Emerging Applications of Large Language Models in Ecology and Conservation Science

Christos Mammides[1,2,3,*], Hao Gu[1,2,3], Thilina Nimalrathna[4], Naufal Rahman Avicena[1,2,3], Harris Papadopoulos[5], Ahimsa Campos-Arceiz[1,2,3,*]

1. Southeast Asia Biodiversity Research Institute, Chinese Academy of Sciences & Center for Integrative Conservation, Xishuangbanna Tropical Botanical Garden, Chinese Academy of Sciences, Mengla, Yunnan, 666303, China.
2. Yunnan International Joint Laboratory of Southeast Asia Biodiversity Conservation & Yunnan Key Laboratory for Conservation of Tropical Rainforests and Asian Elephants, Menglun, Mengla, Yunnan, 666303, China.
3. Yunnan International Joint Laboratory for the Conservation and Utilization of Tropical Timber Tree Species, Xishuangbanna Tropical Botanical Garden, Chinese Academy of Sciences, Mengla, Yunnan, 666303, China
4. School of Biological Sciences, The University of Hong Kong, Hong Kong SAR, China
5. Department of Electrical Engineering, Computer, Engineering and Informatics, Frederick University, Nicosia, 1036, Cyprus

* Corresponding authors: Christos Mammides; cmammides@outlook.com / Ahimsa Campos-Arceiz; ahimsa@xtbg.ac.cn

Acknowledgements:

We acknowledge the use of ChatGPT-4o for brainstorming ideas and editing the manuscript, including identifying areas for improvement and enhancing overall readability. We also used Gemini 2.5 Pro to generate the glossary and the descriptions of the prompting strategies, as detailed in the text. CM is thankful to the Marie Curie Alumni Association for providing access to Coursera, which enabled him to complete courses on generative AI, prompt engineering, and fine-tuning of large language models. This work was financially supported by the Southeast Asia Biodiversity Research Institute, Chinese Academy of Sciences (CAS-SEABRI: #Y4ZK111B05). CM was supported by The Wisdom Yunnan Project (202503AM140014), ACA by the High-End Foreign Experts Program of the Yunnan Revitalization Talents Support Plan, and TN and NRA by the CAS-ANSO Scholarship for Young Talents.

Emerging Applications of Large Language Models in Ecology and Conservation Science

Abstract

The emergence of large language models (LLMs) marks a major development in artificial intelligence, with potentially transformative implications for ecology and conservation science. Built on advanced deep-learning architectures, these models can support a wide range of tasks, from analysing unstructured texts to enhancing biodiversity monitoring and generating policy-relevant insights. This article synthesises emerging applications of LLMs across ecology and conservation, drawing on the wider literature and practical use cases. We highlight the potential of LLMs to streamline ecological workflows and accelerate evidence-based conservation, while also discussing key technical and ethical challenges, such as inaccurate and biased outputs, and unequal access. We offer recommendations for addressing these challenges to support the reliable and responsible use of LLMs, including strategies for improving output accuracy and ensuring proper validation. When implemented thoughtfully, LLMs can serve as a valuable addition to the ecologists' toolkit, enhancing scientific capacity and supporting broader efforts towards achieving biodiversity goals.

Keywords: BERT; ChatGPT; DeepSeek; Foundation Models; Generative AI; LLaMa; LLMs; Prompt Engineering

1. Introduction

Large Language Models (LLMs; see Box 1) represent an exciting breakthrough in **Artificial Intelligence** (AI) and are receiving growing attention due to their transformative potential (White et al. 2023a; Lam et al. 2024). Initially developed for natural language processing tasks, they are now being applied across a range of domains to support complex workflows, including summarising clinical trial reports (White et al. 2023b), analysing interview transcripts (Tai et al. 2024), and delivering localised climate insights to non-specialists (Koldunov & Jung 2024). LLMs are advanced deep-learning systems built on transformer neural network architectures and trained through self-supervised learning on vast and diverse text datasets (Lam et al. 2024; Morera 2024). **Foundation models**, such as GPT-4o, LLaMa 2, and DeepSeek-V2, are capable of capturing complex linguistic patterns and semantic relationships, enabling advanced language understanding and generation (Morera 2024). Chat-based LLMs, such as ChatGPT and LLaMA-Chat, which are further refined through supervised learning and reinforcement techniques, are able to support sophisticated conversational applications. LLMs, used in this article to refer broadly to foundation models and their chat-based counterparts, have recently demonstrated remarkable capabilities across a range of tasks (Lam et al. 2024), including extracting and synthesising large volumes of information and generating computer code (Cooper et al. 2024; Jhonnerie et al. 2024).

Given their strong performance and rapid development, LLMs are increasingly being integrated into the workflows of researchers and practitioners (Charness et al. 2025), including those in ecology and conservation science. However, their increased adoption also raises concerns, including the risk of generating inaccurate information and the underrepresentation of minority voices due to biases in training data (Reynolds et al. 2024; Urzedo et al. 2024). These issues are especially relevant in ecology and conservation science, which require not only sound ecological knowledge but also a deep understanding of the social and cultural contexts (Sandbrook 2024). Addressing these challenges will be essential to ensure that the use of LLMs is both effective and equitable. Yet, despite the growing interest in LLMs, practical guidance on their potential uses, strengths, and limitations remains scattered, and often in resources beyond those typically consulted by ecologists and conservation practitioners. Moreover, while much of the current attention on LLMs focuses on their chat-optimised interfaces, emerging evidence suggests that some of the most impactful and innovative uses will likely come from embedding LLMs into backend systems supporting specialised workflows. Collectively, these knowledge gaps highlight the need for a timely synthesis.

In this article, we provide an overview of emerging applications of LLMs in ecology and conservation science, highlighting key opportunities, current challenges, and offering practical recommendations to support their effective adoption. Our analysis draws primarily from use cases in the wider academic literature and examples from our own experience, including training. To identify relevant literature, we searched Web of Science, Scopus, and Google Scholar using the terms: [“LLMs” OR “Large Language Model*”] AND [“Ecology” OR “Biodiversity” OR “Conservation”]. We also reviewed the reference lists of all retrieved articles to record additional sources (see Supplementary Materials). After initial screening for relevance and removing duplicates, we identified 123 documents relevant to our topic (Figure 1). Recognising the fast-evolving nature of this field, in addition to peer-reviewed journal articles (n = 83), we also considered preprints (n = 34) and conference proceedings (n = 6). In the sections that follow, we review promising and fast-developing applications of LLMs in ecology and conservation science (Section 2), examine key technical and ethical challenges (Section 3), and conclude with practical solutions and recommended best practices (Section 4).

2. Emerging Applications of LLMs

LLMs can support a variety of tasks, from automating labour-intensive processes such as data extraction to assisting with data analysis, improving communication and outreach, and informing policy. They hold the potential to streamline workflows, expand access to information, and accelerate evidence-based conservation.

2.1 Extracting Ecological Data

A major bottleneck in ecological research lies in extracting key insights from unstructured sources (Marcos et al. 2025), such as scientific publications and reports. While these sources hold valuable information, accessing and synthesising it is often laborious and time-consuming. Recent applications suggest that LLMs may offer a promising solution (Farrell et al. 2024; Gurr et al. 2024; Castro et al. 2024; Elliott & Fortes 2024), stemming from their ability to rapidly process large volumes of unstructured text and identify relevant content using patterns learned from modelling complex linguistic structures (Castro et al. 2024).

Gougherty and Clipp (2024) used text-bison-001, a publicly available LLM part of Google's PaLM 2 family of models, to extract data on plant pathogens and their host plants from the academic literature, achieving faster extraction speeds than those of human reviewers while maintaining high accuracy. Similarly, Keck et al. (2025) used OpenAI's GPT-4o to extract species interactions from over 80,000 scientific articles, and Marcos et al. (2025) developed an LLM-based workflow to extract species traits, with both studies reporting encouraging levels of accuracy. Fu et al. (2025) employed DeepSeek-R1 to analyse 247 Chinese court cases, successfully identifying hotspots of wildlife crime involving sea turtles. Scheepens et al. (2024) evaluated GPT-4's ability to extract taxonomic information from article abstracts. Despite some inaccuracies due to **hallucinations**, the preliminary results were promising, demonstrating the potential of LLMs for large-scale data extraction from unstructured sources. It should be noted, however, that accuracy can vary depending on the model used and the specific task, underscoring the need for careful model selection (Castro et al. 2024). Moreover, effective information extraction will oftentimes require sophisticated workflows that go beyond the use of standalone foundation models, incorporating robust error-mitigation mechanisms (Section 4) and reliable retrieval strategies (Iyer et al. 2025).

Nonetheless, LLMs can also assist in extracting information from structured sources, such as online databases. In a recent example, researchers used OpenAI's GPT-4 model to develop a chatbot enabling users to query the Integrated Digitized Biocollections (iDigBIO) database using natural language (Elliott et al. 2024). The ability to interact with databases through natural language is an active area of research beyond ecology and conservation (Miao et al. 2024), and holds significant promise for expanding access to information by reducing reliance on specialised technical skills.

2.2 Accelerating Literature Reviews and Evidence Synthesis

LLMs could also streamline how researchers and practitioners conduct literature reviews, synthesise evidence (Berger-Tal et al. 2024; Reynolds et al. 2024), and even identify emerging topics (Gurr et al. 2024; Ji et al. 2025), without being constrained by siloed disciplinary thinking. With the volume of scientific publications growing exponentially, identifying and reviewing relevant documents is becoming increasingly challenging. Yet, comprehensive reviews are critical for advancing science and guiding evidence-based conservation (Berger-Tal et al. 2024; Iyer et al. 2025). In a recent study, Chang et al. (2024) developed a machine learning pipeline that incorporates LLMs to process over two million scientific articles and assess how nature-based solutions promote human well-being and biodiversity conservation. Krishna Moorthy et al. (2025) implemented a GPT-based workflow to process large volumes of scientific literature to synthesise information on study locations, biome types, and quantitative metrics. By combining *iterative prompting* (Table 2; Section 4.2) with manual validation, they demonstrated how LLMs can support efficient and scalable ecological literature reviews while reducing errors.

