDF to extract text, figures, and tables from PDF/abstract and supplementary materials, using heuristics to detect visual components and GPT-4o to create synthetic labels, content generation based on prompts and content structuring. Content is designed for academics, practitioners, and funding organizations, and uses customizable themes to increase engagement.

Text is divided into coherent sections and slides are created with integrated images and subtitles. FAISS embeddings facilitate rapid information retrieval, while python-pptx ensures consistent, high-quality presentations. Effective error handling ensures results despite processing complications and facilitates the development of professional, audience-focused slideshows.

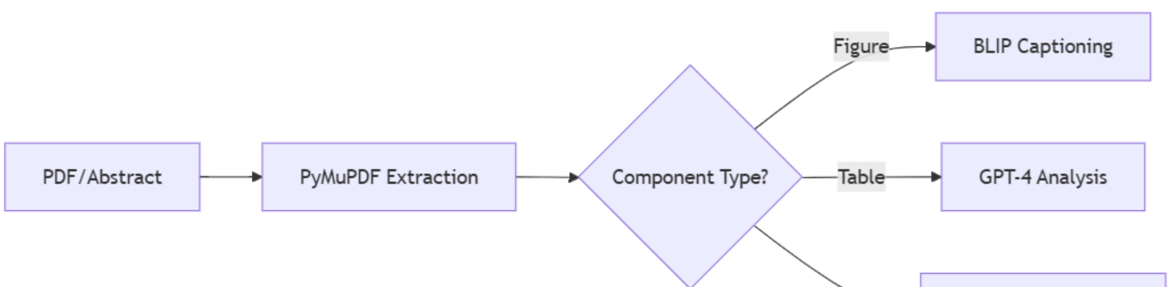

Figure 9: Core technical pipeline of AutoDeck-AI showing PDF/abstract processing through PyMuPDF extraction, component classification with GPT-4, and FAISS vectorization for content structuring.

2.6 Future Recommendations

Future work should focus on further enhancing retrieval precision by incorporating hierarchical vector stores and context-specific filtering techniques. In addition, extending the system to support multilingual queries and integrating voice recognition could make the tool accessible to a broader audience. Real-time data integration from IoT devices and environmental sensors would also allow the system to provide dynamic, up-to-date presentations. Finally, offering customizable slide templates could further empower users to create presentations that best suit their specific needs.

References

[1] Karel Charvát et al. *Big Data in Agriculture – From Foodie Towards Data Bio*. Tech. rep. Plan4all, Oct. 2017.

[2] Sándor Neubauer. "The place of data in precision agricultural data asset management". In: *Institutiones Administrationis* 1 (2021), pp. 52–61. DOI: 10.54201/iajas.v1i2.23.

[3] Mark W. Rosegrant et al. *Climate Change and Agricultural Development*. International Food Policy Research Institute (IFPRI), 2021.

[4] Tom B. Brown et al. “Language Models are Few-Shot Learners”. In: *Advances in Neural Information Processing Systems* 33 (2020), pp. 1877–1901.

[5] Patrick Lewis et al. “Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks”. In: *Advances in Neural Information Processing Systems* 33 (2020), pp. 9459–9474.

3 Agri Chatbot: From Science to Soil

Authors: Vamsi Krishna Kommineni, Anne Peter, Caren Daniel, Alexander Espig

3.1 Problem Addressed

Farmers often encounter an overwhelming amount of information from diverse sources, including scientific literature, research data, and agricultural advice platforms [1, 2]. This information overload complicates the process of extracting actionable, relevant insights that can aid in decision-making [3, 1]. Farmers require a system that simplifies the discovery of critical information, enabling them to apply scientific knowledge effectively to their specific agricultural contexts [4, 5]. This challenge calls for a solution that makes accessing, interpreting, and applying scientific research, structured data, and expert advice easier.

3.2 Motivation

The agricultural sector faces increasing demands to improve efficiency, sustainability, and resilience to climate change [6]. To meet these challenges, farmers need timely and relevant scientific insights. However, navigating the vast landscape of scientific articles, datasets, and advisory resources can be daunting, especially when the information is fragmented or complex [5]. Motivated by the need to bridge this gap, our goal was to build a solution that would empower farmers with quick, science-backed answers to their questions, thus enabling informed decision-making while minimizing the complexity of accessing and interpreting scientific data.

3.3 Solution

This solution provides farmers with a simple user interface (UI) connected to a three-stream pipeline powered by Large Language Models (LLMs) and Retrieval-Augmented Generation (RAG). The pipeline processes agricultural data from three sources: 1) 201 scientific publications and articles, 2) Over 500,000 scientific tabular data entries, including crop information, pesticide information, and some environmental datasets, and 3) Direct Query Responses (LLM answers the query based on its trained data). A Streamlit UI allows farmers to interact with the system, input queries, and refine their prompts for better results, ensuring they receive scientifically-backed and actionable information. We used LLaMA 3.3-70B as the LLM and a FAISS vector store in our proposed workflow.

3.4 Non-technical Description

Farmers are often overwhelmed by the information available, making it difficult to find valuable insights for improving their practices. Our solution simplifies this process by providing an intelligent chatbot combining scientific research publications and tabular data to give farmers clear and actionable answers. Through a user-friendly chatbot interface, farmers can ask specific queries and receive responses grounded in both the latest research and practical data. Moreover, the system offers practical guidance on improving query formulation, enabling farmers to get better results over time.

3.5 Technical Description

From a technical standpoint, the solution implements a three-stream pipeline (Figure 1) that leverages RAG to enhance the accuracy and relevance of responses:

- Vector Database for Scientific Publications: A vector store is created from 201 scientific publications and articles, and an RAG model retrieves the most relevant documents in response to a user query. An LLM then processes the retrieved documents to generate detailed and contextually accurate answers.
- Vector Database for Scientific Tabular Data: Similarly, the system creates a vector store for more than 500,000 scientific tabular data (e.g., agricultural datasets). It uses RAG-based retrieval to fetch relevant data points in response to specific queries. The system processes these data points to provide a comprehensive data-backed answer.

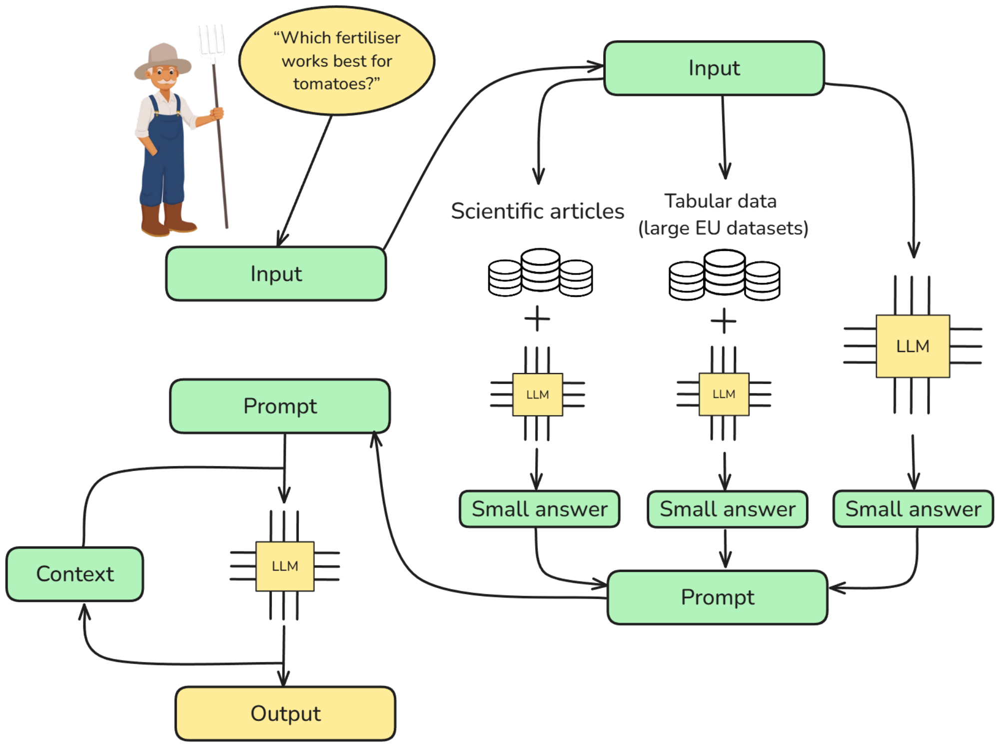

Figure 10: Overview of the approach

- Direct Query Processing with LLM: When a user asks a question, the LLM processes the query and generates a response based on its trained knowledge. This is done through the LLM's built-in capabilities, using its pre-existing knowledge. This part of the pipeline is more helpful when the information about the query is not available in scientific publications and larger datasets.

Fusing answers from all three streams is key to the solution, ensuring the final response is scientifically accurate and practically applicable. The system also uses a prompt to refine and streamline the responses by removing redundancy and providing clarity.

