

1        **Harnessing Large Language Models for Ecological Literature**

2                                **Reviews: A Practical Pipeline**

3        Sruthi M. Krishna Moorthy<sup>1,2,3\*</sup>, Man Qi<sup>2</sup>, Alice Rosen<sup>2</sup>, Yadvinder Malhi<sup>3,4</sup>, Roberto

4    Salguero-Gomez<sup>1,2</sup>

5

6        <sup>1,2</sup> Pembroke College, University of Oxford, Oxford, United Kingdom

7        <sup>2</sup> Department of Biology, University of Oxford, Oxford, United Kingdom

8        <sup>3</sup> Environmental Change Institute, School of Geography and the Environment, University  
9        of Oxford, Oxford, OX1 3QY, United Kingdom

10        <sup>4</sup>Leverhulme Centre for Nature Recovery, University of Oxford, Oxford, OX1 3QY, United  
11        Kingdom

12

13        \*Corresponding author: [sruthi.krishnamoorthyparvathi@pmb.ox.ac.uk](mailto:sruthi.krishnamoorthyparvathi@pmb.ox.ac.uk)

14

## 15        **1. Abstract**

16        Hundreds of thousands of peer-reviewed articles and grey literature reports are published every  
17        year in ecology and conservation biology. This ever-growing body of knowledge presents new  
18        challenges. Indeed, it is becoming increasingly challenging for researchers to stay current on new  
19        information and to identify knowledge gaps. Here, we argue that Large Language Models (LLMs)  
20        such as OpenAI's GPT-4o mini offer a powerful yet accessible solution to help overcome this  
21        challenge, as LLMs require only effective prompt engineering rather than specialized AI expertise.  
22        We present a streamlined LLM-driven pipeline for filtering and extracting information from large  
23        volumes of literature, illustrating its potential through two case studies. Our findings show that, by  
24        combining LLMs with short, iterative prompting workflows and targeted manual validation checks,  
25        researchers can rapidly obtain structured outputs—such as study locations, biome types, or  
26        quantitative measures—while minimizing model hallucinations and misinterpretations. We  
27        emphasize that domain experts remain integral for shaping prompts, verifying results, and  
28        ensuring the extracted information aligns with real-world research and conservation needs.  
29        Overall, this pipeline underscores the synergy between human expertise and LLM capabilities,  
30        promising more efficient literature reviews for a broad range of ecological and conservation  
31        applications.

32        **Key-words:** large language models, literature review, prompt engineering, information extraction,  
33        evidence synthesis, research gaps and trends

## 34 2. Introduction

35 Ecology and conservation science are data-driven disciplines that rely on synthesizing research  
36 across diverse fields to shape management and policy decisions. Researchers in both disciplines  
37 are very active, producing high volumes of peer-reviewed articles, technical reports, and grey  
38 literature that continue to grow exponentially each year. In the last decade there has been a two-  
39 fold increase in the number of peer-reviewed publications per year just in ecology (from Web of  
40 Science). These efforts can allow us to better address challenges caused by habitat loss,  
41 ecosystem degradation, climate change, invasive species and other threats. But there is a dark  
42 side to such high productivity: keeping up with it all. Traditional literature reviews often struggle to  
43 keep pace with the proliferation of studies (Tsertsvadze et al. 2015; Qureshi et al. 2023). As a  
44 result, an inadvertent consequence of such exponential increase is that keeping track of the  
45 scientific advancements and gaps in ecology and conservation science is becoming more  
46 challenging.

47 Large Language Models (LLMs) such as openAI's GPT models, Meta's llama, *etc.* are  
48 increasingly recognized for their potential to revolutionize the literature review process in ecology  
49 (Gougherty and Clipp 2024; Castro et al. 2024), conservation (Reynolds et al. 2024) and other  
50 research fields. ChatGPT and its use in academia has been at the centre of an active debate  
51 regarding the ethics of its usage since its launch, in 2022. This debate has predominantly focused  
52 on its use in academic writing (Johnson et al. 2024), coding (Cooper et al. 2024), education  
53 (Extance 2023), and its energy and water consumption (Ren et al. 2024). In the meantime,  
54 context-based information retrieval has long been a core pursuit in AI (Coutaz et al. 2005). LLMs  
55 now streamline this process by inferring information that may not be explicitly stated in the text.  
56 Although the medical and clinical sciences have made strides in using LLMs or specialized  
57 Natural Language Processing (NLP) pipelines for systematic reviews and rapid meta-analyses

58 (Huang et al. 2024; Sacoransky, Kwan, and Soboleski 2024; Idnay et al. 2021), ecology and  
59 conservation research has only begun to explore these possibilities. Recent publications have  
60 highlighted tools such as ChatGPT (primarily for summarizing) and AI-enabled research assistant  
61 software like Elicit (Whitfield and Hofmann 2023) (for semi-automated data extraction) that could  
62 enable efficient synthesis of existing literature in conservation contexts (Berger-Tal et al. 2024;  
63 Reynolds et al. 2024). Furthermore, while tools like Elicit harness underlying LLM technologies  
64 (including OpenAI's GPT models), there remains a notable gap in practical, consensus-driven  
65 guidelines for applying LLMs to literature reviews in ecology and conservation. Some recent  
66 ecological studies already report that simple data extraction tasks, such as identifying a study's  
67 location, can exceed 90% accuracy using LLM-based approaches (Gougherty and Clipp 2024;  
68 Castro et al. 2024). However, these efforts have largely focused on relatively straightforward  
69 extractions. Here, we go beyond location-level extractions to demonstrate how LLMs can also  
70 extract and interpret more complex, context-based information — an area where robust prompting  
71 guidelines and domain-expert oversight become essential. Here, we offer concrete strategies to  
72 help researchers effectively incorporate LLMs into their review workflows.