Although these early applications tend to require complex workflows and high technical expertise, they nevertheless demonstrate the potential of LLMs to accelerate synthesis (Reynolds et al. 2024), particularly when paired with mechanisms to mitigate inaccuracies. As these tools become more accessible and reliable, their role in synthesising information will likely grow. One especially promising application is their potential to support living evidence syntheses, i.e., continuously updated reviews that integrate new findings as they emerge

(Mitchell et al. 2025). Moreover, commercial AI-powered research assistants, such as Elicit and Consensus, can further support literature reviews through their LLM-based, user-friendly interfaces. However, issues around affordability and access, especially for researchers with limited resources, raise concerns about digital inequalities (Section 3.5).

2.3 Leveraging Publicly Available Data for Ecological Insights

In addition to information syntheses, LLMs show promise in supporting a range of other analytical tasks (Frazier & Song 2025). For instance, they can help analyse ecologically relevant information from publicly available sources such as social media (Giebink et al. 2024), including performing sentiment analysis to better understand human–nature interactions and inform conservation planning (Wei & Hou 2023). In a recent study, researchers used GPT-3.5 to analyse social media posts and study public perceptions of urban green spaces in Singapore, finding generally positive sentiments and identifying the factors shaping those perceptions (Zhang & Su 2024). In China, researchers used LLMs to analyse 1,849 online travelogues posted by visitors to National Forest Parks, providing insights into forest experiences that can inform park management and design (Wei & Hou 2023). Similarly, in Brazil, researchers used ChatGPT to analyse Tripadvisor reviews of two protected areas, finding high overall satisfaction, but also identifying concerns regarding outdated information and sanitation, offering actionable insights for park management (de Souza et al. 2024). Together, these examples demonstrate the potential of LLMs for scalable analysis of people's experiences with nature, offering valuable insights for landscape planning and management (Frazier & Song 2025).

Another promising application of LLMs is their use in monitoring media to provide real-time, relevant ecological insights. NewsPanda and NewsSerow are two such examples that support conservation efforts by automating the detection of environmentally relevant media articles. NewsPanda uses a fine-tuned BERT-based model with active learning to classify articles and extract key information, and has been deployed by WWF in multiple countries to monitor thousands of conservation sites (Keh et al. 2023). NewsSerow focuses on low-resource languages, using summarisation, few-shot classification, and self-reflection with LLMs to identify conservation-relevant content with minimal training data (Jain et al. 2024).

LLMs can also contribute to combating illegal trade by enabling the monitoring of online advertisements. In a novel use case, researchers developed a cost-effective approach to detect wildlife trafficking on e-commerce platforms by using LLMs to generate pseudo-labels for a small subset of ads (Barbosa et al. 2025). These labels were then used to train specialised classifiers that accurately identify wildlife-related ads. This approach can potentially enable scalable monitoring of illicit wildlife trade and support real-world enforcement and research efforts. Importantly, the use of LLMs to generate pseudo-labels for training classifiers holds promise for a broader range of applications beyond online information monitoring.

2.4 *Supporting Code Generation and Programming in Ecology*

Ecological research often involves complex analyses requiring advanced statistical and programming skills. Yet many training programs worldwide offer limited, if any, formal training in these areas (Mammides and Papadopoulos, 2024). With appropriate oversight, LLMs can support researchers in conducting analyses, including generating the necessary code (Campbell et al. 2024; Cooper et al. 2024; Jhonnerie et al. 2024). Specialised tools, such as GitHub Copilot, DeepSeek-Coder, and the "ellmer" package in the R Programming Language, are already helping facilitate this process (Guo et al. 2024). LLMs can also be effective in troubleshooting and explaining existing code (Merow et al. 2023), thereby supporting the learning process. They can also help tidy and comment code, making it easier for researchers to potentially share their code in publications (Mammides & Papadopoulos 2024), a practice which remains uncommon despite its importance for reproducible research.

LLMs have proven useful for translating code between programming languages. This can be especially helpful for ecologists, who are often more familiar with the R Programming Language, while many cutting-edge applications, such as machine learning tools, are developed in Python. By easing this transition, LLMs can expand access to a wider range of analytical tools. More broadly, they can help bridge disciplinary divides by offering personalised, around-the-clock guidance on unfamiliar concepts and methods (Mammides & Papadopoulos 2024).

2.5 *Experimental Applications and Multimodal Innovations*

Recent studies have also begun exploring more experimental applications of LLMs, with potentially transformative implications. Sastry et al. (2023) succesfully integrated LLaMa 2 into a species distribution modelling (SDM) framework to improve predictions of species ranges. Leblanc et al. (2025) used Pl@ntBERT, a transformer-based LLM developed specifically for ecological applications, to classify and map habitats across Europe at very high resolution, leveraging species distribution maps generated using deep-SDMs.

As language models evolve into multimodal tools capable of processing and generating formats beyond text, such as images and audio, their potential for novel ecological applications will further grow (Miao et al. 2024). Currently, there is strong interest in integrating multimodal LLMs with edge devices, such as camera traps, drones, and acoustic sensors, to advance biodiversity monitoring applications (Robinson et al. 2024; Zhao et al. 2024). Dussert et al. (2025) tested whether pre-trained multimodal LLMs could identify animal behaviour patterns from images obtained through camera traps, a widely used tool for monitoring biodiversity. They found that LLMs could potentially automate this typically laborious task, notably without model training or labelled datasets, which are often costly to produce but required by conventional machine learning classification techniques (Dussert et al. 2025).

In another novel use case involving camera trap images, Fergus et al. (2024) showed that integrating multimodal LLMs, in this case Microsoft's Phi-3.5, with **retrieval-augmented generation (RAG)** techniques (Section 4.4), can improve species identification and even provide contextually rich information about the species detected, going beyond what the camera alone can capture. For example, their LLM-integrated workflow could be potentially used to generate reports detailing the species in an area, as well as relevant information about their ecology and conservation status (Fergus et al. 2024). Similarly, NatureLM-audio, an LLM-based model designed for bioacoustic applications, can be used to detect, classify, and interpret animal vocalisations using natural language prompts (Robinson et al. 2024). While the long-term value of these applications and the extent to which they can be scaled up remain to be seen, the potential of LLMs is already evident.

2.6 Enhancing Science Communication and Outreach

LLMs are already supporting science communication (Richards et al. 2024), including by reducing language barriers and promoting the democratisation of science. While English is the dominant language in scientific publishing, most researchers globally are non-native speakers. Studies have shown that language proficiency can affect citation rates and overall career advancement (Hannah et al. 2025). LLMs have proven effective in improving, for example, scientific writing by refining grammar and enhancing clarity, tasks that often require costly editing services or reliance on personal networks (Zenni & Andrew 2023).

LLMs could further reduce language barriers by making research published in foreign languages accessible to the global scientific community (Valdez et al. 2024). While not yet perfect, LLM-based translation tools generally outperform earlier technologies, enabling researchers to engage with relevant studies. This is especially valuable in conservation, where important ecological insights are often published in local journals. As LLMs continue to advance, they are likely to transform also other aspects of science communication, including conference presentations.

Beyond reducing language barriers, LLMs can support science communication more broadly by facilitating, for example, the dissemination of targeted information (Richards et al. 2024). Lewers (2024) demonstrated that LLM-powered chatbots can simplify complex biodiversity informatics standards, easing their adoption by researchers. Similarly, Wang et al., (2024) developed ChatBBNJ, a customised question-answering system built using a prompt-based configuration to focus on topics related to biodiversity beyond national jurisdiction. The dissemination of specialised information can also be supported by domain-specific LLMs, which are **fine-tuned** through further training. For example, OceanGPT is a LLaMA-based model retrained on more than 67,000 ocean science documents to answer oceanographic questions and serve as a virtual expert (Bi et al. 2023). Another example is MarineGPT, which was fine-tuned on a marine-specific dataset to provide contextual and accurate information on marine environments (Zheng et al. 2023). Because these models are refined or adapted using domain-specific information, they can deliver more accurate results for specialised tasks compared to general-purpose models such as ChatGPT, Gemini, and Qwen, provided, though, that key issues, including biases associated with the training data, are adequately addressed (see Ziegler et al. 2024; Section 3).

LLMs also show potential for engaging broader audiences with ecological topics (Richards et al. 2024), an important yet often overlooked dimension of science communication. Ecologists are typically trained to communicate in technical language that may not resonate with non-experts. LLMs can help adapt messages for diverse audiences and platforms, making research more accessible (D'Souza et al. 2025). Moreover, specialised models, such as those mentioned above, can deliver tailored ecological information to the public, stakeholders, and policymakers. Much like business chatbots designed for customer support, these models can be adapted to deliver region- or topic-specific ecological insights.

General-purpose pre-trained foundation models can also contribute to the dissemination of broader scientific knowledge with careful use (Rodas-Trejo & Ocampo-González 2024). In an interesting use case, researchers found that widely used LLMs, such as ChatGPT, can counter sensationalised media reports about wildlife risks to humans and livestock by providing more accurate assessments, potentially reducing human-wildlife conflict (Santangeli et al. 2024). Similarly, researchers in Mexico evaluated ChatGPT as a source of natural history information on wild mammals, finding it useful, though its outputs required validation (Rodas-Trejo & Ocampo-González 2024), which should be a standard practice when using foundation models to generate specialised knowledge (Section 4.3).