Finally, a Streamlit UI is used to interact with the solution, which not only allows farmers to input questions via a chatbot interface but also includes a side panel designed to guide farmers in writing effective prompts for better results

3.6 Recommendations for Future Extensions

- Vector Store and Retrieval Improvement: Enhance retrieval with techniques like hierarchical vector stores or context-specific filters to improve speed and precision.
- Multilingual Support and Voice Commands: Add support for multiple languages and integrate voice recognition, allowing farmers to interact using natural language in their preferred language.
- Real-time Data Integration: Incorporate real-time data (e.g., weather, IoT devices) for more accurate and personalised responses.

References

[1] Sándor Neubauer. "The place of data in precision agricultural data asset management". In: *Institutiones Administrationis* 1 (2021), pp. 52–61. DOI: 10.54201/iajas.v1i2.23.

[2] Karel Charvát et al. *Big Data in Agriculture – From Foodie Towards Data Bio*. Tech. rep. Plan4all, Oct. 2017.

[3] Mickaël Chizallet, Laurence Prost, and Françoise Barcellini. "Presented at the 2nd International Symposium on Work in Agriculture". In: *2nd International Symposium on Work in Agriculture*. Clermont-Ferrand, France, Mar. 2021.

[4] Laura Bridle et al. *J-PAL (MIT) and CEGA*. Tech. rep. J-PAL (MIT) and CEGA, 2019.

[5] Lee-Ann Sutherland and Pierre Labarthe. "Professional identity and the future of advisory work in agriculture". In: *The Journal of Agricultural Education and Extension* 28.5 (2022), pp. 525–547.

[6] Mark W. Rosegrant et al. "Agricultural Development: New Perspectives in a Changing World". In: *Agricultural Development: New Perspectives in a Changing World*. Vol. 19. Washington, DC: International Food Policy Research Institute, 2021, pp. 629–660.

4 FarmGuide: A bridge between scientists and farmers for natural agriculture practices

Authors: Bartolome Ortiz-Viso, Lorenz Gunreben, Mir Nafis Sharear Shopnil, Nayanika Das

4.1 Problem Addressed

Scientific findings on crop protection, pollination, and pest management are often scattered across numerous research papers, making it challenging for farmers to apply this knowledge in their daily practices effectively. The absence of easily accessible, science-based insights results in suboptimal decision-making, which can negatively impact crop yields, sustainability, and economic outcomes. Ultimately, this situation undermines farmers' ability to make well-informed decisions. Furthermore, the way information is presented on this topic is crucial. Insects, for example, can be seen both as beneficial pollinators and as harmful pests [1, 2]. With the ongoing decline in insect populations, explaining their essential role and the importance of their preservation for agriculture has become a significant challenge that needs to be addressed.

4.2 Motivation

Insect-plant interactions are fundamental to agricultural ecosystems [3, 4], directly influencing crucial processes such as pollination, pest control, and crop health. These interactions are key to improving crop yields and promoting sustainable farming practices [5]. However, farmers face significant barriers when accessing and applying scientific research. The sheer volume of literature, the specificity and context dependability of it [6] combined with the technical language and complexity of ecological studies, makes it difficult for farmers to extract relevant and actionable insights.

As a result, the valuable knowledge from scientific research often remains out of reach for those who need it most. Bridging this gap is necessary and critical to equipping farmers with the tools and knowledge required to make informed, science-backed decisions that promote crop protection, increase productivity, and ensure environmental sustainability [1, 7]. By making complex research more accessible and understandable, we can drive positive change in agricultural practices and improve outcomes for farmers, biodiversity and ecosystems.

Figure 11: Farmguide logo.

4.3 Solution and Non-technical Description

The solution we propose for this problem is called Farmguide (Figure 11). It leverages large language models (LLMs) to extract features that help identify which research articles might be helpful for farmers across various scenarios. Once these articles are identified, we use LLMs to enable farmers to interact with the information in natural language. This allows them to ask questions, better understand the articles, and make informed decisions about whether to implement the recommendations and how to do so effectively. To achieve this, we have developed a graphical interface that provides a service where farmers can input basic information about their crops, ecosystem and cultivation method.

The system then suggests relevant articles, provides tailored advice, and activates a natural language chat feature. Through this chat, users can ask questions and receive answers as if engaging in a conversation (Figure 13).

4.4 Technical Description

We develop a pipeline (See Figure 12 for a complete diagram of the pipeline) designed to process a collection of scientific articles and perform unsupervised label extraction using a Large Language Model (LLM). The system leverages each article's title, abstract, and results sections to identify and extract key attributes such as geographic location, ecosystem type, agricultural dynamics, and crop species. This automated process generates an initial dataset (Table 2) that includes the original articles, a set of extracted labels, and complementary textual responses produced by the LLM.

Paper title	**Country**	**Crop**	**Ecosystem**	**Agriculture**	**Region**	**Additional data**
Functional land cover scale for three insect pests with contrasting dispersal strategies in a fragmented coffee-based landscape in Central Kenya	Kenya	Coffee	Agrosystem	Smallholder agriculture	Eastern Africa highlands	*Abstract, results, questions...*
Making biodiversity work for coffee production. A case study of Gayo Arabica coffee in Indonesia	Indonesia	Coffee	Tropical rain forests and savannah type ecosystems	Not stated	Sumatra (Gayo highlands)	*Abstract, results, questions...*

Table 2: Extract from the data frame generated as the first step of the Farmguide pipeline.

To support user interaction, we developed a Flask[1]-based web service (screenshots in Figure 13) that allows dynamic querying and ranking of articles. The service is modular and can connect either to a locally hosted LLM (e.g., Llama 3.[2] served via Ollama[3]) or a cloud-based model through API keys (e.g., GPT-4[4]). Within our web interface, users configure a profile selecting their thematic or research interests; this profile collect users crop, location, surrounding ecosystem, etc.

Once a profile is selected, the system computes the semantic similarity between the user's profile and all extracted labels using sentence embeddings (all-MiniLM-v6[5]). Each label type (e.g., crop, location) is assigned a configurable weight to reflect its relevance to the user query. These weighted similarities are then aggregated using a weighted average to generate a ranked list of articles most relevant to the selected profile.

This ranked list is used in two ways:

- As contextual grounding for prompt generation sent to the LLM.
- As a visual or summarized output for the user, either as a ranked article list or a concise evidence summary.

Finally, the system integrates both user-defined parameters and the top-ranked contextual information into the LLM prompt, enabling customized, context-aware response generation.

4.5 Recommendations for Future Extensions

Multiple extensions could offer valuable scientific knowledge from this project; we highlight some of them:

[1] https://flask.palletsprojects.com/en/stable/
[2] https://ollama.com/library/llama3.1
[3] https://github.com/ollama/ollama
[4] https://platform.openai.com/docs/models
[5] https://huggingface.co/sentence-transformers/all-MiniLM-L6-v2

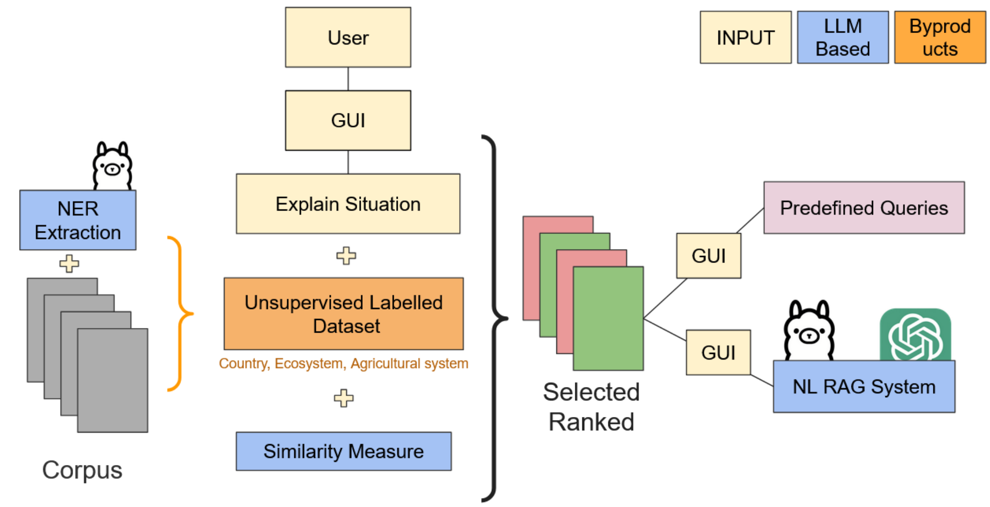

Figure 12: Farmguide schema, describing the system's steps and the system's different inputs and outputs.

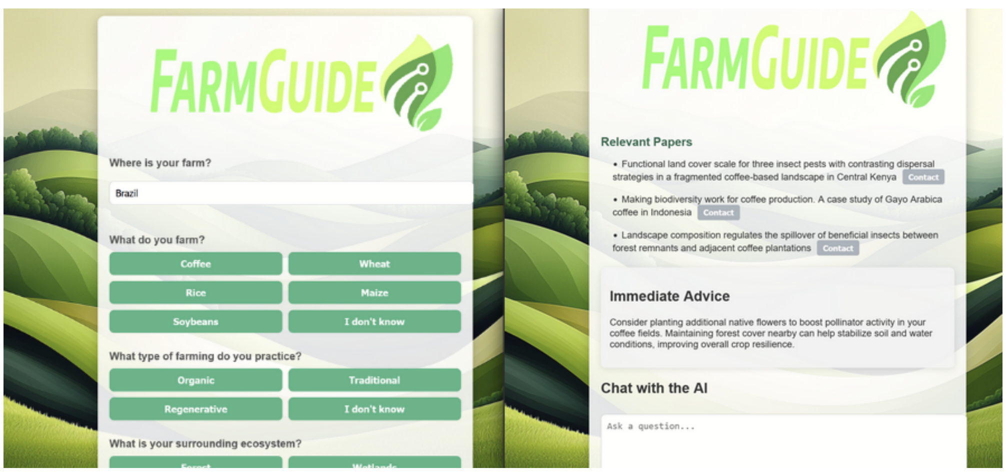

Figure 13: Farmguide screenshots from profile gathering and LLM interaction after it.