73 We introduce a streamlined, LLM-enabled pipeline for efficient literature reviews in ecology and  
74 conservation science. Importantly, our pipeline is complete with practical prompting guidelines  
75 and reusable code that non-AI experts can readily adopt. We demonstrate the pipeline using  
76 OpenAI's GPT-4o mini model, chosen for its popularity and widespread adoption, using two  
77 literature review case studies with different objectives: one on uncovering patterns in autonomous  
78 ecosystem monitoring (e.g., drones, ground or underwater robots) and another on evidence  
79 synthesis for protected-area effectiveness. By setting the model's temperature parameter to zero,  
80 a feature found in all major LLMs, we ensure more deterministic outputs, minimizing variability  
81 across repeated runs (Peeperkorn et al. 2024). We focus in particular on automating two key time-  
82 consuming tasks in any systematic literature review: (1) Filtering relevant publications after they

83 have been gathered from conventional or AI-based literature-search tools (e.g., Elicit, Web of  
84 Science), and (2) Extracting both qualitative and quantitative data from location or taxa to specific  
85 effect sizes used in meta-analyses. Through these two detailed case studies, we illustrate how  
86 researchers can apply our pipeline and guidelines to enhance their review workflows without  
87 needing specialized AI expertise. We further discuss common pitfalls, highlighting that careful  
88 design of prompts and domain-expert oversight can overcome many of the known limitations of  
89 LLMs. While specific coding implementations and libraries may change depending on the choice  
90 of the LLM, the pipeline is designed to be adaptable and applicable to any LLM beyond the  
91 demonstrated OpenAI's GPT-4o mini, such as Meta's Llama and Google's gemini, etc. In addition,  
92 we provide all the final prompts and the code used to run the LLM in the supplementary materials  
93 for transparency and reproducibility.

### 94 **3. Literature review pipeline and LLM integration**

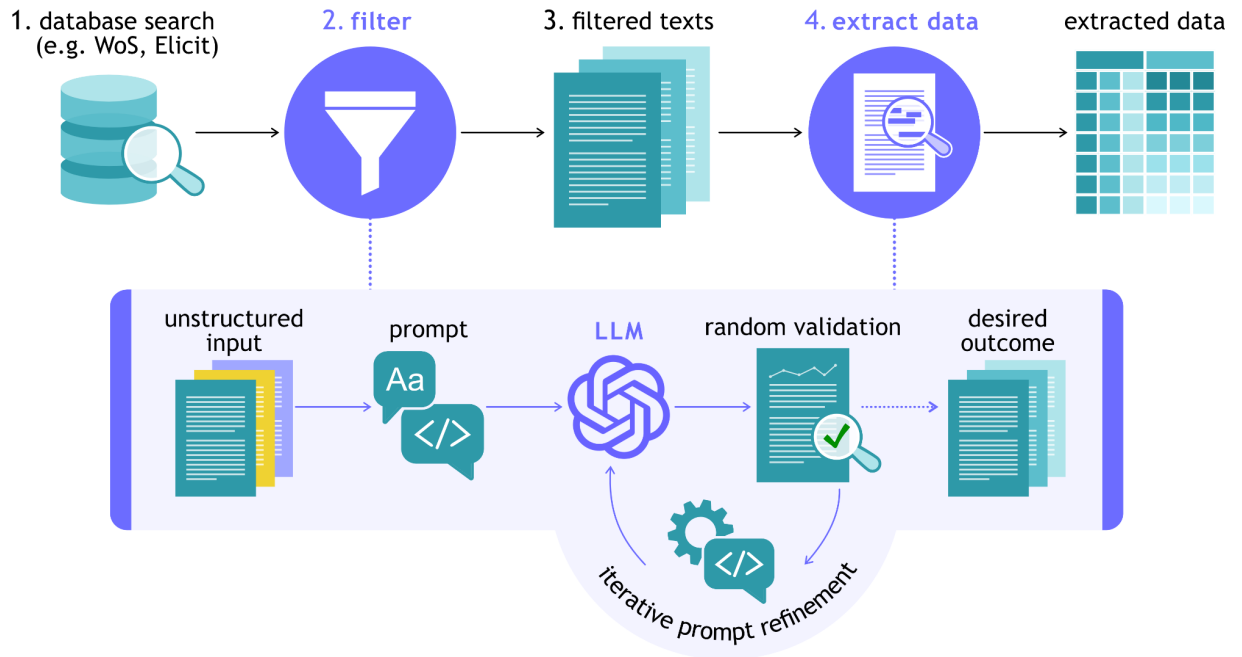
95 A typical literature review pipeline in ecology or conservation often follows these major steps: (1)  
96 Defining the objective or research question. For illustration, here we explore two case studies:  
97 (Case 1) "*What is the state-of-the-art of drones and autonomous robots in monitoring*  
98 *biodiversity?*", and (Case 2) "*Are terrestrial protected areas effective in increasing species*  
99 *richness or abundance compared to unprotected areas?*" (2) Searching/collecting publications –  
100 using databases like Web of Science, Scopus, or Elicit (Whitfield and Hofmann 2023) (3)  
101 Screening and filtering – deciding which studies are relevant for your review's objectives (4) Data  
102 extraction – gathering key information from each selected study (e.g., location, ecosystem type,  
103 effect sizes, outcomes). The extracted data is then summarised for qualitative literature reviews  
104 or quantitatively analysed when the goal is a meta-analysis, highlighting implications for future  
105 research directions, policy, or management.