2.7 Informing Policy and Decision-Making Processes

There is often a disconnect between academic research and policymaking, shaped not only by differing priorities but also by contrasting norms in how information is communicated. LLMs can help bridge this gap by translating scientific findings into policy-relevant insights and synthesising information in ways that align with the needs of decision-makers (Reynolds et al. 2024). In a novel application, DeSantis et al. (2024) used OpenAI's GPT-3.5 model to analyse the national biodiversity targets of individual countries and assess how well they align with the adopted global biodiversity targets, finding crucial gaps and specific areas for improvement.

While policy applications of LLMs in ecology and conservation science are still in the very early stages, their use in fields such as law and environmental science is more advanced and offers useful examples (Larosa et al. 2025). Many of these applications involve adapting the core capabilities of LLMs highlighted above, such as summarisation, information retrieval, and natural language interaction, for policy-focused tasks (Larosa et al. 2025). Examples include distilling complex scientific documents into actionable insights, deploying chatbots to communicate policy guidance to stakeholders, and enhancing public engagement through clearer, more accessible communication (Larosa et al. 2025).

A particularly innovative and promising application is the integration of LLMs into multi-agent systems for policy simulation (Kalyuzhnaya et al. 2025; Sreedhar et al. 2025). Socioecological systems are characterised by complex interactions, feedback loops, and diverse stakeholder perspectives, all of which can be difficult to map and incorporate into policy decisions. LLM-powered agents, each representing the views and behaviours of different actors, such as policymakers, land users, or conservation advocates, can simulate these interactions and help decision-makers anticipate how policies may unfold in practice (Kalyuzhnaya et al. 2025; Sreedhar et al. 2025). This approach enables a more nuanced assessment of policy effectiveness, including potential unintended consequences, supporting the development of policies that are both robust and context-appropriate.

2.8 Evolving Roles and Broader Applications

The use cases and opportunities outlined above, ranging from data extraction and information synthesis to habitat mapping and biodiversity monitoring, illustrate only part of the wide-ranging and rapidly evolving potential of LLMs and how these models are beginning to reshape ecological workflows. Broader applications discussed in the wider academic literature, such as the use of LLMs in education (Tupper et al. 2024), law, and social sciences, can further amplify their relevance and value in ecology and conservation science. Moreover, the emergence of **agentic LLM systems** (Plaat et al. 2025), such as AutoGPT (built on top of OpenAI's GPT-4), which are designed to perform complex tasks with minimal human input (Plaat et al. 2025), may further accelerate the automation of ecological workflows, although concerns arise about oversight, reliability, and equity.

3. Current Challenges

While LLMs show clear promise, their use also brings a range of practical, technical, and ethical challenges, many of which remain underexamined amid rapid development and must be addressed to ensure **Trustworthy AI** (Liu et al. 2023). Key concerns include inaccurate or even fabricated content, limited transparency and reproducibility, and high technical and computational demands, especially for more advanced applications. Ethical considerations include digital inequalities, environmental impact, and the risk of eroding foundational skills as reliance on these tools grows.

3.1 Hallucinations and Fabricated Content

A major concern that emerged with the public release of LLMs is their tendency to hallucinate, i.e., to generate plausible but factually incorrect information. A typical example in research contexts is fabricated references that appear credible but are entirely fictitious, often combining unrelated author names and article titles. This issue was especially pronounced when popular models like ChatGPT were first released in 2022 without internet access, but it persists even today with web-enabled LLMs, especially for queries on less common or out-of-distribution topics not covered in the models' training data. Because hallucinations can go undetected without manual verification, they present a barrier to the effective use of LLMs.

3.2 Inaccurate and Biased Information

Inaccuracies in LLM outputs are not solely due to hallucinations; they can also result from conflicting, incomplete, or outdated training data or from a model's inability to apply contextual or domain-specific reasoning (Berger-Tal et al. 2024). For example, earlier versions of ChatGPT were trained only on data up to 2021, excluding more recent information. Even with current models, many of which have access to more up-to-date data and real-time retrieval capabilities, inaccurate information remains common, underscoring the need for cautious use and robust validation mechanisms (Section 4.3).

More broadly, LLMs can produce inaccurate or contextually inappropriate content also when they fail to account for geographic or temporal relevance, an issue of particular concern in fields such as conservation science, where context is crucial for effective decision-making. LLM outputs have also been shown to reflect biases in their training data (Berger-Tal et al. 2024; Urzedo et al. 2024). For example, when queried on ecological restoration topics, ChatGPT's responses were found to prioritise the dominant perspectives from researchers in high-income countries, while underrepresenting voices from lower-income regions and Indigenous communities (Urzedo et al. 2024). Similarly, specialised LLM-based chatbots, designed to deliver targeted information on a specific topic, can instead propagate biases (Ziegler et al. 2024), depending on the data used to refine the models and the design choices made during their development.

Inaccurate output can also result from technical limitations. For instance, model parameters, such as the **context window** and maximum output length, can influence the quality, accuracy, and completeness of a response. Generally, the longer the user prompt or the more extensive the requested output, the higher the likelihood of erroneous responses. That said, these limitations are becoming less prominent with **long-context LLMs**, which are designed to process large amounts of text in a single input, and **memory-augmented LLMs**, which can store, retrieve, and update information as needed beyond the fixed-size input context window. Nonetheless, users should be mindful of these constraints and how they influence the performance of the different models used, and apply effective strategies, such as prompt engineering (Section 4.2), to mitigate errors, ensure content accuracy, and reduce the risk of flawed outputs influencing future model training.

3.3 Lack of Transparency and Reproducibility

Improving the reproducibility of scientific research has become a major focus in recent years. However, the spread of LLMs introduces new challenges. Three key issues contribute to this concern. Firstly, LLMs and deep learning models more broadly often operate as black boxes, making it difficult to trace how outputs are generated. Although ongoing efforts towards **Explainable AI** (Liu et al. 2023) could reduce this issue in the future, it is currently a major concern. Secondly, LLM responses are non-deterministic, meaning that the same prompt can yield different outputs. While this variability can be partially controlled by adjusting the model's **hyperparameters** (e.g., by setting the temperature to zero to increase consistency; Iyer et al. 2025), doing so may limit the nuance in the model's responses. Thirdly, LLMs evolve rapidly, with frequent releases. Outputs may differ across versions, and even if the model used is documented, it may later become inaccessible or unsupported.

3.4. Technical Complexity and Computational Demands

Another major challenge associated with the use of LLMs is the technical complexity and computational resources needed for many applications. While simpler tasks, such as scientific writing improvements, can be handled through the available user-friendly chat-optimised interfaces (Mammides & Papadopoulos 2024; Santangeli et al. 2024), more advanced and potentially impactful applications necessitate access to models via application programming interfaces (APIs) and their integration into sophisticated workflows (e.g., Chang et al., 2024). These workflows typically demand advanced programming skills and computational resources that many researchers and practitioners lack. While such barriers are not unique to LLMs, they are amplified by the models' transformative potential and the skills required to fully leverage them. These issues highlight the importance of accessible training as well as interdisciplinary collaboration between ecologists, who can offer domain-specific insight, and machine learning experts who can provide the technical expertise (Murray et al. 2025) (Section 4.6).

3.5 Ethical Concerns

The technical complexities and computational demands also raise ethical concerns. While these models can support the democratisation of research, e.g., by helping to overcome language barriers (Valdez et al. 2024), they may also exacerbate digital inequalities if access becomes unequal through premium pricing, limited infrastructure, or lack of technical skills (Murray et al. 2025). There is already evidence suggesting that resource limitations in biodiverse but economically disadvantaged regions contribute to scientific inequities (Mammides et al. 2016; Campos-Arceiz et al. 2018); unequal access to powerful LLMs risks deepening this divide (Reynolds et al. 2024).

Another major ethical concern relates to how these models are trained and the sources used in the process (Liu et al. 2023). Beyond the widely discussed issues, such as copyright infringement, where content may be used without the creators' consent, there is a broader problem of representativeness (Sandbrook 2024; Urzedo et al. 2024). Training data often draws from readily available sources that do not necessarily reflect all perspectives (Sandbrook 2024). For example, viewpoints expressed in low-resource languages or by Indigenous communities are frequently underrepresented (Urzedo et al. 2024), which means LLM outputs may overlook these critical voices, potentially biasing decision-making (Gewin 2025). Addressing this issue while respecting data sovereignty (Section 4.4) is essential for the trustworthy and equitable application of LLMs (Gewin 2025).

There is also valid concern that the current hype around LLMs, and AI more broadly, may divert resources from essential, conventional practices in ecology and conservation (Reynolds et al. 2024). Field-based studies already represent a declining proportion of the literature, and ecology students are spending less time in the field, potentially leading to a disconnect with nature and its processes (Soga & Gaston 2025). If the rise of LLMs further shifts attention and resources away from the collection of primary field observations, which represent the backbone of ecological knowledge, it could represent a negative development.

A further concern is the risk of deskilling (Reynolds et al. 2024), for example, in tasks related to computer coding. Learning to code involves hands-on engagement (e.g., writing, debugging, and refining code), which fosters deeper understanding. If ecologists and conservation scientists increasingly rely on LLMs to perform these tasks, they may miss essential opportunities to develop the skills needed to critically evaluate and troubleshoot code on their own. This concern extends beyond coding and applies to other key tasks, such as scientific writing and statistical analyses.

Finally, there is growing concern regarding the environmental impact of training, deploying, and using LLMs (Rillig et al. 2023; Sandbrook 2024). Training large models requires significant energy and generates substantial emissions, while data centres consume vast volumes of water for cooling (Sandbrook 2024). Although training is typically more resource-intensive than usage (inference), the impact varies with model size and efficiency. Mitigation efforts include using renewable energy, optimising deployment, and adopting energy-efficient architectures like DeepSeek's Mixture of the Experts, which activates only a subset of model parameters during inference. However, these issues underscore the need to use LLMs thoughtfully, reserving them for tasks that truly require their capabilities. Otherwise, applications intended to support conservation may ironically end up harming the environment.