- Label extraction in FarmGuide enables experimentation with various approaches. A comparative analysis of these methods would provide valuable insights and offer clear guidelines for selecting optimal strategies for future projects.
- The initial weights for the similarity calculation were set based on an ecologist's recommendation, with higher importance given to crop type and ecosystem over country and type of agriculture. This reflects the overlap of ecosystems within countries and the difficulty of clearly defining farming types. In future versions, these weights could be shaped by input from multiple users and experts, allowing for more diverse and context-specific configurations. Additional factors provided by farmers—such as irrigation methods, pesticide use, or crop rotation—could also be integrated to refine the similarity assessment.
- Another noteworthy feature is the evaluation through similarity metrics. How we handle the different weights and which embeddings and metrics we use could offer valuable insights for future research projects. From a more functional perspective, FarmGuide has the potential to expand across multiple crops and ecosystems by upgrading and enlarging the corpus of articles utilised.
- Additionally, farmers could gain access to information on how to reach scientists while providing feedback on the usefulness and accuracy of the responses and suggestions for improvement. Simultaneously,

scientists could play a crucial role in reviewing the feature extraction process and recommending additional articles for inclusion in the system.

4.6 Supplementary material

- The code, explanations, documentation, visuals, and everything we generated in the project are hosted in this GitHub repository: https://github.com/Gunreben/FarmersGuide.
- Demo video link: we link a demo video of the project prototype. In the video, you can see the capabilities of our prototype, the graphical interface and the additional interactions a user may have with it. https://www.youtube.com/watch?v=g3sFBekraBA

References

[1] M. Busse et al. "How farmers think about insects: Perceptions of biodiversity, biodiversity loss and attitudes towards insect-friendly farming practices". In: *Biodiversity and Conservation* 30.11 (2021), pp. 3045–3066. DOI: 10.1007/s10531-021-02235-2. URL: https://doi.org/10.1007/s10531-021-02235-2.

[2] A. B. Gurung. "Insects – a mistake in God's creation? Tharu farmers' perception and knowledge of insects: A case study of Gobardiha Village Development Committee, Dang-Deukhuri, Nepal". In: *Agriculture and Human Values* 20.4 (2003), pp. 337–370. DOI: 10.1023/B:AHUM.0000005149.30242.7f. URL: https://doi.org/10.1023/B:AHUM.0000005149.30242.7f.

[3] G. Sharma, P. A. Malthankar, and V. Mathur. "Insect–Plant Interactions: A Multilayered Relationship". In: *Annals of the Entomological Society of America* 114.1 (2021), pp. 1–16. DOI: 10.1093/aesa/saaa032. URL: https://doi.org/10.1093/aesa/saaa032.

[4] D. Giron et al. "Promises and challenges in insect–plant interactions". In: *Entomologia Experimentalis et Applicata* 166.5 (2018), pp. 319–343. DOI: 10.1111/eea.12679. URL: https://doi.org/10.1111/eea.12679.

[5] I. Merle, J. Hipólito, and F. Requier. "Towards integrated pest and pollinator management in tropical crops". In: *Current Opinion in Insect Science* 50 (2022), p. 100866. DOI: 10.1016/j.cois.2021.12.006. URL: https://doi.org/10.1016/j.cois.2021.12.006.

[6] G. Tamburini et al. "Pollination contribution to crop yield is often context-dependent: A review of experimental evidence". In: *Agriculture, Ecosystems & Environment* 280 (2019), pp. 16–23. DOI: 10.1016/j.agee.2019.04.022. URL: https://doi.org/10.1016/j.agee.2019.04.022.

[7] L. M. Beltrán-Tolosa et al. "Mestizo Farmers' Knowledge of Entomofauna Is Reflected in Their Management Practices: A Case Study in the Andean-Amazon Foothills of Peru". In: *Frontiers in Sustainable Food Systems* 4 (2020). DOI: 10.3389/fsufs.2020.539611. URL: https://doi.org/10.3389/fsufs.2020.539611.

5 DiversiTeam at EcoHack-2025: EcoSearch

Authors: Sarah T. Bachinger, Daphne Frederike Auer, Edward Gow-Smith

5.1 Motivation and Problem Addressed

When searching for research papers in ecology, the distribution of returned papers with respect to certain diversity metrics may not be particularly heterogenous. For example, research paper search engines (such as Google Scholar) may return papers predominantly from the global north, or from a westernised scientific viewpoint, rather than research representing indigenous or local ecological knowledge. Such biased distributions will reduce the diversity of research easily accessible to ecologists, limiting the well-roundedness of conclusions, and thus potentially hindering the quality of subsequent research. We develop a tool which allows the user to specify the desired geographical distribution of papers, and thus diversify the results.

5.2 Solution

We develop a tool (EcoSearch) which allows users to search for papers in the field of invasion biology based on keywords, and then to diversify the results based on the continent of affiliation of the first author. Our work involves: (1) expert interviews to identify key attributes, focusing on the first author's continent; (2) indexing 37,000 invasion biology papers, partially from [1], extracting affiliation and hypothesis data; (3) implementing search with BM25 and a Bi-Encoder, optimized for CPU use. This is based on previous work from [2]. (4) A Streamlit-based UI with sliders for filtering and result diversification.

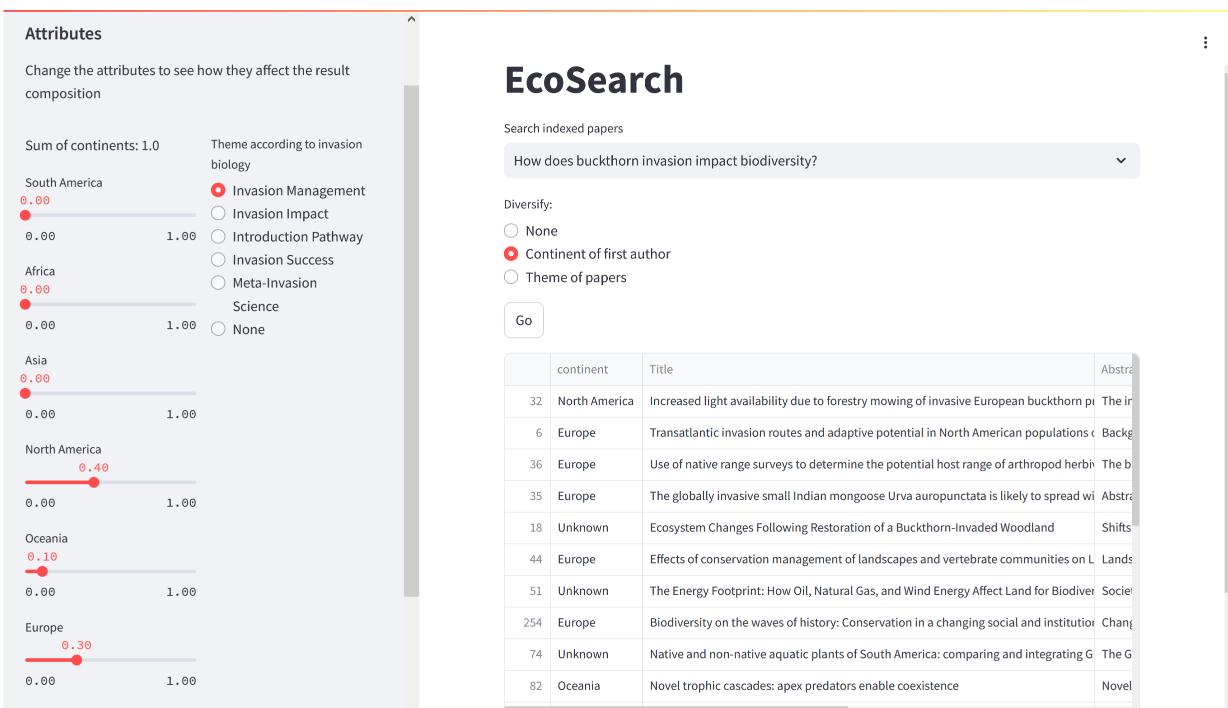

Figure 14: EcoSearch dashboard prototype

5.3 Non-technical Description

First, a sample query can be selected. This would be the search term a user would enter in a running system, i.e. the keywords of interest for finding papers. Then, the search results are shown, as well as the information about some metadata, which is the continent of the first author's affiliation in our example. If desired, the results will be re-ordered to show publications from a new continent distribution, allowing the results to be easily diversified. Therefore, the user changes the distribution for the different continents on the left. We also generate summaries from the first five results of the rankings with an LLM.