106 While LLMs can play a major role in searching and collecting the relevant publications (Berger-  
107 Tal et al. 2024), here we demonstrate how LLMs can also be seamlessly integrated into the  
108 traditional literature review pipeline for filtering (step 3) and extracting data from the literature (step  
109 4) (Figure 1).

- 110 • Filtering: After obtaining a corpus of papers, LLMs can help quick determination of whether  
111 each paper meets the inclusion criteria for the literature survey.
- 112 • Data extraction: LLMs can enable the automatic extraction of simple metadata (e.g.,  
113 region of study) or more complex information (e.g., effect sizes, sampling design)

114 While LLMs are not a substitute for domain expertise, we argue that LLMs have the potential to  
115 accelerate routine tasks with minimal expert supervision. Domain experts in ecology and  
116 conservation biology remain in charge of crafting the prompts and validating the model responses,  
117 especially when extracting nuanced or specialized data (Reynolds et al. 2024), as we will highlight  
118 below in our case studies. As such, the AI landscape has evolved in a direction where we now do  
119 not need AI experts to use AI; rather, we need domain experts to instruct these systems  
120 effectively. We enlist the following steps to ensure quality control and mitigate potential biases or  
121 inaccuracies when integrating LLMs into the literature review pipeline:

- 122 • Random checks: Randomly select a portion of the LLM-generated outputs (e.g., 10%) for  
123 manual verification and calculate the accuracy of your LLM responses on this subset.
- 124 • Iterative prompt refinement: Since LLM's responses rely heavily on the prompts provided,  
125 if your manual validation checks reveal recurring issues or systematic errors such as  
126 misunderstood terms, overlooked taxa, *etc.*, revise your prompt accordingly. Repeat the  
127 random checks and prompt refinement until you arrive at your desired outcome. In our  
128 case, we stopped refining the prompt when subsequent prompt refinements no longer  
129 resulted in significant improvements in accuracy.



130

131

132

133

134

135

136

137

**Figure 1: Overview of our Large Language Model (LLM)-powered literature review pipeline to synthesize knowledge and seek gaps of knowledge in ecology and conservation biology.** After collecting publications via conventional or AI-based databases (e.g., Web of Science, Elicit), relevant articles are filtered (with or without LLM) before the LLM extracts key information (e.g., study location, effect sizes). An iterative feedback loop — with random validation checks and prompt refinements — ensures accuracy and results in a structured final dataset for further analyses.

138

#### **4. Case Study 1: Uncovering research trends in autonomous biodiversity monitoring**

139

140

141

142

143

As new technologies like drones and other uncrewed robots proliferate, researchers seek to understand how these tools are used for monitoring ecosystems (Qi et al. 2025). Rather than manually reading hundreds of abstracts, an LLM can rapidly classify and summarize the focus of each study, highlighting the gaps and trends in the field of autonomous ecosystem monitoring.

144 Here, our goal was to explore how LLMs can extract categorical data from studies using robots  
145 to monitor ecosystems. To that end, we used traditional keyword-based searches to gather 1154  
146 abstracts on drone or robot usage in ecosystem monitoring. The detailed search string and  
147 publication selection criterion is explained in Qi et. al., (Qi et al. 2025). We then randomly selected  
148 225 abstracts to validate our LLM-based information retrieval pipeline. We specifically tested  
149 OpenAI's GPT-4o mini model on the abstracts of 225 publications describing novel applications  
150 of autonomous monitoring of ecosystems. Given that all 225 abstracts mention drone or robot-  
151 based monitoring, we did not apply LLM-based filtering for this objective. We then designed  
152 prompts asking GPT-4o mini to extract (1) the country of the study (2) the biome in which the  
153 study was conducted (*e.g.*, forests, savannahs), and (3) the taxonomic kingdom monitored by the  
154 autonomous system (*e.g.*, animal, plant). We sorted the target information into three levels of  
155 extraction difficulty—easy, moderate, and complex:

- 156 1. Easy – Explicitly stated details, such as the country or region where drones/robots were  
157 deployed.
- 158 2. Moderate – Implicit but easily inferred details, such as the biome where drones/robots  
159 were deployed.
- 160 3. Complex – Context-based inferences requiring careful prompts, such as distinguishing the  
161 organism of primary interest from the organism actually monitored by drones/robots.

162 We applied a straightforward prompt design on all 225 abstracts, then evaluated the performance  
163 of the LLM by manually validating the LLM's outputs for a set of 40 randomly selected abstracts.  
164 We identified a few systematic issues via this validation step, which we then improved by  
165 iteratively refining the prompts.

- 166 (1) Inconsistent model outputs complicate analysis: For instance, without explicitly specifying  
167 the expected output format, the model may generate different terms for the same category



168 across abstracts. For countries, explicitly asking to respond with the standardized ISO3  
169 codes in prompts (e.g., "USA" for the United States of America) can prevent  
170 inconsistencies, such as generating "United States" as response for one abstract and  
171 "United States of America" for another.

172 (2) Lack of granularity causes ambiguity: Another issue related to inconsistent model outputs  
173 is the lack of specification of predefined labels for certain categorical data. For instance, if  
174 the abstracts involve monitoring tropical forests, a lack of clear guidance may result in the  
175 model outputting varying labels such as "forests" or "tropical forests," even though they  
176 refer to the same biome. This ambiguity underscores the importance of domain experts or  
177 researchers defining the level of granularity required for their specific use case. By  
178 providing a predefined list of acceptable labels (Figure 2b), and iteratively refining them if  
179 needed, users can ensure uniformity and reduce ambiguity in the outputs.