4. Recommendations for more effective and responsible use of LLMs

Effective use of LLMs requires recognising their risks and addressing them through both technical solutions and thoughtful use. While model advances may reduce challenges like hallucinations, many issues depend on how users interact with these tools. Techniques such as **prompt engineering**, **retrieval-augmented generation (RAG)**, and **human-in-the-loop** oversight, described below, can improve reliability and inclusivity (Reynolds et al. 2024). Sharing prompts, workflows, and best practices supports transparency, reproducibility, and impact. In this section, we offer practical guidance to help maximise benefits and mitigate risks in ecological and conservation applications.

4.1 Detecting Inaccuracies and Measuring Uncertainty

As noted in Section 3, a key limitation of LLMs is their tendency to generate inaccurate information, including hallucinations (Kumar et al. 2025; Iyer et al. 2025). Although ongoing research on detecting and mitigating such errors may reduce their frequency, they are unlikely to be fully eliminated (Mora-Cross & Calderon-Ramirez 2024). To address this issue, various strategies can be employed to improve model reliability. These include post-hoc fact-checking (e.g., cross-referencing output with authoritative sources), self-verification (prompting the model to assess or justify its own responses), and confidence estimation (assessing the model's uncertainty through internal metrics or response patterns) (Mora-Cross & Calderon-Ramirez 2024; Kumar et al. 2025). Importantly, output accuracy will vary depending on the model used and the specific task, making careful model selection crucial for optimal results. Tools such as LMArena (https://lmarena.ai), which allow users to compare different LLMs across multiple benchmarks and performance metrics, support more informed task-specific decisions.

In practice, users can also adopt simpler techniques that serve as practical forms of confidence estimation. For example, submitting the same query multiple times and comparing the variability of outputs can help flag potentially hallucinated content, which tends to lack consistency. Another approach involves using a second LLM to fact-check the first, though this can be computationally expensive and increases the environmental footprint of LLM usage. More broadly, effective prompt engineering, discussed in the next section, can also help reduce inaccuracies by providing clearer instructions and more targeted context.

4.2 Prompt Engineering

Prompt engineering involves crafting input prompts to guide LLMs toward relevant and reliable outputs (White et al. 2023a). Well-designed prompts improve clarity, reduce ambiguity, and better align responses with the user's intent and domain-specific needs (Sahoo et al. 2024). For instance, explicitly asking the model to respond with "NA" when information is not clearly available can reduce hallucinations (Marcos et al. 2025), which often arise from the assumption that an answer must always be provided. Other best practices include: (a) using clear, specific language to define the task (e.g., in *instruction-based prompting*; Table 1); (b) breaking down complex tasks into manageable steps (*prompt chaining*); and (c) providing examples of the desired output and format (Sahoo et al. 2024), a highly effective approach known as *few-shot prompting* (White et al. 2023a; Table 1).

Two particularly effective strategies to improve accuracy are *role prompting* and *contextual prompting* (White et al. 2023a; Sahoo et al. 2024) (Table 1). Role prompting involves instructing the model to adopt a specific identity or perspective (e.g., "Assume you are an ecologist designing an experiment…"), which can improve both the relevance and factual correctness of the response (Sahoo et al. 2024). Contextual prompting entails providing detailed background information or situational context within the prompt, enabling the model to generate more targeted, nuanced, and reliable responses. Context may include domain-specific details, constraints, or examples that anchor the model's reasoning process (Sahoo et al. 2024).

Another widely recommended technique is *Chain-of-Thought (CoT) prompting* (Table 1), which requires the model to generate a step-by-step explanation of its reasoning (Farrell et al. 2024). This method enhances interpretability by making the model's logic explicit, thereby facilitating evaluation of the correctness and coherence of its conclusions. CoT prompting is particularly effective for complex reasoning tasks and helps reduce the risk of shallow or unsupported answers (Sahoo et al. 2024). Additional techniques include *iterative prompting*, where prompts are refined step-by-step based on previous outputs to improve accuracy and depth, and *prompt ensembling*, which involves using multiple prompts to generate varied responses for the same task and combining them to produce a more robust and consistent result (Table 1).

Curated prompt catalogues are becoming increasingly common (White et al. 2023a; Sahoo et al. 2024), especially outside the fields of ecology and conservation science. As LLM use expands, researchers are encouraged to share their prompt techniques, for instance, in the supplementary materials of their publications. This practice can promote transparency, reproducibility, and collective innovation, much like the open sharing of computer code. Additionally, emerging frameworks such as DSPy provide a programmatic approach to prompt engineering by structuring prompts as modular, reusable components (Barbosa et al. 2025), allowing for systematic tuning, better scalability and more maintainable LLM workflows.

4.3 Validation

To use LLMs reliably, users must integrate validation measures into their workflows (Reynolds et al. 2025). One practical technique, known as *citation prompting*, is to instruct models to include footnotes or citations with their responses that can be manually verified. Another effective validation technique involves cross-checking LLM outputs against established datasets (Vaghefi et al. 2023). In more formal evaluations, users could employ benchmarking to assess a model's performance against task-specific, curated datasets, allowing for the calculation of accuracy and other relevant performance metrics (Bi et al. 2023). More advanced applications may also link models to external databases (Kumar et al. 2025), a form of retrieval-augmented generation (see Section 4.4). For example, if a model is asked to recommend strategies for addressing threats to a particular species, its output can be validated through authoritative sources, either manually or using automated methods. **Self-consistency** testing is also a useful method for validating the reliability of LLM outputs (Kumar et al. 2025). It involves generating multiple responses using the same or slightly varied prompts and assessing the extent to which the answers converge on a consistent response.

Crucially, these technical approaches should be complemented by **human-in-the-loop (HITL)** validation, in which domain experts review (Berger-Tal et al. 2024), interpret and correct LLM outputs as needed before they are used in research or policy (Reynolds et al. 2024). This layer of oversight helps ensure that the outputs are not only technically accurate but also ecologically relevant and ethically sound, which is especially important in conservation applications where decisions can affect multiple stakeholders.

4.4. Retrieval-Augmented Generation (RAG)

RAG can improve the output of LLMs by dynamically incorporating external documents at inference time, addressing the limitations of foundation models that rely solely on pre-trained knowledge (Pichai 2023; Kumar et al. 2025). Real-time access to up-to-date and specialised information can enhance the accuracy and relevance of the responses (Shelby & da Silva 2024; Kumar et al. 2025). This is especially important in fields like conservation science, where up-to-date, context-specific knowledge is critical for effective interventions.

RAG frameworks can also help address some of the ethical concerns mentioned above, such as the underrepresentation of minority groups and Indigenous perspectives. By incorporating curated, diverse, and context-specific knowledge bases, such as Indigenous ecological knowledge, they can enable more inclusive and contextually accurate outputs (Pichai 2023). When proprietary data is involved, RAG can be implemented using locally run, offline language models to respect data sovereignty. An added advantage is that these smaller models, which may be equally effective depending on the task, tend to have a lower environmental impact.

4.5 Fine-tuning and Specialised Models

While powerful, pre-trained foundation models such as GPT and Qwen often lack the specialised knowledge required for ecological applications. **Fine-tuning** these models using domain-specific datasets can significantly improve the accuracy, relevance, and contextual appropriateness of their outputs (Abdelmageed 2023; Bi et al. 2023). An example is OceanGPT, a LLaMa-based model fine-tuned on over 67,000 documents related to ocean science (Bi et al. 2023). However, not all foundation models are available for fine-tuning, and in some cases, retraining can lead to "catastrophic forgetting", where the model loses some of its previously acquired knowledge when updated with new data. Perhaps more importantly, retraining a model is resource-intensive and computationally demanding, making it impractical in many cases.

Alternatively, customised models, such as ChatBBNJ (Wang et al. 2024), can be developed for targeted purposes without retraining. A common approach involves appending predefined instructions or contextual information to each prompt, which is essentially what custom GPTs represent. These lightweight adaptations can support domain-specific applications, such as public awareness campaigns, educational outreach, or policy support, offering more accessible and effective tools.

4.6 Capacity Building

While the strategies outlined above can largely enhance the reliability and effectiveness of LLM applications, it is crucial that users are provided with training that builds capacity (Murray et al. 2025) and disseminates evolving best practices. Without adequate training, researchers may use these transformative tools without fully recognising their limitations or how to address them. Formal training can also help unlock the full potential of LLMs, enabling more innovative and impactful applications (Mammides & Papadopoulos 2024). Such training could take various forms, including integration into university curricula or shorter workshops offered at conferences, for example. Equally important is fostering a culture of transparency around how LLMs are used and evaluated, as well as inclusivity in how they are developed (Murray et al. 2025).

5. Conclusion

LLMs will play an increasingly integral role in how ecological and conservation science is conducted, communicated, and applied. From automating labour-intensive tasks, such as data extraction and evidence synthesis, to enhancing biodiversity monitoring and improving communication across languages and disciplines, LLMs have the potential to increase the efficiency, accessibility, and impact of research. However, these opportunities come with significant challenges, ranging from technical limitations associated, such us inaccuracies and difficulties in scaling up applications, to broader concerns about equity, reproducibility, and environmental sustainability. Addressing these issues will require technological advances and a strong commitment to best practices, ethical and regulated standards, and inclusive design. Moving forward, the integration of LLMs into ecological and conservation research should be guided by trustworthy use and transparent reporting, interdisciplinary collaboration, and ongoing critical evaluation of their strengths and limitations. Training and support for researchers, especially those in under-resourced settings, will be essential to ensure equitable access and meaningful participation in this technological shift. With thoughtful usage, LLMs can accelerate scientific discovery and help democratise knowledge production and application in support of global biodiversity goals.

References

Abdelmageed N. 2023. BiodivBERT: a Pre-Trained Language Model for the Biodiversity Domain.