5.4 Technical Description

The hackathon project was coded in Python and all dependencies are listed in the file environment.yml, which can be installed with Conda.

DATA The folder data/invasionBio contains all relevant documents. The DOIs of the publications used can be found in corpus_dois.csv. These DOIs were then used to download the title and abstract. Labels for the countries and continent of the first author's affiliation and labels for the hypothesis theme can be added as metadata with the files code/DatasetCreation.ipynb and code/adding_attribute_labels.ipynb.

SEARCH Executing the file retrieval_base.py creates a ranking using BM25 and the Transformer-based Bi-Encoder. The first 500 results can then be reranked based on a new target distribution by running rerank.py. The target distribution can be passed as a terminal argument. To prepare the result list for the user interface, run create_final_data.py. The code is licensed under MIT.

USER INTERFACE The whole graphical user interface is saved in the gui/streamlit.py file. For the demo, we used pre-calculated queries and attribute distributions. The search results are stored in the folder out.

5.5 Recommendations for Future Extensions of Your Solution

The majority of extensions address the integration of further attributes that are important in the ecology domain. As it stands, our tool only diversifies based on the continent (of affiliation) of the first author. But there are many other metrics with which one may want to diversify the search results. Some avenues are:

- To show context-based expertise, for example, one could show the affiliations of the partners. This could be used for clustering institutions that are active in a certain type of research, which could be valuable for both researchers to show areas for collaboration, as well as to write grants.
- An interesting topic is to show search results containing indigenous knowledge. This includes both academic research and grey literature. The former could be labelled as such using keywords to search in the abstract like local, expert knowledge, indigenous, and community-specific keywords like Inuit, Cree, etc. The latter is more difficult to find, but potential sources include practitioner reports from agricultural institutes. As a website, `https://www.ipbes.net/` was mentioned as a further potential source.
- Using the study site for the continent representation in contrast to the author affiliation might highlight underrepresented areas. New attributes could include ecosystems like wetlands, forest for countries or continents.

We experimented with LLM-based summaries of the first five papers, but decided against implementing that in the first prototype. If used, one could have the summaries for both unfiltered and diversified results and compare both, which would be of interest to a researcher looking for two different viewpoints on an issue. The LLM-based summaries would provide, in general, an efficient overview of the returned results.

It would be beneficial to perform some evaluation of our tool. In particular, we could conduct a user study of the GUI and of the reranking in order to assess the benefits, drawbacks, and utility of our tool.

Another idea would be to automatically generate features for the domain of interest. Currently, we have restricted our search results to the field of invasion biology, for which we have a number of defined subfields. However, one could use for example GraphRAG (where a knowledge graph is generated with LLMs: [3]) to extract features from the dataset in an unsupervised way. These features would then be used for diversification, allowing our tool to be used in a domain-agnostic fashion.

References

[1] Daniel Mietchen et al. *Invasion Biology Corpus 2024-07*. Zenodo. June 2024. DOI: 10.5281/zenodo.12518037. URL: https://doi.org/10.5281/zenodo.12518037.

[2] Daphne Frederike Auer, Divyasha Sunil Naik, and Birgitta König-Ries. "Adjusting Fairness and Diversity in Search Results: User-Driven Re-Ranking for Water Research Literature". In: *Datenbanksysteme für Business, Technologie und Web (BTW 2025)*. Research Track, 3–7 March 2025. Bamberg, Germany: Gesellschaft für Informatik, Mar. 2025, pp. 129–152. DOI: 10.18420/BTW2025-06. URL: https://doi.org/10.18420/BTW2025-06.

[3] D. Edge et al. *From Local to Global: A Graph RAG Approach to Query-Focused Summarization*. arXiv preprint arXiv:2404.16130. 2024. arXiv: 2404.16130. URL: https://arxiv.org/abs/2404.16130.

6 Matchbox at EcoHack-2025: Empowering Ecology Research with Efficient Concept Mapping

Authors: Marc Brinner, Nadeen Fathallah, Tarek Al Mustafa

6.1 Problem Addressed

The increasing scale of ecological research presents challenges in systematically identifying and mapping concepts across scientific literature. Current methods for concept extraction and ontology mapping are inefficient, lack contextual disambiguation, and fail to account for semantic variation, resulting in inconsistent or incomplete connections between research papers and domain-specific ontologies.

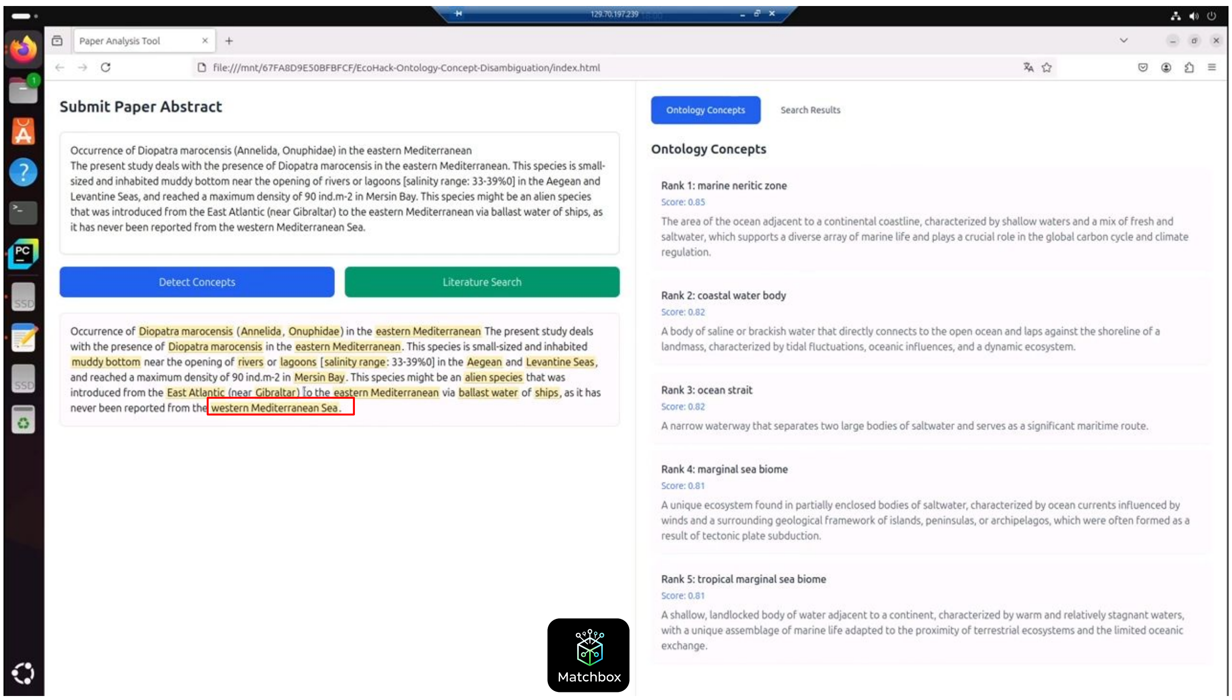

Figure 15: MatchBox user interface showcasing the ontology-concept disambiguation pipeline in action. The tool highlights relevant concepts in a user-submitted scientific abstract using a trained DeBERTa model and semantic embedding techniques. The left panel shows a user-submitted scientific abstract with automatically highlighted candidate concepts extracted using a DeBERTa-based token classification model. The top five ontology concepts retrieved for the selected term are displayed on the right panel with definitions and confidence scores, enabling fine-grained semantic alignment between ecological texts and ENVO/INBIO ontologies. The figure shows the concept disambiguation process for the selected term "western Mediterranean Sea," along with its top-ranked ontology matches, such as marine neritic zone and coastal water body. Top-ranked ontology matches are retrieved using embedding-based matching and LLM-based reranking to determine conceptual alignment. The system supports interactive exploration and real-time reranking through a fine-tuned LLM-based concept verifier.

6.2 Motivation

Efficient concept mapping is critical for advancing ecology research, particularly in areas like invasion biology, where systematic literature reviews play a central role. Existing approaches rely heavily on keyword-based search and manual curation, which are time-consuming and error-prone. By leveraging modern advances in large language models (LLMs) and transformer-based architectures, we aim to create an automated pipeline that streamlines concept extraction, semantic embedding, and ontology alignment. This enables researchers to locate, interpret, and integrate scientific knowledge more effectively.

6.3 Solution

We developed an end-to-end pipeline that combines advanced LLMs and transformer-based models to perform accurate and efficient concept extraction and ontology mapping. Our approach involves:

1. Using an LLM (Llama-3-8B-Instruct) to extract and define relevant concepts from ecological abstracts.
2. Training a series of DeBERTa-based models for token-level classification, definition embedding, and semantic similarity scoring.
3. Implementing a reranker model to improve precision by verifying semantic alignment between text concepts and ontology definitions.
4. Creating an interactive matching tool and a concept-based search engine that enables exploration of semantically related literature.