180 (3) Vague prompts lead to wrong inferences: For instance, the initial prompt designed for  
181 identifying the taxonomic kingdom directly monitored by drones/robots was vague, such  
182 that it inadvertently conflated the primary study organism with the organism monitored by  
183 drones/robots. An example of this accidental conflation in our set of 225 publications  
184 occurred for a study (Proudfoot et al. 2023) where the authors examined how fish diversity  
185 is influenced by eelgrass meadows. The authors used drones to monitor eelgrass meadow  
186 structure, not the fish, although the latter was often identified as the monitored taxonomic  
187 group by the LLM. We refined the prompt by adding disclaimers and explicitly instructing  
188 the model to avoid conflating the different organisms (see Fig 2b). Additionally, asking the  
189 model to justify its responses helped us identify the reasoning behind specific outputs,  
190 enabling targeted refinements to the prompt and addressing systematic issues effectively.

191 Once iterative refinements no longer resulted in additional improvements in the 40 randomly  
192 selected subset of abstracts, we evaluated the accuracy of the final prompt on all 225 abstracts.

193 To extract country of study location from the abstracts, OpenAI's 4o-mini model performed with a  
194 high accuracy of 97%, which is in agreement with the performance of LLMs as reported by other  
195 recent studies (Castro et al. 2024; Gougherty and Clipp 2024). Traditional Natural Language  
196 Processing (NLP) models are most often used for extracting named entities like country or city  
197 names from unstructured text (Mohit 2014). When compared to LLMs, in our case study,  
198 traditional NLP models like Python's spaCy achieved an accuracy of 63% for the same task.  
199 LLMs, which are recent deep learning-based advanced models in the field of NLP (Farrell et al.  
200 2024), have a significant advantage over traditional NLP models like Python's spaCy for extracting  
201 the names of the study locations *i.e.* countries. While spaCy relies on country or city names being  
202 explicitly mentioned in the original source, so they can be recognized by named entity recognition  
203 (Vasiliev 2020), LLMs can infer countries from indirect location references (Lin et al. 2024), such  
204 as nature reserves or regional landmarks. For example, in our case study, a paper with the  
205 statement "The occurrence of a peculiar phenomenon called soil balls was observed in Dingbian  
206 County, northern Loess Plateau,(Yang et al. 2023)" LLM inferred "China," whereas spaCy did  
207 not. Similarly, although the biome was not explicitly stated in a study about blue whale foraging  
208 (Torres et al. 2020), GPT-4o mini successfully inferred the study corresponds to "marine", based  
209 on contextual clues present in the abstract. In fact, the accuracy achieved for extracting biomes  
210 in our case study by GPT-4o mini was 80%. In the case of complex-context-based inference, we  
211 evaluated GPT-4o mini's ability to differentiate between the primary study organism and the  
212 organism monitored by the drones/robots. Through creative prompting (Fig 2), our pipeline  
213 achieved an overall accuracy of 81% with the 4o-mini model.

214 The performance of GPT against our expert-validated data in this case study demonstrates how  
215 LLMs can successfully categorize large sets of abstracts effectively. In turn, this approach could  
216 offer ecological researchers a broad overview of trends, biases and potential knowledge gaps  
217 (*e.g.*, which geographic regions are most and least represented respectively).

## 218 **5. Case Study 2: Evidence synthesis for terrestrial protected areas**

219 Terrestrial protected areas are among the most widely used and longstanding conservation  
220 strategies worldwide, designed to safeguard biodiversity, preserve critical habitats, and maintain  
221 ecosystem functions in the face of habitat loss and other anthropogenic threats (McNeely 1994).  
222 As a result, understanding the effectiveness of protected areas in enhancing species richness or  
223 abundance compared to unprotected sites is of crucial importance as it directly informs policy  
224 decisions, resource allocation, and future conservation planning on a global scale. In this case  
225 study, we demonstrate the potential of LLMs to aid various stages of evidence synthesis, including  
226 filtering relevant studies and extracting complex quantitative data for meta-analyses. Building on  
227 the non-quantitative extraction capabilities showcased in Case Study 1, we next examine the  
228 synthetic evidence for how effective terrestrial protected areas are at enhancing species richness  
229 or abundance.

230 Evidence syntheses and meta-analyses require robust quantitative comparisons (*e.g.*, effect and  
231 sample sizes) across multiple studies. While AI-based tools like Elicit or conventional search  
232 engines like Web of Science can retrieve numerous publications, not all publications meet the  
233 criteria for quantitative synthesis. We used the top ten most relevant open-access articles (as  
234 ranked by Elicit) discussing terrestrial protected areas and biodiversity. We then crafted a prompt  
235 asking GPT-4o mini to determine whether each paper quantitatively compared protected and non-  
236 protected areas measuring species richness or abundance. The LLM provided “yes/no” answers  
237 with justification in a structured (JSON) format.

238 From the subset of relevant articles, we tasked GPT-4o mini with extracting:

- 239 1. The effect size metric (*e.g.*, Cohen’s *d*, log response ratio).
- 240 2. The numeric value and confidence interval (or standard error).
- 241 3. What the effect size specifically represents (*e.g.*, species abundance, species richness).