Barbosa J, Gondhali U, Petrossian G, Sharma K, Chakraborty S, Jacquet J, Freire J. 2025. A Cost-Effective LLM-based Approach to Identify Wildlife Trafficking in Online MarketplacesDOI: 10.48550/ARXIV.2504.21211. arXiv. Available from https://arxiv.org/abs/2504.21211 (accessed June 16, 2025).

Berger-Tal O, Wong BB, Adams CA, Blumstein DT, Candolin U, Gibson MJ, Greggor AL, Lagisz M, Macura B, Price CJ. 2024. Leveraging AI to improve evidence synthesis in conservation. Trends in Ecology & Evolution. Elsevier.

Bi Z, Zhang N, Xue Y, Ou Y, Ji D, Zheng G, Chen H. 2023. OceanGPT: A Large Language Model for Ocean Science Tasks. arXiv. Available from https://arxiv.org/abs/2310.02031 (accessed December 25, 2024).

Campbell H, Bluck T, Curry E, Harris D, Pike B, Wright B. 2024. Should we still teach or learn coding? A postgraduate student perspective on the use of large language models for coding in ecology and evolution. Methods in Ecology and Evolution **15**:1767–1770. Wiley Online Library.

Campos-Arceiz A, Primack RB, Miller-Rushing AJ, Maron M. 2018. Striking underrepresentation of biodiversity-rich regions among editors of conservation journals. Biological Conservation **220**:330–333.

Castro A, Pinto J, Reino L, Pipek P, Capinha C. 2024. Large language models overcome the challenges of unstructured text data in ecology. Ecological Informatics **82**:102742.

Chang CH et al. 2024. Global evidence of human well-being and biodiversity impacts of natural climate solutions. Nature SustainabilityDOI: 10.1038/s41893-024-01454-z. Available from https://www.scopus.com/inward/record.uri?eid=2-s2.0-85211482907&doi=10.1038%2fs41893-024-01454-z&partnerID=40&md5=afc0c6b114e5875ea1b0bdbc6b7ad49d.

Charness G, Jabarian B, List JA. 2025. The next generation of experimental research with LLMs. Nature Human Behaviour **9**:833–835.

Cooper N, Clark AT, Lecomte N, Qiao H, Ellison AM. 2024. Harnessing large language models for coding, teaching and inclusion to empower research in ecology and evolution. Methods in Ecology and Evolution. Wiley Online Library.

de Souza AMF, Brandão ACL, Di Paolo ÍF, de Souza Conte TNM, da Silva Miranda J, da Silva Quinto WA. 2024. Use of ChatGPT and Tripadvisor data for discovering positive and negative aspects of the BioParque Vale Amazônia and the Carajás National Forest between 2014 and 2022. Revista de Gestão e Secretariado **15**:e3796–e3796.

DeSantis N, Supples C, Phillips L, Pigot J, Ervin J, Wade T. 2024. Leveraging AI for Enhanced Alignment of National Biodiversity Targets with the Global Biodiversity Goals. Nature-Based Solutions:100198. Elsevier.

D'Souza J et al. 2025, April 5. Reflections from the 2025 EcoHack: AI & LLM Hackathon for Applications in Evidence-based Ecological Research & Practice. Computer Sciences. Available from https://ecoevorxiv.org/repository/view/8910/ (accessed June 16, 2025).

Dussert G, Miele V, Van Reeth C, Delestrade A, Dray S, Chamaillé-Jammes S. 2025. Zero-shot animal behaviour classification with vision-language foundation models. Methods in Ecology and Evolution:2041–210X.70059.

Elliott M, Luciano M, Fortes J. 2024. Integrating Large Language Models and the iDigBio Portal for Conversational Data Exploration and Retrieval. Biodiversity Information Science and Standards **8**:e142696.

Elliott MJ, Fortes JAB. 2024. Toward Reliable Biodiversity Information Extraction From Large Language Models. Pages 1–10 2024 IEEE 20th International Conference on e-Science (e-Science). IEEE, Osaka, Japan. Available from https://ieeexplore.ieee.org/document/10678666/ (accessed December 25, 2024).

Farrell MJ, Le Guillarme N, Brierley L, Hunter B, Scheepens D, Willoughby A, Yates A, Mideo N. 2024. The changing landscape of text mining: a review of approaches for ecology and evolution. Proceedings B **291**:20240423. The Royal Society.

Fergus P, Chalmers C, Matthews N, Nixon S, Burger A, Hartley O, Sutherland C, Lambin X, Longmore S, Wich S. 2024. Towards Context-Rich Automated Biodiversity Assessments: Deriving AI-Powered Insights from Camera Trap Data. Sensors **24**:8122.

Frazier AE, Song L. 2025. Artificial intelligence in landscape ecology: recent advances, perspectives, and opportunities. Current Landscape Ecology Reports **10**:1–13. Springer.

Fu C, Ren W, Pagani-Núñez E, Han Y, Yang JC, Qiao H, Li Z. 2025. Ai-Assisted Data Extraction Helps Uncover Spatiotemporal Patterns and Socioeconomic Drivers of Wildlife Crime Involving Sea Turtles. Available at SSRN: https://ssrn.com/abstract=5320068. SSRN. Available from https://ssrn.com/abstract=5320068.

Gewin V. 2025. AI carves out a niche in ecology and conservation research. Nature Methods **22**:895–900.
Giebink N, Gupta A, Verìssimo D, Chang CH, Chang T, Brennan A, Dickson B, Bowmer A, Baillie J. 2024. Automating the Analysis of Public Saliency and Attitudes towards Biodiversity from Digital Media. arXiv. Available from https://arxiv.org/abs/2405.01610 (accessed December 25, 2024).
Gougherty AV, Clipp HL. 2024. Testing the reliability of an AI-based large language model to extract ecological information from the scientific literature. npj Biodiversity **3**:13. Nature Publishing Group UK London.
Guo D, Zhu Q, Yang D, Xie Z, Dong K, Zhang W, Chen G, Bi X, Wu Y, Li Y. 2024. DeepSeek-Coder: When the Large Language Model Meets Programming--The Rise of Code Intelligence. arXiv preprint arXiv:2401.14196.
Gurr GM, Liu J, Pogrebna G. 2024. Harnessing artificial intelligence for analysing the impacts of nectar and pollen feeding in conservation biological control. Current Opinion in Insect Science:101176. Elsevier.
Hannah K, Fuller RA, Smith RK, Sutherland WJ, Amano T. 2025. Language barriers in conservation science citation networks. Conservation Biology **n/a**:e70051. John Wiley & Sons, Ltd.
Iyer R, Christie AP, Madhavapeddy A, Reynolds S, Sutherland W, Jaffer S. 2025. Careful design of Large Language Model pipelines enables expert-level retrieval of evidence-based information from syntheses and databases. PLOS One **20**:e0323563.
Jain S et al. 2024. Where It Really Matters: Few-Shot Environmental Conservation Media Monitoring for Low-Resource Languages. arXiv. Available from https://arxiv.org/abs/2402.11818 (accessed December 25, 2024).
Jhonnerie R, Oktorini Y, Volcherina Darlis V, Pebriandi P, Prianto E, Fatmawati R, Nofrizal N, Ramses R, Miswadi M, Rahmatdillah R. 2024. Integration of generative artificial intelligence and Google Earth Engine for mangrove land cover mapping. BIO Web of Conferences **136**:03002.
Ji X, Wu X, Deng R, Yang Y, Wang A, Zhu Y. 2025. Utilizing large language models for identifying future research opportunities in environmental science. Journal of Environmental Management **373**. Available from https://www.scopus.com/inward/record.uri?eid=2-s2.0-85211460324&doi=10.1016%2fj.jenvman.2024.123667&partnerID=40&md5=52ef2acb04b8d0e31ee321e3e7cde3ef.
Kalyuzhnaya A, Mityagin S, Lutsenko E, Getmanov A, Aksenkin Y, Fatkhiev K, Fedorin K, Nikitin NO, Chichkova N, Vorona V. 2025. LLM Agents for Smart City Management: Enhancing Decision Support Through Multi-Agent AI Systems. Smart Cities (2624-6511) **8**.
Keck F, Broadbent H, Altermatt F. 2025, January 27. Extracting massive ecological data on state and interactions of species using large language models. Ecology. Available from http://biorxiv.org/lookup/doi/10.1101/2025.01.24.634685 (accessed January 31, 2025).
Keh SS et al. 2023, April 30. NewsPanda: Media Monitoring for Timely Conservation Action. arXiv. Available from http://arxiv.org/abs/2305.01503 (accessed June 17, 2025).
Koldunov N, Jung T. 2024. Local climate services for all, courtesy of large language models. Communications Earth & Environment **5**:13.
Krishna Moorthy S, Qi M, Rosen A, Malhi Y, Salguero-Gomez R. 2025, February 7. Harnessing Large Language Models for Ecological Literature Reviews: A Practical

Pipeline. Life Sciences. Available from https://ecoevorxiv.org/repository/view/8516/ (accessed March 20, 2025).

Kumar A, Kim H, Nathani JS, Roy N. 2025. Improving the Reliability of LLMs: Combining CoT, RAG, Self-Consistency, and Self-Verification. arXiv. Available from https://arxiv.org/abs/2505.09031 (accessed June 30, 2025).

Lam HYI, Ong XE, Mutwil M. 2024. Large language models in plant biology. Trends in Plant Science. Elsevier.

Larosa F, Hoyas S, Conejero HA, Garcia-Martinez J, Nerini FF, Vinuesa R. 2025. Large language models in climate and sustainability policy: limits and opportunities. arXiv. Available from https://arxiv.org/abs/2502.02191 (accessed June 24, 2025).

Lewers K “Kit.” 2024. Comparative Methods for Building Chatbots: Open Source, Hybrid, and Fully Integrated Large Language Models. Biodiversity Information Science and Standards **8**:e143032.