6.4 Non-technical Description

Our solution simplifies how ecologists discover and connect concepts in scientific texts. Imagine reading a research paper and wondering how its ideas align with existing scientific frameworks like the ENVO or INBIO ontologies. Our tool highlights key terms, identifies their meanings, and connects them to related concepts in ecological databases. For instance, if a paper mentions “invasive species,” the tool identifies this term and links it to similar concepts and definitions across various scientific texts. It also lets users search for papers that share conceptual overlaps, even when the language differs. This system reduces the manual effort required to find and interpret related research, empowering ecologists to focus on advancing their field.

6.5 Technical Description

As the basis for our experiments, we use a subset of 6000 paper titles and abstracts that address invasion biology from a dataset collected using a Wikidata query [1]. We then proceeded to extract concepts from the abstracts by prompting an LLM (Llama-3-8B-Instruct) to identify single-word or multi-word concepts that potentially match concepts contained in ontologies like the ENVO. Then, we again used the same model to generate five definitions for each extracted term, with the corresponding scientific abstract being provided as additional context to ensure that the model generates a definition that explains the actual meaning of the concept as used in the abstract. As a second part of our dataset, we used the INBIO ontology [2], and the ENVO ontology [3], which contain the concepts we want to link to the concepts that occur in the scientific texts. For these concepts from the ontologies, we created five additional definitions using the same LLM. Again, we ensured that the definitions match the exact meaning of the concept from the ontology by defining the term that is provided by the ontology as context for the LLM. The resulting dataset contains many concepts from the ontology or concepts extracted from abstracts, both in combination with corresponding definitions. We used this dataset to create a pipeline leveraging four different models to perform automatic term matching between ontology and abstracts:

We trained a DeBERTa-base model to recognize relevant concepts (as identified by the LLM) in a scientific abstract. We treated this as a token-level classification problem so the model could predict scores for each token in the abstract, indicating whether it was part of a relevant concept. We trained the model to predict label 1 for the first token of every word that the LLM extracted and to predict label 2 for each subsequent token of that concept to determine exact concept boundaries even if two concepts are next to each other. The loss function we used is the categorical cross entropy for the two “positive” labels (1 and 2). For label 0, we decided not to use the categorical cross-entropy since it would strongly punish if additional tokens were selected as relevant if the LLM did not. This would be undesirable since we assume that the LLM might not have extracted all relevant concepts. Thus, we instead included an $L2$ penalty that drives the average probability of labels 1 and 2 to zero for all tokens not marked as relevant by the LLM. This “softer” method of driving relevance scores of non-relevant tokens to 0 ensures that the model is not punished much if it predicts additional concepts as relevant, as long as this is rarely done.

The second model we created is a definition embedding model that transforms definitions of concepts into dense vector representations. We again used the DeBERTa base and trained it by predicting embeddings for two definitions of the same concept (with both originating from the same abstract or the same concept in the ontology to avoid potentially mixing definitions of different concepts that have the same name (e.g., invasion can be used in the medical context, which differs from invasion in ecology)) and training the model to predict similar embeddings for those two definitions, while predicting different embeddings for unrelated concepts. For that, we used a margin-based triplet loss: Given an anchor definition A, a positive sample B (i.e., a definition of the same concept as A,) and a negative definition C (defining a different concept), we defined two distances $d_1 = ||A - B||$ and $d_2 = ||A - C||$ and defined the loss $L = relu(d_1 - d_2 + 1)$ We used in-batch negatives and computed this loss for all possible pairs of positive and negative samples in the batch for a single gradient update. The resulting model embeds definitions into an embedding space that places definitions of similar concepts close to each other.

The definition embedding model allows for matching concepts from texts to concepts from the ontology by first generating a definition for the concept from the text, embedding it, and comparing that embedding to the embeddings for the definitions from the concepts in the ontology. However, this process is inefficient since it requires generating a definition for each concept of interest in a given text. We, therefore, trained an additional embedding predictor that takes a scientific abstract and directly predicts the embedding of each token, which can then be quickly matched with the ontology embeddings. To do that, we again leveraged our existing dataset: For a given abstract and relevant terms extracted by the LLM, we first embedded the definitions of those concepts using the model trained in the previous section and then used those embeddings as ground truth for training the DeBERTa model (using an $L2$ loss) to directly predict that embedding for the corresponding token in the abstract without ever seeing the definition. In this way, we created a model that directly predicts semantic embeddings for every relevant concept in the text in a single forward pass.

The resulting models are able to quickly find relevant concepts from the ontology for a given concept in an abstract. A potential problem with this method is that similar concepts will have similar embeddings, so multiple similar concepts are retrieved without a precise relevance ranking. The model never saw the textual and ontology concepts simultaneously (because their embeddings are predicted in separate models). Therefore, we trained a reranker model, a fine-tuned version of Llama-3.2-1B-Instruct. This model receives three pieces of information as input:

- A sentence from the abstract
- The concept that we are interested in (i.e., a word from the sentence)
- A definition.

The model is then fine-tuned to return "yes" if the given definition matches the concept from the text and "no" otherwise. The training was done using the definitions generated for the abstract terms as positive samples and random definitions generated for other abstract concepts as negative samples. The resulting model can thus identify if a given concept from the ontology (with corresponding definition) matches a concept from the text.

We created a matching tool that uses all models to perform real-time term matching between user-provided input texts and the INBIO and ENVO ontologies. If the user provides a text, the tool first uses the concept extraction model to extract relevant terms. Then, it uses the abstract token embedding model to generate semantic embeddings for each concept in the abstract. These are then compared to the embeddings of all ontology concepts (which were precomputed using the definition embedding model). Finally, the top 5 candidates that were retrieved in this way for each concept are plugged into the reranker model (together with the sentence in which the concept appears in the text) to create a final ranking of the semantic relatedness of the concepts. This ranking is presented to the user if they click on one of the detected concepts, which we highlight in the text.

Finally, we used our models to create a concept-based scientific search engine. If the user provides an abstract, we again use the abstract token embedding model to generate semantic embeddings for each concept from the abstract so that we essentially have n embeddings of relevant concepts representing a given abstract. We did the same for each abstract in our dataset. To check the relatedness of the two abstracts, we compute the pairwise distances between all concept embeddings that represent the two abstracts. To decide if a candidate paper matches our query paper, we take the minimum distance from each query-paper-embedding to any other embedding from the candidate to get a score for each concept from our query that represents if that concept is also addressed in the other abstract. We then average all concepts in the query to get a single score that measures the overall concept overlap between the two papers. We found this method highly successful at matching related concepts, even if the same words in the abstracts do not represent them.

6.6 Recommendations for Future Extensions

1. Since this project was coded in the context of a hackathon, the time for model training was sparse. Therefore, it should be possible to enhance the system further by enhancing the models used and how intricately they were trained.

2. Expand the system to process paper abstracts and full-text papers.

3. When talking to domain experts, it became clear that literature search is a huge practical issue in the domain. The manual effort needed for systematic literature reviews could be reduced drastically using similarity embeddings instead of keyword search approaches. We imagine a future system to look like this:

   - Input: 'I'm looking for papers that contain the concept [invasive species]. Return papers that use the concept like it is used in this example [pasted paper abstract by user].'
   - Output: Papers containing a concept with an embedding similar to the embedding for [invasive species] produced from the user query.

6.7 Supplementary material

- All code, documentation, and project materials are available in our GitHub repository: `https://github.com/EcoWeaver/EcoHack-Ontology-Concept-Disambiguation`

- Demo video- A walkthrough of the prototype, showcasing its features, user interface, and interactive capabilities: `https://www.youtube.com/watch?v=ffp0jGeaZlI`

References

[1] F. Brunner et al. *Invasion Biology Literature Dataset*. Zenodo. 2024. DOI: 10.5281/zenodo.12518037. URL: https://doi.org/10.5281/zenodo.12518037.

[2] INBIO Ontology. *BioPortal*. https://bioportal.bioontology.org/ontologies/INBIO. Accessed: March 26, 2025. 2025.

[3] Environmental Ontology (ENVO). *BioPortal*. https://bioportal.bioontology.org/ontologies/ENVO. Accessed: March 26, 2025. 2025.

7 EcoSci Recommender

Author: Samira Korani

7.1 Problem Addressed

The primary aim is to address the challenge of organizing, analyzing, and understanding research contributions, methodologies, and expertise across disciplines to facilitate innovation, detect research gaps, and foster collaboration. The problem can be broken down into the following aspects:

- Mapping Research Contributions
- Understanding Methodologies
- Validating Research Outcomes
- Interrelating Research Topics
- Institutional Research Focus
- Domain and Skill Mapping

By addressing these problems, this framework aims to build a robust, structured, and actionable knowledge base that not only enhances understanding of the research landscape but also facilitates targeted discovery and collaboration.

7.2 Motivation

The motivation for developing a framework to map research contributions, methodologies, expertise, and institutional focus stems from the increasing complexity and volume of academic research. As the global research ecosystem grows, significant challenges emerge in identifying innovation opportunities, detecting research gaps, and fostering interdisciplinary collaboration.