242 Eight of the ten top ranked publications by Elicit met the criteria for the quantitative meta-analysis.  
243 Simple presence/absence checks were highly reliable with all relevant publications correctly  
244 identified. However, extracting numeric effect sizes proved challenging due to the diverse ways  
245 they are reported.

246 Extracting effect sizes from full-text studies posed two consistent issues: (1) misidentifying non-  
247 standard metrics (e.g., model interactions) as effect sizes, and (2) overlooking one or more  
248 relevant effect sizes when multiple comparisons were reported in the same document. For  
249 instance, in one study, the LLM incorrectly flagged an interaction term as a standardized effect  
250 size (Lehikoinen et al. 2019), while in another, it missed several effect-size measures altogether  
251 (Cazalis et al. 2020). We addressed these problems by iteratively refining our prompts and  
252 requesting standardized outputs by explicitly specifying the type of acceptable effect size metrics  
253 such as Hedge's  $g$  or log response ratios and by providing few-shot examples, e.g., "Hedge's  $g =$   
254  $0.65$ , 95% CI [ $0.45$ ,  $0.85$ ]" (see Supplementary Figure S1 for more details on few-shot prompting  
255 style) . This approach significantly reduced errors and inconsistencies, guiding the LLM to ignore  
256 regression interactions and search carefully for real effect sizes within the same text. However,  
257 domain experts play a crucial role in defining the criteria for extracting effect sizes from studies,  
258 as they are best positioned to determine what is relevant for their analysis. For example, experts  
259 need to decide whether percentage differences in effectiveness are acceptable as effect sizes or  
260 if specific standardized metrics are required (Kallogjeri and Piccirillo 2023). In publications where  
261 numerous effect-size metrics are scattered throughout long or complex result sections, occasional  
262 omissions may persist. This reinforces the need for human validation and domain-expert oversight  
263 in large-scale evidence syntheses (Gougherty and Clipp 2024). One suitable way to minimize  
264 these omissions is to split the large results sections of publications into smaller segments of fixed  
265 size (e.g. 500 words) (Arefeen, Debnath, and Chakradhar 2024). Importantly, we found that  
266 asking the LLM to justify its responses helps identify the reasoning behind specific outputs. Once

267 effect size data were validated, we could quickly compile them into a spreadsheet for meta-  
268 analytic calculations. This points to a scalable approach for large-scale evidence syntheses in  
269 conservation.

## 270 **6. Prompting guidelines and best practices**

271 The perfect prompt does not exist. By embracing this reality, domain experts (e.g., ecologists,  
272 conservation biologists, etc.) aiming to use LLMs to extract information from high volumes of  
273 sources can view effective prompting as a continuous, iterative process. Based on our case  
274 studies above, here we summarize a series of prompting guidelines that have consistently proven  
275 useful in achieving highly accurate extractions of data from published literature. These principles  
276 focus on structuring the model instructions, specifying acceptable outputs, and demanding  
277 justification from the model, helping you to build trust in the LLM-assisted workflow.

- 278 1. System prompts: Some LLM interfaces let the user set a “system” prompt that defines the  
279 model’s role and constraints. An example system prompt used in our case studies: “You  
280 are an ecological research assistant specifically focused on extracting information from  
281 scientific text. Do not fabricate data.” This system prompt helps maintain consistency and  
282 reduce confusion as you iterate on user prompts.
- 283 2. Delimit the prompt: Create clear and structured prompts by using delimiters, breaking  
284 down complex requests into manageable steps. For instance, we provided step-by-step  
285 instructions in our prompts when tasking the model with extracting the data, as can be  
286 seen in Figure 2b.
- 287 3. Be explicit: Avoid ambiguity by asking for specific output formats (Entity 1 of Step 2 in  
288 Figure 2b) or predefined labels for categorical data where possible (Entities 2 and 3 of  
289 Step 2 in Figure 2b). Provide explicit instructions. Ask the model to explicitly check for

290 relevant information before generating a response from a given source (for instance,  
291 scientific content in our case study, Step 3 in Figure 2b).

292 4. Diversify your prompt: Utilize various prompting styles to explore different outputs,  
293 including:

294 ○ Zero-shot prompting: Pose a question or task without providing examples, allowing  
295 the model to generate a response based solely on its training (Figure 2a).

296 ○ Few-shot prompting: Provide a few examples of desired inputs and outputs to  
297 guide the model toward generating relevant responses (Supplementary Figure  
298 S1).

299 ○ Chain-of-thought prompting: Encourage the model to think through a problem step-  
300 by-step, enhancing reasoning and producing more thorough answers (Entity 3 of  
301 Step 2 in Figure 2b).

302 5. Demand justification: Request justification for the model's responses, including evidence  
303 to support its claims ("Comment" key in Step 4).

304 6. Request structured output: Whenever possible for easier post-processing of the model's  
305 responses, ask for structured output in JSON format (Step 4). OpenAI's recent model  
306 releases (gpt-4o model from 2024-08-06) ensures the output follows JSON schema if  
307 requested. JSON (JavaScript Object Notation) is a lightweight data interchange format  
308 that is commonly used to represent structured data in a key-value pair format.

**a. System prompt**

You are a helpful assistant.

**User prompt**

Extract the following information from the abstract on the application of drones or robots or UAVs to study any organisms and/or the environment. Identify and return the countries where the study was conducted. Identify the biomes of the study. Identify the study organisms monitored by drones in the abstract. Return the kingdoms the organisms belong to.

The abstract text is: "Eelgrass..."