Liu Y, Yao Y, Ton J-F, Zhang X, Guo R, Cheng H, Klochkov Y, Taufiq MF, Li H. 2023. Trustworthy LLMs: a Survey and Guideline for Evaluating Large Language Models’ Alignment. arXiv. Available from https://arxiv.org/abs/2308.05374 (accessed July 1, 2025).

Mammides C et al. 2016. Increasing geographic diversity in the international conservation literature: A stalled process? Biological Conservation **198**:78–83.

Mammides C, Papadopoulos H. 2024. The role of large language models in interdisciplinary research: Opportunities, challenges and ways forward. Methods in Ecology and Evolution **15**:1774–1776. Wiley Online Library.

Marcos D, Van De Vlasakker R, Athanasiadis IN, Bonnet P, Goëau H, Joly A, Kissling WD, Leblanc C, Van Proosdij ASJ, Panousis KP. 2025. Fully automatic extraction of morphological traits from the web: Utopia or reality? Applications in Plant Sciences:e70005.

Merow C, Serra-Diaz JM, Enquist BJ, Wilson AM. 2023. AI chatbots can boost scientific coding. Nature Ecology & Evolution **7**:960–962.

Miao Z et al. 2024, August 1. New frontiers in AI for biodiversity research and conservation with multimodal language models. Biodiversity. Available from https://ecoevorxiv.org/repository/view/7477/ (accessed December 25, 2024).

Mitchell E, Are EB, Colijn C, Earn DJD. 2025. Using artificial intelligence tools to automate data extraction for living evidence syntheses. PLOS ONE **20**:e0320151.

Mora-Cross M, Calderon-Ramirez S. 2024. Uncertainty Estimation in Large Language Models to Support Biodiversity Conservation. Pages 368–378 Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 6: Industry Track). Association for Computational Linguistics, Mexico City, Mexico. Available from https://aclanthology.org/2024.naacl-industry.31 (accessed December 25, 2024).

Morera A. 2024. Foundation models in shaping the future of ecology. Ecological Informatics **80**:102545.

Murray KI, Hockings KJ, Hodgson D. 2025. The potential for AI to divide conservation. Trends in Ecology & Evolution:S0169534725000588.

Pichai K. 2023. A retrieval-augmented generation based large language model benchmarked on a novel dataset. Journal of Student Research **12**.

Plaat A, van Duijn M, van Stein N, Preuss M, van der Putten P, Batenburg KJ. 2025. Agentic Large Language Models, a survey. arXiv. Available from https://arxiv.org/abs/2503.23037 (accessed June 25, 2025).

Reynolds SA et al. 2025. Conservation changed but not divided. Trends in Ecology & Evolution **40**:424–425.

Reynolds SA, Beery S, Burgess N, Burgman M, Butchart SH, Cooke SJ, Coomes D, Danielsen F, Di Minin E, Durán AP. 2024. The potential for AI to revolutionize conservation: a horizon scan. Trends in Ecology & Evolution. Elsevier.
Richards D, Worden D, Song XP, Lavorel S. 2024. Harnessing generative artificial intelligence to support nature-based solutions. People and Nature **6**:882–893. Wiley Online Library.
Rillig MC, Ågerstrand M, Bi M, Gould KA, Sauerland U. 2023. Risks and Benefits of Large Language Models for the Environment. Environmental Science & Technology **57**:3464–3466.
Robinson D, Miron M, Hagiwara M, Pietquin O. 2024. NatureLM-audio: an Audio-Language Foundation Model for Bioacoustics. arXiv. Available from https://arxiv.org/abs/2411.07186 (accessed July 1, 2025).
Rodas-Trejo J, Ocampo-González P. 2024. Assessment of ChatGPT's potential as an innovative tool in searching for information on wild mammals. Ecological Informatics **83**. Available from https://www.scopus.com/inward/record.uri?eid=2-s2.0-85203417039&doi=10.1016%2fj.ecoinf.2024.102810&partnerID=40&md5=13b89383421fe335c89805f76cece9ff.
Sahoo P, Singh AK, Saha S, Jain V, Mondal S, Chadha A. 2024. A Systematic Survey of Prompt Engineering in Large Language Models: Techniques and Applications. arXiv. Available from https://arxiv.org/abs/2402.07927 (accessed June 30, 2025).
Sandbrook C. 2024. Beyond the Hype: Navigating the Conservation Implications of Artificial Intelligence. Conservation Letters:e13076. Wiley Online Library.
Santangeli A, Mammola S, Nanni V, Lambertucci SA. 2024. Large language models debunk fake and sensational wildlife news. Integrative Conservation **3**:127–133. Wiley Online Library.
Sastry S, Xing X, Dhakal A, Khanal S, Ahmad A, Jacobs N. 2023. LD-SDM: Language-Driven Hierarchical Species Distribution Modeling. arXiv. Available from https://arxiv.org/abs/2312.08334 (accessed December 25, 2024).
Scheepens D, Millard J, Farrell M, Newbold T. 2024. Large language models help facilitate the automated synthesis of information on potential pest controllers. Methods in Ecology and Evolution. Wiley Online Library.
Shelby L, da Silva RVMA. 2024. Retrieval-augmented Generation: Empowering Landscape Architects with Data-driven Design. Journal of Digital Landscape Architecture **2024**:267–276.
Soga M, Gaston KJ. 2025. Extinction of experience among ecologists. Trends in Ecology & Evolution:S0169534724003185.
Sreedhar K, Cai A, Ma J, Nickerson JV, Chilton LB. 2025. Simulating cooperative prosocial behavior with multi-agent llms: Evidence and mechanisms for ai agents to inform policy decisions. Pages 1272–1286.
Tai RH, Bentley LR, Xia X, Sitt JM, Fankhauser SC, Chicas-Mosier AM, Monteith BG. 2024. An Examination of the Use of Large Language Models to Aid Analysis of Textual Data. International Journal of Qualitative Methods **23**:16094069241231168.
Tupper M, Hendy IW, Shipway JR. 2024. Field courses for dummies: To what extent can ChatGPT design a higher education field course? Innovations in Education and Teaching International:1–15.
Urzedo D, Sworna ZT, Hoskins AJ, Robinson CJ. 2024. AI chatbots contribute to global conservation injustices. Humanities and Social Sciences Communications **11**:204.
Vaghefi SA et al. 2023. ChatClimate: Grounding conversational AI in climate science. Communications Earth & Environment **4**:480.

Valdez J, Damasceno G, Oh RRY, Quintero Uribe LC, Barajas Barbosa MP, Amado TF, Schmidt C, Fernandez M, Sharma S. 2024. Strategies for advancing inclusive biodiversity research through equitable practices and collective responsibility. Conservation Biology **38**:e14325.

Wang X, Zhang M, Liu H, Ma X, Liu Y, Chen Y. 2024. ChatBBNJ: a question–answering system for acquiring knowledge on biodiversity beyond national jurisdiction. Frontiers in Marine Science **11**:1368356. Frontiers Media SA.

Wei Y, Hou Y. 2023. Forest Visitors' Multisensory Perception and Restoration Effects: A Study of China's National Forest Parks by Introducing Generative Large Language Model. Forests **14**.

White J, Fu Q, Hays S, Sandborn M, Olea C, Gilbert H, Elnashar A, Spencer-Smith J, Schmidt DC. 2023a, February 21. A Prompt Pattern Catalog to Enhance Prompt Engineering with ChatGPT. arXiv. Available from http://arxiv.org/abs/2302.11382 (accessed January 29, 2025).

White R, Peng T, Sripitak P, Rosenberg Johansen A, Snyder M. 2023b. CliniDigest: A Case Study in Large Language Model Based Large-Scale Summarization of Clinical Trial Descriptions. Pages 396–402 Proceedings of the 2023 ACM Conference on Information Technology for Social Good. Association for Computing Machinery, New York, NY, USA. Available from https://doi.org/10.1145/3582515.3609559.

Zenni RD, Andrew NR. 2023. Artificial Intelligence text generators for overcoming language barriers in ecological research communication. Austral Ecology **48**:1225–1229.

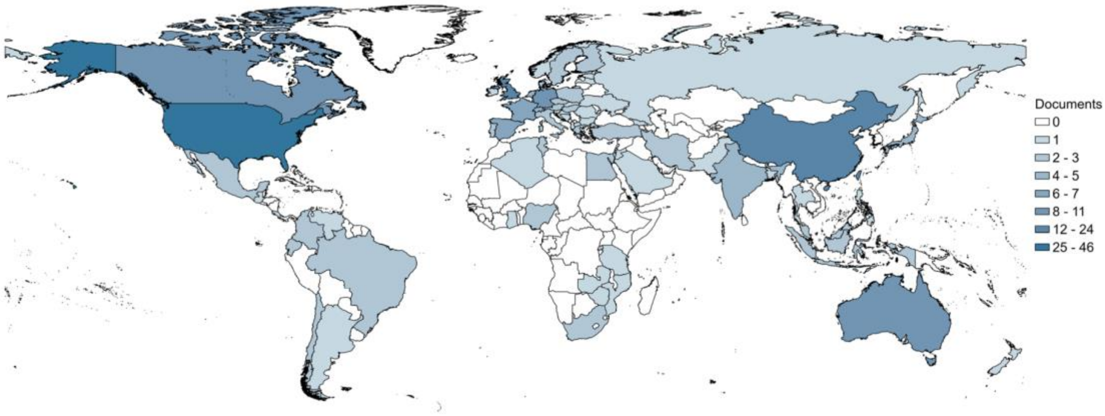
Zhang W, Su Y. 2024. Perception study of urban green spaces in Singapore urban parks: Spatio-temporal evaluation and the relationship with land cover. Urban Forestry & Urban Greening **99**:128455.