7.3 Core Components of the Solution

1. Entity and Relationship Extraction
2. Knowledge Graph Construction
3. Data Sources and Integration, APIs and datasets from scholarly repositories (e.g., Semantic Scholar, PubMed, CrossRef, OpenAlex).
4. LLM-Powered Agent
5. Web Service applications

7.4 Non-technical Description

Imagine a tool that works like a highly intelligent research assistant, designed to help scientists, researchers, and institutions navigate the overwhelming amount of research being published every day. This tool is powered by advanced artificial intelligence (AI) and focuses on simplifying the process of discovering new ideas, understanding research trends, and finding gaps where important questions are yet to be explored. Here's how it works:

- Finding Key Contributions
- Understanding Methods and Approaches
- Spotting Research Gaps
- Connecting the Dots
- Highlighting Expertise

7.5 Technical Description

- Data Collection Layer
- NLP Layer - Relation and Entity Extraction
- Knowledge Graph Construction
- User subgraph construction
- API integration
- LLM-Powered Agent (Task-Specific Tuning)
- Update the Graph
- Web app service

7.6 Future Extensions

Ecology Semantic Search Engine

8 BioSim: A Multi-Agent Framework for Biological Invasion Simulation

Authors: Zijian Ling, Shuhan Miao

8.1 Problem Addressed

Biological invasions pose severe ecological and economic challenges worldwide. Invasive species can outcompete native species, disrupt ecosystems, and cause significant financial damage to agriculture, fisheries, and water systems [1]. Traditional simulation models for biological invasions often lack adaptability, scalability, and generalization, making it difficult to predict species interactions in diverse ecological contexts [2]. To address these limitations, we propose an LLM-based multi-agent framework that enhances simulation scalability and generalizability.

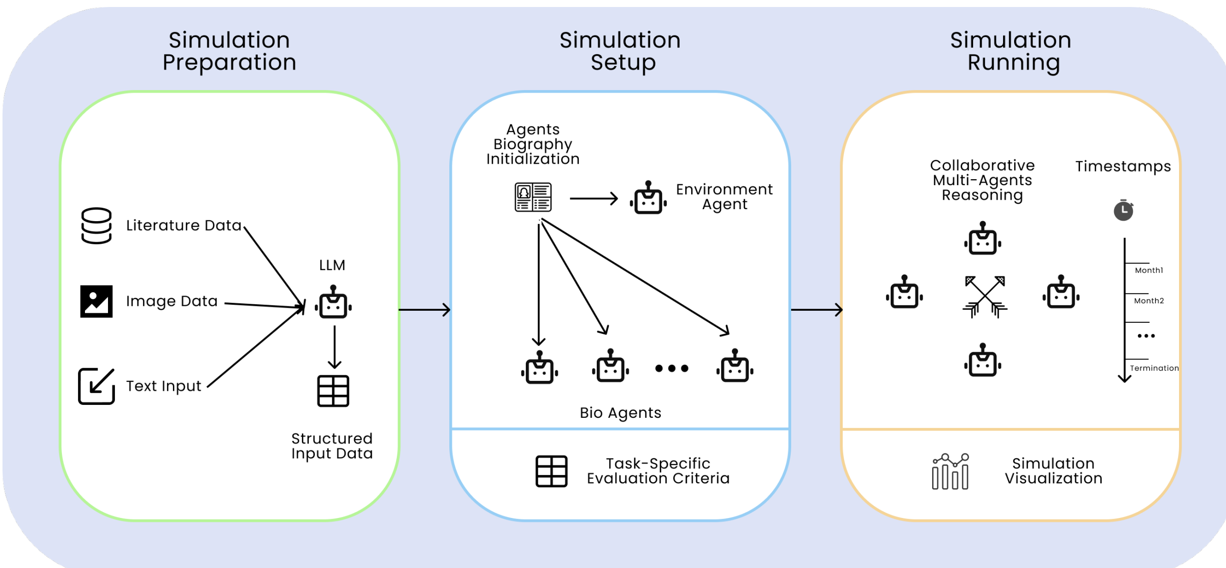

Figure 16

8.2 Motivation

The increasing prevalence of invasive species has made it essential to develop more robust and dynamic simulation models. Traditional models are often rigid, requiring extensive parameter tuning and predefined rules, limiting their ability to adapt to new ecological scenarios. Recent advancements in Large Language Models (LLMs) and multi-agent systems offer a promising alternative, allowing simulations to integrate real-time data, reason collaboratively, and dynamically adjust based on environmental conditions [3]. Our motivation stems from the need for a scalable, generalizable, and data-driven solution that supports conservation efforts and invasive species management.

8.3 Solution

Our proposed framework, BioSim, leverages LLMs and a multi-agent system to simulate biological invasions. BioSim consists of interconnected components, including structured input data, literature data integration, agent initialization, simulation execution, and evaluation. The model enables agents (representing invasive and native species) to interact, adapt, and reason collaboratively based on ecological rules and real-time data. The system also provides simulation visualization and analysis tools, facilitating decision-making in conservation and ecological management.

8.4 Non-technical Description

The BioSim framework models the interactions between invasive and native species in a simulated environment. Consider the case of zebra mussels vs. native freshwater mussels in the Great Lakes. Zebra mussels, being invasive, reproduce rapidly and outcompete native mussels for food and habitat. Our framework simulates these interactions by allowing agents (representing the species) to behave based on predefined ecological principles and real-time learning. Over time, the simulation demonstrates how native species decline, how invasive species adapt, and how environmental changes influence population dynamics. The results help ecologists and policymakers develop more effective mitigation strategies.

8.5 Technical Description

BioSim employs a multi-agent architecture where each agent represents a species with specific behavioral rules. The system integrates:

- LLM-driven reasoning: Agents utilize an LLM for decision-making, adjusting behaviors dynamically based on environmental inputs.
- Collaborative multi-agent interactions: Agents simulate group dynamics, resource competition, and ecological shifts.
- Data-driven modeling: The framework incorporates literature-based ecological models and real-world data.
- Simulation workflow: The process involves setup, initialization, running the simulation over multiple timestamps, and evaluation.
- Visualization tools: Results are presented through charts and models to aid analysis.

8.6 Recommendations for Future Extensions

While BioSim provides a strong foundation for biological invasion simulation, future enhancements could include:

- Advanced framework: Optimize multi-agent framework to support more realistic, comprehensive, and large-scale simulations.
- Expanded species interactions: Incorporate more species to model complex ecological webs including one-to-one invasion and many-to-many scenarios.
- Geospatial modeling: Integrate GIS-based environmental data for region-specific predictions.
- Climate impact integration: Simulate how climate change alters invasion dynamics.
- Policy testing and intervention strategies: Develop scenarios where different control measures (e.g., eradication efforts, habitat restoration) are tested in the simulation.

References

[1] Daniel Simberloff et al. "Impacts of biological invasions: what's what and the way forward". In: *Trends in Ecology & Evolution* (2013).

[2] Ana Buchadas et al. "Identifying priority areas for research and management of invasive species in a river basin by combining species distribution models and multi-criteria decision analysis". In: *Journal of Environmental Management* (2017).

[3] Ziyi Yang et al. *Query2Summary: Teaching LLMs to Generate Query-Focused Scientific Summaries via Graph-RAG*. arXiv preprint arXiv:2411.11581. 2024. arXiv: 2411.11581. URL: https://arxiv.org/abs/2411.11581.

9 Healing Factor at EcoHack-2025: Forecasting Ecosystem Recovery Efforts in Ukraine

Author: Andrii Krutsylo

9.1 Problem Addressed

The war in Ukraine has caused severe damage to local ecosystems, including burned forests, polluted waters, and disrupted wildlife habitats [1]. Traditional restoration efforts are labor-intensive and rely on low-tech methods, making it difficult to efficiently assess and manage restoration progress across vast affected areas. In addition, strict government protocols limit the integration of innovative machine learning solutions into disaster recovery operational planning and resource management.

9.2 Motivation

Effective ecological restoration in conflict-affected regions such as Ukraine requires strategic allocation of limited resources to maximize restoration outcomes. Given the scale of environmental degradation and regulatory constraints, there is a critical need for data-driven insights to guide where additional interventions will have the greatest impact on ecosystem recovery. Optimizing the balance between intensity of intervention and breadth of coverage can improve the efficiency and effectiveness of ecosystem restoration.

9.3 Solution

Healing Factor addresses this need by using Microsoft Planetary Computer's Sentinel-2 satellite imagery to predict the Normalized Difference Vegetation Index (NDVI), a key indicator of vegetation health [2]. The system analyzes NDVI trends to identify areas in Ukraine that are on track for ecological recovery and highlights regions that require further intervention. This approach provides strategic insight to optimize resource allocation and focus restoration efforts where they are most needed and effective.

9.4 Non-technical Description

Healing Factor is a tool designed to help restore Ukraine's damaged ecosystems by using satellite imagery to monitor vegetation health. It processes these images to create maps that show which areas are recovering effectively and which need additional support. These maps allow organizations and decision-makers to allocate their efforts and resources more efficiently, ensuring that restoration initiatives are both effective and timely.

9.5 Technical Description

9.5.1 Data Collection

Healing Factor uses Sentinel-2 satellite data accessed through the Microsoft Planetary Computer [3, 4]. The data retrieval focuses on four spectral bands: B02 (blue), B03 (green), B04 (red), and B08 (near infrared). Monthly satellite images within specified bounding boxes and date ranges are collected to monitor targeted regions over time.