Based on the provided abstract, here is the extracted information:

- **Countries where the study was conducted:** Canada (specifically British Columbia)
- **Biomes of the study:** Temperate coastal areas
- **Study organisms monitored by drones:** Fish
- **Kingdoms the organisms belong to:** Animalia

If you need further assistance or additional information, feel free to ask!

Country names not consistent or standardized

Biomes are not from clearly defined categories

Incorrect response due to missing nuance

## b. System prompt

You are a research assistant specialized in ecology. You carefully read the text you are given, extract the requested information from the text and you only respond in JSON format as instructed. You never make up data that is not present in the text.

Provide system prompt defining the model's constraints. Helps with hallucinations.

## User prompt

Extract information from the abstract on the application of drones or robots or UAVs to study any organisms and/or the environment. Your task is to perform the following actions:

Step-by-step instructions

1. Read the content of the abstract in its entirety.

2. Specifically, extract the structured information on the following entities from the abstract:

Ask for standardized names

Entity 1 - Country: Identify and return the standardized ISO3 names of the countries where the study was conducted. Do not include countries mentioned in the text that are not directly related to the study location.

Entity 2 - Biomes: Identify the biomes of the study. Check if the biomes belong to the following categories: Forests, Shrublands, Grasslands, Savannas, Mangroves, Wetlands, Coastal, Coral reef, Marine. Wetlands refers to all terrestrial water-logged ecosystems including peats, swamps, lakes, rivers and streams. Arid deserts should be classified as "NA" and semi-arid desert shrublands should be classified as "Shrublands". Coastal refers to the coastal ecosystems like delta, intertidal zones, coastal aquaculture, sea grass communities, estuarine, tidal marshes, lagoons, rocky shores. Return "NA" for otherwise.

Give background information and clarifying details

Entity 3 - Kingdom: Identify the study organisms monitored by drones in the abstract. Return the kingdoms the organisms belong to. The values for this key should be one of those within the square brackets: ["Animalia", "Plantae", "Fungi", "Protista", "Archaea", "Bacteria", "NA"]. For instance, consider a study with the main objective of understanding how the bat population is affected by forest structure, measured by drones. Since the drone is deployed to measure the forest structure directly and not the bats, the correct response for this question is Plantae and not Animalia.

State the level of granularity needed as a response

Important: Carefully read and understand whether the drones/robots/UAVs are directly studying/monitoring/mapping an organism or only the abiotic environments.

Be explicit and provide example responses for scenarios the model might encounter

3. Do not interpret what the robots could technically measure or monitor. First find the relevant information in the abstract and then generate the response for a specific entity based on the information. If the information is not present in the text, use "NA" as the value.

4. Output your response as a JSON object as in the following format.

Request structured output, e.g. in JSON format

- Country: List of countries
- Biomes: ["Forests", "Shrublands", "Grasslands", "Savannas", "Mangroves", "Wetlands", "Coastal", "Coral reef", "Marine", "NA"]
- Kingdom: ["Animalia", "Plantae", "Fungi", "Protista", "Archaea", "Bacteria", "NA"]
- Comment: Justify your responses for all the entities.

Request justification

If more than one value is present or identified for an entity, return all the values in a comma-separated form. Do not return values that are not listed within square brackets.

The abstract text is delimited with triple backticks. ``Eelgrass...``

Use delimiters for clarity



```
{
  "Country": ["CAN"],
  "Biomes": ["Coastal"],
  "Kingdom": ["Plantae"],
  "Comment": "The study was conducted in British Columbia, Canada, which is why 'CAN' is listed as the country. The biomes identified include 'Coastal' because the study focuses on eelgrass meadows, which are coastal ecosystems. The kingdom is 'Plantae' since the primary organism studied is eelgrass (Zostera marina), a plant species."
}
```



311 **Figure 2. Illustration of incorporating our proposed guidelines in crafting effective prompts**  
312 **for extracting data from scientific studies** (a) Initial prompt for extracting country, biome, and  
313 kingdom information from an abstract. Although straightforward, the prompt lacks explicit  
314 instructions on acceptable outputs (e.g., standardized country codes) and precision in defining  
315 possible biome categories, leading to inconsistent or ambiguous results. (b) Refined prompt  
316 incorporating the guidelines: a clearly defined system role, step-by-step instructions, standardized  
317 outputs (ISO3 for countries, predefined biome list), structured JSON formatting, and justification  
318 requests. This structured approach substantially improves the accuracy and consistency of the  
319 extracted data.

320 To highlight the effectiveness of iterative prompt refinement, we showcase how various prompting  
321 guidelines—such as the use of delimiters, explicit instructions, step-by-step breakdowns,  
322 justification requests, and structured JSON outputs—achieved the desired outcome compared to  
323 an initial novice prompt when applied to a complex abstract (Fig. 2). Among the prompting styles,  
324 chain-of-thought prompting, where we explicitly guide the model through our thought process,  
325 proved most effective for extracting complex context-based information (e.g., the kingdom of  
326 organisms) from this example, as illustrated in Figure 2b. This approach demonstrates how  
327 guiding the model through reasoning steps enhances its ability to process complex requests.  
328 Nevertheless, the number of iterations required for prompt refinement before reaching the desired  
329 outcome is inherently difficult to quantify, as it depends on the complexity of the problem and the  
330 user's experience. However, by following the prompting guidelines outlined in this study,  
331 ecologists can begin with a stronger foundation, minimizing the number of iterations needed to  
332 reach their desired results.