Zhao A, Fratila A, Zhang Y, Zhai Z, Liu Z, Shang Y. 2024. Automatic Waterfowl and Habitat Detection using Drone Imagery and Deep Learning. Pages 1–6 2024 IEEE International Conference on Consumer Electronics (ICCE). IEEE, Las Vegas, NV, USA. Available from https://ieeexplore.ieee.org/document/10444338/ (accessed December 25, 2024).

Zheng Z, Zhang J, Vu T-A, Diao S, Tim YHW, Yeung S-K. 2023. MarineGPT: Unlocking Secrets of Ocean to the Public. arXiv. Available from https://arxiv.org/abs/2310.13596 (accessed December 25, 2024).

Ziegler M, Lothian S, O'Neill B, Anderson R, Ota Y. 2024. AI Language Models Could Both Help and Harm Equity in Marine Policymaking: The Case Study of the BBNJ Question-Answering Bot. arXiv. Available from https://arxiv.org/abs/2403.01755 (accessed December 25, 2024).

Figure 1. Countries contributing to research on LLM applications in ecology and conservation science, based on the distribution of author affiliations in the reviewed articles, preprints, and conference proceedings. The USA, China, the UK, Australia, and Germany were the top five contributors (see Fig. S1 for more details).



Box 1. Glossary of key technical terms related to large language models (LLMs), highlighted in bold in the text. The glossary was created using Gemini 2.5 Pro through **role prompting** and a **human-in-the-loop** approach to ensure accuracy and brevity. The following prompt was used: *"You are an AI expert tasked with creating a glossary for a manuscript. Your goal is to make complex AI concepts intuitive for ecologists, field researchers, and conservation policymakers who are experts in their own fields but novices in AI"*.

Artificial Intelligence: The broad field of creating computer systems that can perform tasks that typically require human intelligence, such as natural language understanding.

Agentic LLMs: Large Language Models (LLMs) that can perform multi-step tasks autonomously to achieve a goal.

Context Window: The amount of text (the "context") an LLM can "remember" or consider at one time when generating a response.

Explainable AI: A set of tools and techniques that aim to make AI decisions understandable to humans, opening the "black box."

Fine-tuning: Taking a large, pre-trained "foundation model" and providing additional, specialised training on a smaller, domain-specific dataset.

Foundation Models: Very large, powerful AI models (like GPT-4) that are pre-trained on a massive, general dataset (e.g., a large portion of the internet). They can be adapted to a wide range of tasks through fine-tuning.

Hallucinations: When an LLM generates text that is factually incorrect, nonsensical but presents it confidently as fact.

Human-in-the-loop (HITL): A system design where humans are intentionally included in the AI decision-making process, typically to review, correct, or validate the AI's outputs.

Hyperparameters: The settings and knobs of a machine learning model that are set by the researcher *before* the training process begins.

Large Language Models: Advanced deep learning models, typically utilising transformer architectures, pre-trained on extensive text and code datasets to comprehend and generate human-like language for diverse natural language processing applications.

Long-context LLMs: A new generation of Large Multimodal Models with a very large Context Window, allowing them to process and "remember" entire books, long videos, or hours of audio at once.

Memory-augmented LLMs: Models augmented with external memory that persists across sessions, allowing them to store and retrieve information beyond the limits of their fixed context window.

Prompt Engineering: The technique of crafting effective inputs ("prompts") to get the desired output from an LLM.

Retrieval-Augmented Generation (RAG): A technique that allows an LLM to access and pull information from a specific, trusted database *before* answering a question.

Self-consistency: A technique to improve an AI's reasoning by generating multiple different answers to the same complex question and then choosing the most common or logically consistent answer.

Trustworthy AI: An umbrella term for AI that is lawful, ethical, and technically robust. Key pillars include fairness (unbiased), explainability, reliability, and accountability.

Table 1. Overview of key prompting strategies that can be used to enhance the accuracy, reliability and relevance of the outputs generated by LLMs. Hypothetical examples of how they could be used in the context of ecology and conservation science are provided in the supplementary materials. The descriptions and examples were generated using Gemini 2.5 Pro through **role prompting** and a **human-in-the-loop approach** to ensure accuracy and clarity. The following prompt was used: *"You are an AI expert writing a section for a manuscript on AI applications in ecology and conservation science. The target audience is ecologists, field researchers, and policymakers—experts in their domain but novices in AI. Your task is to define and illustrate eight key prompt engineering strategies. The tone must be formal, clear, and respectful of the reader's expertise. Please provide the following: strategy name, definition, and example use relevant to ecology and conservation science."*

Prompting Strategy	**Description**
Chain-of-Thought (CoT) Prompting	Instructs the model to break down its process into a series of intermediate, sequential steps. By asking the model to "think step-by-step," the user can encourage a more rigorous and transparent reasoning process, which often leads to more accurate final answers.
Citation Prompting	Instructs the model to provide sources, references, or citations for the information it generates. Crucially, this strategy must be paired with independent verification, as models can "hallucinate"—i.e., generate plausible but non-existent citations.
Few-shot Prompting	Provides the model with several examples (the "shots") of the desired input-output pattern before presenting the actual task. This is highly effective for tasks requiring a specific format, structure, or style, as the model learns from the provided exemplars.
Instruction-based prompting	Involves a clear and concise command that states the desired task. It forms the basis of all prompting and is most effective for straightforward, well-defined objectives.
Iterative Prompting	A conversational approach where the initial prompt is a starting point for a dialogue. The user refines, corrects, or expands upon the model's output through a series of follow-up prompts. This dynamic process allows for the progressive shaping of the final output, making it highly effective for nuanced and complex writing tasks.
Prompt Chaining	Involves breaking a large, multifaceted task into a series of smaller, discrete sub-tasks. The output from one prompt is used as the direct input for the next, creating a workflow or "chain." This modular approach improves control and the quality of the final product.
Prompt Ensembling	Involves querying the model with multiple, distinct prompts that all address the same core task. The different prompts might use different phrasing, roles, or constraints. The resulting outputs are then synthesized by the user to produce a more comprehensive, robust, and less biased final answer.

Role Prompting	Instructs the models to assume a specific persona, expertise, or role before posing the query. This primes the model to adopt a particular tone, vocabulary, and conceptual framework, leading to more contextually relevant and nuanced outputs.

Emerging Applications of Large Language Models in Ecology and Conservation Science

Supplementary Materials

Method used to identify and review relevant research:

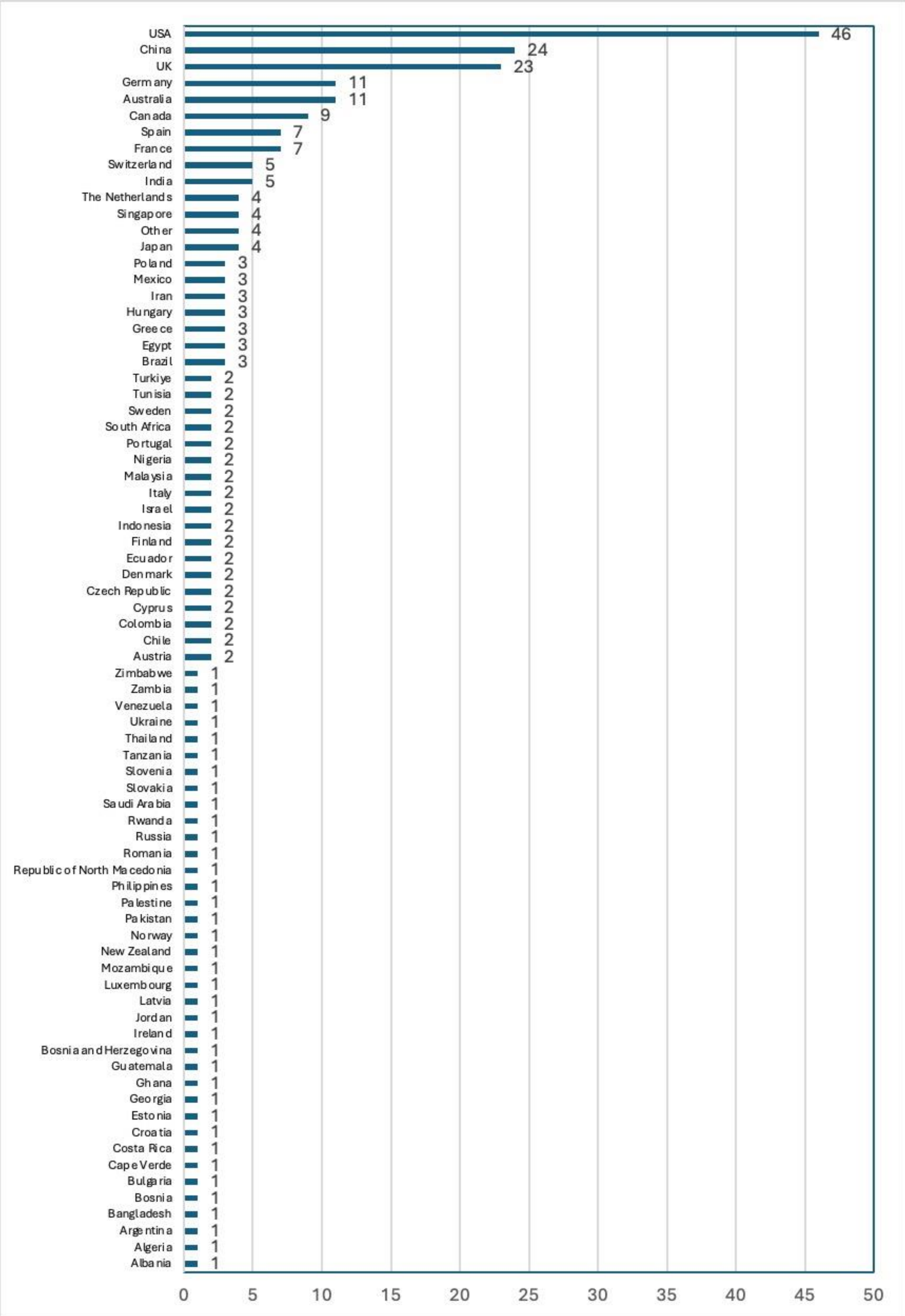
To identify relevant documents, we searched Web of Science, Scopus, and Google Scholar in January 2025 using the following search terms: (“LLMs” OR “Large Language Model*” OR “ChatGPT”) AND (“Ecology” OR “Biodiversity” OR “Conservation”). Given that Google Scholar returns a large number of results, some only tangentially related to the search terms, we limited this search to the first 100 results. This approach is common practice and was justified by confirming that articles beyond the first 100 were not relevant to our purposes.