9.5.2 Preprocessing

- **Monthly data recovery:** Selects images with minimal cloud cover using Planetary Computer's STAC API, limiting to a maximum number per month to ensure data quality.
- **Mosaicking:** Combines multiple images into a single GeoTIFF per month by applying a pixel-wise median function for each spectral band, resulting in a coherent monthly mosaic.
- **Time-series windowing:** Creates sequences of three-month windows from the monthly mosaics to predict the NDVI for the subsequent month.

- **Patch Extraction:** Divides each spatiotemporal sample into manageable patches of size 128×128.

9.5.3 Model Architecture

Healing Factor uses a convolutional long-term memory neural network that effectively handles both spatial and temporal dimensions of the data [5]. Spatial dimensions are captured by convolutional layers that process the height and width of the image patches. Temporal dimensions are handled by LSTM units that process the sequence of monthly data. After processing all time steps, the model outputs a single channel representing the predicted NDVI for the following month.

9.5.4 Training

- **Loss Function:** Mean Squared Error (MSE) is used to quantify the difference between predicted NDVI values and ground truth NDVI.
- **Performance Metrics:** The ConvLSTM model achieves an MSE of less than 0.02 after 50 epochs.
- **Training Duration:** Training takes approximately 2 hours on an NVIDIA GeForce GTX 1050Ti.
- **Validation Strategy:** Utilizes 20% of randomly selected patches from the targeted region for validation to ensure model generalization and prevent overfitting.

9.5.5 Output

The trained model generates predictive NDVI maps that indicate the health and recovery trajectory of vegetation in targeted areas. These maps serve as strategic tools for identifying regions that are recovering well and those that require additional intervention.

9.6 Future Recommendations

1. **Expand ecological metrics:** Include additional indicators such as water quality, soil health, and biodiversity indices to provide a more comprehensive assessment of ecosystem restoration.
2. **Enhance accessibility:** Create a user-friendly web platform to regularly update and distribute restoration maps, making the data easily accessible to stakeholders and the public.
3. **Integrate with field data:** Combine satellite predictions with on-the-ground observations to validate and refine model accuracy [6].

References

[1] D. Basiuk et al. “Environmental impacts of the Russia-Ukraine war: A preliminary assessment”. In: *Environmental Policy and Law* 52.3 (2022), pp. 151–157. DOI: 10.3233/EPL-220048. URL: https://doi.org/10.3233/EPL-220048.

[2] C. J. Tucker. “Red and photographic infrared linear combinations for monitoring vegetation”. In: *Remote Sensing of Environment* 8.2 (1979), pp. 127–150. DOI: 10.1016/0034-4257(79)90013-0. URL: https://doi.org/10.1016/0034-4257(79)90013-0.

[3] M. Drusch et al. “Sentinel-2: ESA’s Optical High-Resolution Mission for GMES Operational Services”. In: *Remote Sensing of Environment* 120 (2012), pp. 25–36. DOI: 10.1016/j.rse.2011.11.026. URL: https://doi.org/10.1016/j.rse.2011.11.026.

[4] Microsoft Planetary Computer. *Planetary Computer Documentation*. https://planetarycomputer.microsoft.com. 2021. URL: https://planetarycomputer.microsoft.com.

[5] X. Shi et al. “Convolutional LSTM Network: A Machine Learning Approach for Precipitation Nowcasting”. In: *Advances in Neural Information Processing Systems (NeurIPS)*. Vol. 28. 2015. URL: https://papers.nips.cc/paper_files/paper/2015/hash/07563a3fe3bbe7e3ba84431ad9d055af-Abstract.html.

[6] N. Pettorelli et al. “Satellite remote sensing for applied ecologists: Opportunities and challenges”. In: *Journal of Applied Ecology* 51.4 (2014), pp. 839–848. DOI: 10.1111/1365-2664.12261. URL: https://doi.org/10.1111/1365-2664.12261.

10 EcoLogic: A Benchmark for Causal and Correlational Reasoning of LLMs Based on Ecological Interactions

Author: Nico Heider

10.1 Introduction

Large Language Models (LLMs) have emerged as powerful tools for reasoning across diverse ecological tasks, including recommendation systems, advanced text analysis, and computational simulations. While adept at capturing complex statistical structures from language, their ability to perform causal and correlational reasoning in domains with intricate interdependencies—such as ecology—remains underexplored. We address this gap by introducing EcoLogic, a benchmark that evaluates LLMs' capacity to reason based on causation and correlation. Unlike existing reasoning benchmarks [1, 2, 3], EcoLogic provides examples grounded in real-world ecological interactions and is explicitly designed to help Ecologists and Computer Scientists select the most appropriate model for their specific tasks. The code and dataset is accessible at `https://github.com/nheider/ecologic`.

10.2 Causal and Correlational Reasoning in Ecology

In the context of ecology, correlational reasoning allows LLMs to infer patterns based on statistical associations. For example, the model might recognize that the population of a predator (e.g., lions) is correlated with the population of its prey (e.g., zebras), based on prior knowledge of their interactions. However, this reasoning does not imply causality—it merely highlights an observed relationship. This type of reasoning is useful in situations where direct causal relationships may not be immediately clear, but there is enough data to identify trends or patterns. Causal reasoning, on the other hand, is more complex and involves understanding the mechanisms through which one event (e.g., the introduction of a new predator) can directly influence another (e.g., a decline in a prey population). It requires understanding the underlying dynamics of the system, such as feedback loops and time-dependent effects. For example, if a keystone species is removed from an ecosystem, causal reasoning would help predict how this would affect the population dynamics of other species in the food web, even if those interactions are not explicitly observed in the training data. Since ecological systems are inherently complex, characterized by multifaceted interactions among species, both causal and correlational reasoning are essential for understanding and predicting outcomes. For instance, understanding a trophic cascade—where changes in one species affect others across multiple levels of the food web—given a list of animals in a habitat, requires correlational reasoning to construct a graph of interactions, based on prior knowledge on the given species, and causal reasoning to model how these changes cascade through the food-web at different time steps. Similarly, in LLM tasks, extracting a novel hypotheses from a paper, and placing it in a causal graph needs the model to both understand how the hypothesis correlates to known knowledge and reason causally how this new fact would interact with other entities. Figure 13 shows a subset of a food-web and shows the intricate interdependencies that form a predator-prey relationship in an ecosystem.

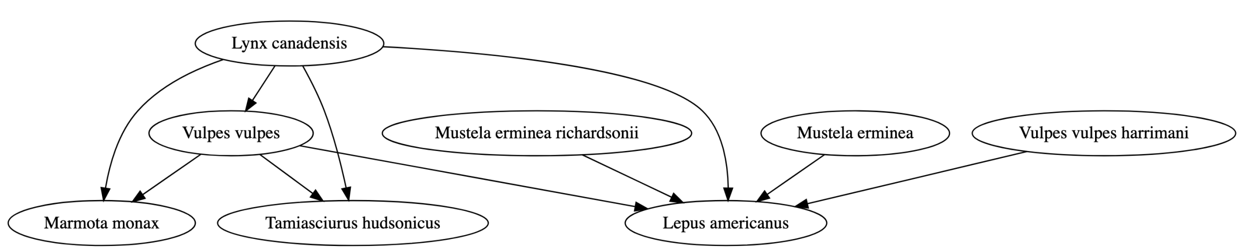

Figure 17: Example of a food-web sub graph generated from the GloBI Dataset [4].

10.3 Benchmark Task

The EcoLogic benchmark introduces three distinct reasoning tasks, each testing a different aspect of reasoning. The causal Graph Reasoning task requires the model to predict the effect of a change in one species on another species, based solely on a given causal structure. For instance, if the predator population increases, how does that affect the prey population? This tests the model's ability to reason through direct causal relationships. In the correlational reasoning task, the model predicts the outcome based on observed correlations, rather than causal structures. It might rely on prior knowledge that, for example, higher predator numbers correlate with lower prey numbers, but without explicitly understanding the causal mechanisms behind it. The mixed reasoning task combines both causal and correlational reasoning, giving the model enough information to use both types of reasoning to make predictions. If LLMs are more than correlational 'stochastic parrots' and also reason causally, we would expect better performance in the mixed task, than in the pure correlational reasoning task. The main contribution of this project is a benchmark dataset consisting of prompts describing food-webs and asking about the impact of species population changes, accompanied by ground truth data.

Graph reasoning task. This task requires the model to predict the effect of a change in one species on another species, based solely on a known causal structure. For instance, if the predator population increases, how does that affect the prey population? We obscure the species and give the model no knowledge of the origin of the task. This tests the model's ability to reason through given causal relationships.

> B lowers C, G, F and A. G lowers C, F and A. H lowers A. E lowers A. D lowers A.
>
> If B declines, what is the immediate effect on the occurrence of C?
>
> Does it a) increase b) decrease, or is there c) no change?

Ecological reasoning task. In this task, the model must predict the outcomes based only on observed correlations from the training data, rather than a given causal structure.

> Lynx canadensis, Marmota monax, Vulpes vulpes, Tamiasciurus hudsonicus, Lepus americanus, Vulpes vulpes harrimani, Mustela erminea richardsonii, Mustela erminea form a food-web. Assume there are no other species present.
>
> If the population of Lynx canadensis declines, what is the immediate effect on the population of Vulpes vulpes?
>
> Does it a) increase b) decrease, or is there c) no change?