333

## 334        **7. Discussion**

335    Large Language Models (LLMs) such as GPT-4o mini offer an accessible way to perform literature  
336    reviews in ecology and conservation with minimal infrastructure, requiring only effective prompt  
337    design. Moreover, these methods are readily adaptable to a wide range of other scientific fields  
338    wherever large volumes of literature need efficient synthesis. The capacity of LLMs to parse  
339    hundreds of thousands of publications has important ramifications for identifying geographical or  
340    taxonomic mismatches in ecological and conservation research. For example, Lin et al. (Lin et al.  
341    2024) recently analyzed over 100,000 published papers using OpenAI's GPT-3.5 Turbo and found  
342    a significant mismatch between the geographical distribution of wildfire research and actual  
343    satellite-derived wildfire locations. Such disparities between conservation research efforts and  
344    real-world conservation priorities are often difficult to uncover, especially when individual  
345    researchers focus on a limited set of species or study areas (Gaulke et al. 2019). In contrast,  
346    using LLMs to extract categorical data from vast swaths of literature can help pinpoint where  
347    additional research or on-the-ground interventions are most urgently needed. Nevertheless,  
348    extracting more complex quantitative information, such as multiple effect sizes buried in long  
349    results sections, can remain challenging and may require further prompt refinement, splitting the  
350    sections into smaller segments and domain-expert oversight.

351    LLMs prioritize the expertise of researchers with local domain knowledge over the need for AI  
352    specialists. By enabling researchers anywhere—including in regions with limited computing  
353    resources—to run powerful models on their local computers, these generalist LLMs help lessen  
354    global scientific inequity, wherein data processing traditionally occurs in well-funded Global North  
355    institutions (Reynolds et al. 2024). LLMs also support multilingual workflows, allowing local-  
356    language publications and indigenous knowledge to be integrated more seamlessly than ever  
357    before. Although GPT-4o mini is a lightweight (lesser model complexity), cheaper, and faster

358 variant of advanced LLMs, its performance and accuracy continue to improve as these models  
359 evolve. Specifically, processing 100,000 abstracts (approximately 500 words each) with the  
360 4omini model only requires \$10, thus significantly enhancing cost efficiency for researchers. In  
361 contrast, AI-enabled software Elicit allows data extraction from only 1,200 publications per year  
362 and further restricts users to extracting just 20 different data types, even with its pro model costing  
363 \$500 annually (<https://support.elicit.com/en/articles/471617>). These constraints may limit its utility  
364 for large-scale, cost-effective literature reviews in any domain.

365 Despite their promise, larger LLMs come with significant computational and environmental costs  
366 (Alzoubi and Mishra 2024). These costs may underscore the value of developing smaller, domain-  
367 specific models in tandem with AI specialists (Reynolds et al. 2024). Meanwhile, hallucinations  
368 remain a risk in generalist models, though tasks centred on extracting information from given  
369 sources are inherently less prone to invented data (Mittelstadt, Wachter, and Russell 2023). Our  
370 case studies illustrate how prompts can be fine-tuned to further mitigate these concerns. This  
371 action can be achieved with explicit system messages and carefully structured user instructions  
372 (Figure. 2b). By combining short, iterative prompting workflows with expert oversight, local  
373 researchers can kick-start the use of LLMs responsibly—gaining exposure to AI-driven reviews  
374 while retaining the option to co-create more specialized, eco-friendly models as needs evolve  
375 (Farrell et al. 2024).

## 376 **8. Conclusion**

377 Large Language Models (LLMs) offer a promising approach to scaling literature reviews in  
378 ecology and conservation. From rapid filtering and classification to extracting structured  
379 quantitative and qualitative data, LLMs can address the bottlenecks posed by vast and ever-  
380 expanding scientific literature. Our two case studies highlight the variety of objectives that can  
381 benefit from AI-driven pipelines, including uncovering trends and performing evidence synthesis

382 for meta-analysis. Thus, rather than replacing human reviewers, we argue that LLMs can  
383 effectively serve as powerful assistants that lighten the burden of large-scale reviews, free up time  
384 for deeper scientific reasoning and synthesis, and expand the scope of evidence considered,  
385 including multiple languages and grey literature (Reynolds et al. 2024). Yet, LLMs are not a  
386 panacea; they must be deployed with robust prompt engineering, strategic checks on accuracy,  
387 and domain-expert oversight. In this way, ecologists and conservation practitioners can embrace  
388 the utility of advanced AI tools without sacrificing rigour or relevance. By following best practices—  
389 such as structured output, iterative prompting, and random validation checks—we can harness  
390 the unprecedented scale and speed of LLMs to advance ecological research and conservation  
391 outcomes more efficiently.

392

## 393 Acknowledgements

394  
395 SMKM was funded by the joint Smithsonian-Pembroke research fellowship. MQ was  
396 supported by a MSCA fellowship (EP/Z00232X/1) hosted by RSG. AR was supported by  
397 the Oxford BBSRC Doctoral Training Partnership (BB/T008784/1). YM was supported by  
398 the Jackson Foundation. RSG was supported by a NERC Pushing the Frontiers grant  
399 (NE/X013766/1). While OpenAI's GPT-4o mini, a large language model, is used for  
400 analysing the data, it was not used for writing or generating the codes used in the  
401 manuscript.

## 402 Conflict of Interest Statement

403  
404 Authors declare no conflict of interest.

## 406 Author Contributions

407  
408 SMKM, YM and RSG conceived the ideas for the study; SMKM and MQ collected and  
409 processed the data; SMKM analysed the data; SMKM and AR designed the figures; AR  
410 verified the codes; SMKM led the writing of the manuscript. All authors contributed  
411 critically to the drafts and gave final approval for publication.