As noted in the main text, we did not restrict our review to peer-reviewed literature but also included preprints and conference proceedings, provided they were indexed in at least one of the three databases searched. Our search initially returned 338 documents: 133 from Web of Science, 105 from Scopus, and 100 from Google Scholar. After removing duplicates, 267 unique documents remained, which were screened for relevance by the lead author. Articles were excluded if they were unrelated to ecology or conservation, or if they did not specifically refer to large language models (e.g., papers using the acronym “LLM” for other meanings). This screening process resulted in 92 articles. To ensure completeness, the lead author also queried Consensus, ScholarGPT, and ChatGPT for additional references, identifying 8 more relevant articles. The final list of 100 articles was then divided among the first four authors, who each reviewed 25 documents and extracted key information using a shared template developed and agreed upon by all authors.

For each article, we extracted key information such as the authorship, title, and dates of submission and publication. We recorded the article type (e.g., Preprint, Primary Research, Review, Conference Proceedings), the countries in which the authors are based, and the categories of their affiliated institutions (e.g., Academia, NGOs). Each article was classified according to whether it described a specific application of LLMs or offered general commentary or review. We noted the study’s objective, the described use(s) of LLMs (e.g., species distribution modelling, sentiment analysis, etc.), the intended user groups (e.g., researchers, practitioners), and whether the application was demonstrated or merely proposed.

We also documented the ecological and geographical focus of each study, the specific LLM(s) mentioned, and the methods used. Additionally, we summarised the main conclusions and reported benefits of LLM use, along with any technical or non-technical limitations identified by the authors. Where applicable, we recorded proposed solutions, assessed the accessibility of the described workflow, and evaluated their usability based on ease of implementation. We also noted any additional relevant references. Finally, prior to submission, the lead author conducted an updated scan of the literature to identify any new documents published between January and June 2025, and reviewed reference lists for additional relevant but previously unlisted documents. This brought the final number of documents reviewed to 123.

Figure S1. Number of times each country appeared across the 123 reviewed articles, preprints, and conference proceedings, based on author affiliations. If multiple authors from the same country were listed in a single publication, the country was counted only once for that document.



Box S1. A glossary of additional technical terms related to LLMs, which are also mentioned in the text. As with Box 1, the definitions were generated using Gemini 2.5 Pro through **role prompting** and a **human-in-the-loop** approach to ensure accuracy and brevity. The following prompt was used: *"You are an AI expert tasked with creating a glossary for a manuscript. Your goal is to make complex AI concepts intuitive for ecologists, field researchers, and conservation policymakers who are experts in their own fields but novices in AI"*

Artificial Neural Network: A computational model inspired by the structure of the human brain, the core building block of Deep Learning.

Data Sovereignty: The principle that data is subject to the laws and governance structures of the nation or community where it is collected.

Deep Learning: A subfield of machine learning that uses very large, multi-layered Artificial Neural Networks to find extremely complex patterns in vast amounts of data.

Digital Inequalities: The disparities in access to and use of digital technologies, including AI, among different social, economic, or geographic groups.

Few-Shot Learning: Training a model to perform a new task using only a handful ("a few shots") of examples.

Living Evidence Syntheses: A systematic review or evidence map that is continuously updated as new research becomes available, often with the help of AI to automate the search and filtering of new studies.

Low-resource Language: A language for which there is very little digital text data available, making it difficult to train AI models like LLMs.

Maximum Output Length: The limit on how much text an LLM can generate in a single response.

Natural Language Processing (NLP): A field of AI focused on enabling computers to understand, interpret, and generate human language. LLMs are the most advanced form of NLP.

Pseudo-labels: Labels for data that are generated by an AI model itself, not by a human. These are then used to train the model further.

Self-reflection: An advanced AI capability where a model can review, critique, and improve its own work.

Summarisation: The task of generating a shorter version of a text while preserving its essential meaning and key information.

Supervised Learning: The most common type of machine learning, where the model learns from data that has been manually labelled with the correct answer.

Unsupervised Learning: A type of machine learning where the model finds hidden patterns or structures in data that has *not* been labelled.

Zero-shot Learning: A model's ability to perform a task it was never explicitly trained to do, without seeing any examples of that task.

Table S1. Hypothetical examples of how the prompting strategies mentioned in the text could be used in the context of ecology and conservation science. Examples were generated using Gemini 2.5 Pro through **role prompting** and a **human-in-the-loop** approach to ensure accuracy and clarity. The following prompt was used: *"You are an AI expert writing a section for a manuscript on AI applications in ecology and conservation science. The target audience is ecologists, field researchers, and policymakers—experts in their domain but novices in AI. Your task is to define and illustrate eight key prompt engineering strategies. The tone must be formal, clear, and respectful of the reader's expertise. Please provide the following: strategy name, definition, and example use relevant to ecology and conservation science."*

Prompting Strategy	Example
Chain-of-Thought (CoT) Prompting	A conservation planner is presented with a new dataset and wants to generate initial hypotheses. **Prompt:** *"A recent camera trap study in a protected Bornean rainforest reserve shows a 30% decline in sun bear* (*Helarctos malayanus*) *sightings over the past five years, while sightings of wild boar* (*Sus* scrofa) *have increased by 50%. Outline the potential ecological drivers for this inverse trend. Please explain your reasoning for each potential driver step-by-step."*
Citation Prompting	A graduate student is writing a literature review. **Prompt:** *"Provide a summary of the current scientific consensus on the effectiveness of wildlife corridors for connecting fragmented populations of large carnivores in North America. Please include citations to key peer-reviewed articles published after 2018."* **User Action:** The student must then use a scholarly database (e.g., Web of Science, Google Scholar) to locate and verify cited articles to confirm their existence and relevance.
Few-shot Prompting	A field researcher has handwritten notes and wants to digitize and structure them into a consistent format for a database. **Prompt:** *"I will provide unstructured field notes, and you will convert them into a structured JSON format. Here are two examples:* <u>*Example 1 Input:*</u> *'Oct 5, 2023, transect 4, observed a golden eagle soaring overhead, adult.'* <u>*Example 1 Output*</u>*: {"date": "2023-10-05", "location": "transect 4", "species": "Aquila chrysaetos", "count": 1, "notes": "Adult, soaring behaviour observed."}* <u>*Example 2 Input:*</u> *'10/5/23, near the creek bed, saw 2 northern leopard frogs by the water's edge.'*

Prompting Strategy	Example
	<u>Example 2 Output:</u> *{"date": "2023-10-05", "location": "creek bed", "species": "Lithobates pipiens", "count": 2, "notes": "Observed at water's edge."}* <u>New Input**:**</u> *'Spotted a black bear, seemed to be a subadult, crossing game trail 2 on Oct 6, 2023.'* *Your Task: Now, please convert the new input into the structured JSON format."*
Instruction-based prompting	A researcher needs to quickly understand the core arguments of a dense government report before further reading. **Prompt:** *"Summarize the key findings and policy recommendations concerning biodiversity loss from the 2023 IPBES report on invasive alien species. The summary should be approximately 300 words and suitable for a non-technical audience."*
Iterative Prompting	An ecology student is researching a topic for an essay. **Prompt 1**: *"Provide a list of the major benefits of mangrove ecosystems for coastal protection."* **LLM Output**: (Provides a general overview mentioning wave attenuation, shoreline stabilization, and biodiversity conservation) **User Prompt 2**: *"That is a good start. Now, please refine this list by focusing specifically on their role in mitigating storm surge associated with tropical cyclones in the Caribbean. Also, frame your response using the concept of 'ecosystem services'."* **LLM Output**: (Returns a more targeted response, highlighting regulating services such as wave energy dissipation, sediment trapping, and flood risk reduction, with references to Caribbean case studies)
Prompt Chaining	A conservation practitioner is developing a comprehensive species management plan. **Prompt 1:** *"Please generate a list of the top five threats to the giant panda* (*Ailuropoda melanoleuca*)*, citing habitat loss, fragmentation, and climate change."* **Prompt 2 (using output from 1):** *"Using the list of threats above, outline a set of specific, measurable, achievable, relevant, and time-bound (SMART) conservation objectives to address the number one threat: habitat loss and fragmentation."*

Prompting Strategy	Example
	Prompt 3 (using output from 2): *"Based on the SMART objectives you just created, propose three potential stakeholder engagement strategies to ensure local community buy-in for the conservation plan."*
Prompt Ensembling	A policymaker needs to understand the multifaceted impacts of a proposed dam project. **Prompt A***: "From the perspective of a freshwater ecologist, outline the potential negative impacts of constructing a hydroelectric dam on riverine biodiversity and fish migration."* **Prompt B***: "From the perspective of an energy policy analyst, outline the primary benefits of the same hydroelectric dam project in terms of renewable energy generation and grid stability."* **Synthesis***: The policymaker would then integrate both outputs to draft a balanced briefing document that acknowledges the ecological trade-offs and the energy benefits.*
Role Prompting	A forest ecologist wants to draft an outreach document to explain their research to the public. **Prompt:** *"Act as a science communicator specializing in forest ecology. Write a 500-word blog post explaining the concept of 'trophic cascades' using the reintroduction of wolves to Yellowstone National Park as the primary example. The tone should be engaging, accessible, and avoid technical jargon."*