Mixture task. The mixed reasoning task combines both causal and correlational reasoning, giving the model enough information to use both types of reasoning to make predictions.

Lynx canadensis preys on Marmota monax, Vulpes vulpes, Tamiasciurus hudsonicus and Lepus americanus. Vulpes vulpes preys on Marmota monax, Tamiasciurus hudsonicus and Lepus americanus. Vulpes vulpes harrimani preys on Lepus americanus. Mustela erminea richardsonii preys on Lepus americanus. Mustela erminea preys on Lepus americanus.

If the population of Lynx canadensis declines, what is the immediate effect on the population of Vulpes vulpes?

Does it a) increase b) decrease, or is there c) no change?

10.4 Technical Implementation

We use a subset of the Global Biotic Interactions (GloBi) [4] dataset to generate smaller sub graphs (Mammals with the interaction type 'preys on') which we then use as an equilibrium state of a discrete simulation with one time step. We simplify the complex real world interactions between species to a basic set of rules: if a species has more predators than pray it decreases, if there is more prey than predators it increases, else there is no change. This is done to create a task we can automatically generate ground truth data for by using a solver. More complex interactions, such as an agent-based simulation, could better reflect ecological reality but would be computationally inefficient and challenging to evaluate LLM performance on, given the stochastic nature of these systems. We find that LLMs generally understand the simplified task correctly. We randomly change the population of one node and use a solver to find the change in a target node to generate ground truth data. The presented approach makes it possible to create a large amount of dataset entries with ground truth, without any labelling by hand. We generate basic prompts, so further refinement can be done via prompt engineering to adapt the benchmark to the idiosyncrasies of the different models.

10.5 Preliminary Results and Future Research

Our preliminary results indicate that the current version of the EcoLogic benchmark is often too easy for state-of-the-art frontier models, such as Claude 3.7 and ChatGPT 4.0. In contrast, smaller models, such as Mistral 7B, frequently struggle with simple tasks like interpreting Latin animal names. We hypothesize that this discrepancy stems from a lack of ecological training data in smaller models, though further investigation is required to confirm this hypothesis. These findings underscore the importance of selecting models that align with the specific requirements of the ecological tasks they are used for. The EcoLogic benchmark is designed to assist ecology practitioners in identifying models that are both energy- and compute-efficient while retaining sufficient capability for their applications. By leveraging large language models responsibly, researchers can maximize their utility while minimizing climate impact. To enhance the benchmark's utility, we plan to increase its difficulty by incorporating a broader range of interaction types and animal classes. Future work could focus on developing methodologies to better disentangle reasoning capabilities from knowledge retrieval, enabling a more precise evaluation of model strengths and limitations.

References

[1] Zhengxuan Jin et al. *Can Large Language Models Infer Causation from Correlation?* arXiv preprint arXiv:2306.05836. 2024. arXiv: 2306.05836. URL: http://arxiv.org/abs/2306.05836.

[2] Zhengxuan Jin et al. *CLadder: Assessing Causal Reasoning in Language Models*. arXiv preprint arXiv:2312.04350. 2024. arXiv: 2312.04350. URL: http://arxiv.org/abs/2312.04350.

[3] Thomas Geffner et al. *Deep End-to-End Causal Inference*. arXiv preprint arXiv:2202.02195. 2022. arXiv: 2202.02195. URL: http://arxiv.org/abs/2202.02195.

[4] GloBI Community. *Global Biotic Interactions: Interpreted Data Products*. Zenodo [Data set]. 2025. DOI: 10.5281/zenodo.14640564. URL: https://doi.org/10.5281/zenodo.14640564.

11 BirdTeam: Bird Alarm Call Classifier

Authors: Vaishnavi Mendu, Moritz Plenz, Will Woof

11.1 Problem Addressed

Existing techniques for monitoring the health of bird populations are time consuming and costly. Often handling individual birds is required, causing undue stress. Even then, insight into the health of the whole population can only be inferred, as monitoring the health of each individual is logistically impossible.

11.2 Motivation

Birds encounter various forms of anthropogenic stress, such as sound pollution, traffic, and loss of natural habitat. Understanding where these stresses occur is crucial to implementing targeted and effective restoration efforts. By pinpointing these areas of stress, conservationists can allocate resources more efficiently to address specific ecological challenges.

Bioacoustic monitoring involves the detection of species from their vocalisations and has expanded rapidly as an ecological monitoring tool, due to recent developments in the field of machine learning. Now, individual species can be identified and located remotely and autonomously, enabling large-scale species-level monitoring. However, current tools only use vocalisations to identify species' presence. This overlooks the potential to identify different calls within a species to infer the behavioural and ecological context of the individual detected. Here, we focus on detecting bird stress via their alarm calls.

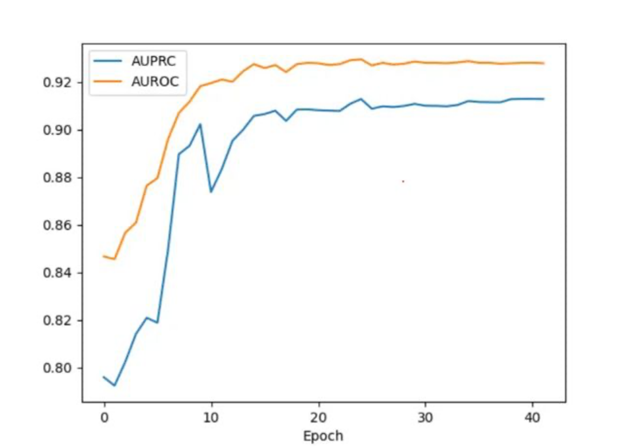

Figure 18: Training curve.

11.3 Solution

Birds are frequently monitored by recording their audio, which offers a cost-effective and scalable method for studying phenomena such as bird migration. Importantly, birds produce distinct calls for different situations, such as alert calls when they feel threatened. By classifying these calls, researchers can use them as proxies to assess bird stress levels. This approach makes it possible to identify locations where birds encounter harmful anthropogenic stress. The data obtained from tracking bird call types can guide efforts to plan and implement ecological restoration projects more effectively.

11.4 Non-technical Description

In this work, we demonstrate the effectiveness of classifying bird call types in a preliminary experiment. We train a classifier that can distinguish alarm calls from other calls or songs from three bird species from

California: the Pied-billed Grebe, California Quail, and Black-necked Stilt. We focused on alarm calls, as we hypothesize that these are most indicative of bird stress.

11.5 Technical Description

Data We use the BirdCLEF 2022 [1, 2] dataset, which includes audio recordings of 152 bird species, their scientific and common name, the latitude and longitude of the recording, and the type of the recorded call. To reduce the task complexity, we selected three species based on location and similarity in alarm calls. We use the call type as the target and aim to predict whether a recording is an alarm call in a binary classification.

Model We finetuned a BirdNet [3] model, which is a residual neural network designed to identify bird species from their audio recordings. It consists of a pre-processing block, four residual stacks and a classification block. The Python script was obtained from the model GitHub repository, and the model was fine-tuned with data segregated as alarm and non-alarm calls. Our trained classifier can be accessed through a simple prototype website that we created with HuggingFace Spaces. The website tags the audio file into either an alarm call or a non-alarm call.

Results Our classifier achieves over 90% AUPRC and AUROC on the training set (c.f. Figure 18), but due to time constraints during the hackathon, we could not extensively evaluate it on a test set. Manual inspection revealed a tendency to over-classify calls as alarm calls, which follow-up work should address. While not deployment-ready, our results provide preliminary evidence that alarm call classification is both feasible and important for restoration ecology.

11.6 Future Extensions

Our preliminary approach classifies the alarm and non-alarm class for three species located in California as proof of concept. Future work could extend the classifier to include more species and a wider variety of calls. With a more nuanced and detailed dataset one could further identify the age or gender of the bird, in order to better understand a species' demography. Of course, our classifier could also be further improved, for example, by including longitude and latitude information which is common practice in SOTA bird species classification models. Another potential way to improve our model's performance is to first classify the bird species, and then include the species information in the call classifier.

Finally, besides classifying birds from the audio files one could also directly classify anthropogenic sounds. This would allow direct linking of bird behavior to human impact by providing fine-grained spatiotemporal information on human activities.

References

[1] S. Kahl et al. "Overview of BirdCLEF 2022: Endangered bird species recognition in soundscape recordings". In: *Working Notes Papers of the CLEF 2022 Evaluation Labs*. 2022. URL: https://ceur-ws.org/Vol-3180/paper-154.pdf.

[2] A. Joly et al. "Overview of LifeCLEF 2022: An evaluation of Machine-Learning based Species Identification and Species Distribution Prediction". In: *Working Notes Papers of the CLEF 2022 Evaluation Labs*. 2022. URL: https://www.imageclef.org/system/files/Lifeclef2022%20%286%29.pdf.

[3] S. Kahl et al. "BirdNET: A deep learning solution for avian diversity monitoring". In: *Ecological Informatics* 61 (2021), p. 101236.