## 413 Data Availability

414  
415 All the data and codes used in the study are made available via the following github  
416 repository: <https://github.com/sruthimoorthy/LLM-Lit-Review-Codes>

# REFERENCES

- 417
- 418 Alzoubi, Yehia Ibrahim, and Alok Mishra. 2024. "Green Artificial Intelligence Initiatives:  
419 Potentials and Challenges." *Journal of Cleaner Production* 468 (143090): 143090.
- 420 Arefeen, Md Adnan, Biplob Debnath, and Srimat Chakradhar. 2024. "LeanContext: Cost-  
421 Efficient Domain-Specific Question Answering Using LLMs." *Natural Language Processing*  
422 *Journal* 7 (100065): 100065.
- 423 Berger-Tal, Oded, Bob B. M. Wong, Carrie Ann Adams, Daniel T. Blumstein, Ulrika  
424 Candolin, Matthew J. Gibson, Alison L. Greggor, et al. 2024. "Leveraging AI to Improve  
425 Evidence Synthesis in Conservation." *Trends in Ecology & Evolution* 39 (6): 548–57.
- 426 Castro, Andry, João Pinto, Luís Reino, Pavel Pipek, and César Capinha. 2024. "Large  
427 Language Models Overcome the Challenges of Unstructured Text Data in Ecology."  
428 *Ecological Informatics* 82 (102742): 102742.
- 429 Cazalis, Victor, Karine Princé, Jean-Baptiste Mihoub, Joseph Kelly, Stuart H. M. Butchart,  
430 and Ana S. L. Rodrigues. 2020. "Effectiveness of Protected Areas in Conserving Tropical  
431 Forest Birds." *Nature Communications* 11 (1): 4461.
- 432 Cooper, Natalie, Adam T. Clark, Nicolas Lecomte, Huijie Qiao, and Aaron M. Ellison. 2024.  
433 "Harnessing Large Language Models for Coding, Teaching and Inclusion to Empower  
434 Research in Ecology and Evolution." *Methods in Ecology and Evolution* 15 (10): 1757–63.
- 435 Coutaz, Joëlle, James L. Crowley, Simon Dobson, and David Garlan. 2005. "Context Is  
436 Key." *Communications of the ACM* 48 (3): 49–53.
- 437 Extance, Andy. 2023. "ChatGPT Has Entered the Classroom: How LLMs Could Transform  
438 Education." *Nature* 623 (7987): 474–77.
- 439 Farrell, Maxwell J., Nicolas Le Guillarme, Liam Brierley, Bronwen Hunter, Daan  
440 Scheepens, Anna Willoughby, Andrew Yates, and Nicole Mideo. 2024. "The Changing  
441 Landscape of Text Mining: A Review of Approaches for Ecology and Evolution."

442 *Proceedings. Biological Sciences* 291 (2027): 20240423.

443 Gaulke, Sarah, Enzo Martelli, Luke Johnson, Carlos G. Letelier, Natalie Dawson, and Cara  
444 R. Nelson. 2019. "Threatened and Endangered Mammals of Chile: Does Research Align  
445 with Conservation Information Needs?" *Conservation Science and Practice* 1 (9).  
446 <https://doi.org/10.1111/csp2.99>.

447 Gougherty, Andrew V., and Hannah L. Clipp. 2024. "Testing the Reliability of an AI-Based  
448 Large Language Model to Extract Ecological Information from the Scientific Literature." *Npj*  
449 *Biodiversity* 3 (1): 13.

450 Huang, Jingwei, Donghan M. Yang, Ruichen Rong, Kuroush Nezafati, Colin Treager,  
451 Zhikai Chi, Shidan Wang, et al. 2024. "A Critical Assessment of Using ChatGPT for  
452 Extracting Structured Data from Clinical Notes." *Npj Digital Medicine* 7 (1): 106.

453 Idnay, Betina, Caitlin Dreisbach, Chunhua Weng, and Rebecca Schnall. 2021. "A  
454 Systematic Review on Natural Language Processing Systems for Eligibility Prescreening in  
455 Clinical Research." *Journal of the American Medical Informatics Association: JAMIA* 29 (1):  
456 197–206.

457 Johnson, Thomas F., Benno I. Simmons, Joseph Millard, Tanya Strydom, Alain Danet,  
458 Amy R. Sweeny, and Luke C. Evans. 2024. "Pressure to Publish Introduces Large-  
459 language Model Risks." *Methods in Ecology and Evolution* 15 (10): 1771–73.

460 Kallogjeri, Dorina, and Jay F. Piccirillo. 2023. "A Simple Guide to Effect Size Measures."  
461 *JAMA Otolaryngology-- Head & Neck Surgery* 149 (5): 447–51.

462 Lehtikoinen, Petteri, Andrea Santangeli, Kim Jaatinen, Ari Rajasärkkä, and Aleks  
463 Lehtikoinen. 2019. "Protected Areas Act as a Buffer against Detrimental Effects of Climate  
464 Change-Evidence from Large-Scale, Long-Term Abundance Data: XXXX." *Global Change*  
465 *Biology* 25 (1): 304–13.

466 Lin, Zhengyang, Anping Chen, Xuhui Wang, Zhihua Liu, and Shilong Piao. 2024. "Large  
467 Language Models Reveal Big Disparities in Current Wildfire Research." *Communications*

468 *Earth & Environment* 5 (1): 1–6.

469 McNeely, Jeffrey A. 1994. “Protected Areas for the 21st Century: Working to Provide  
470 Benefits to Society.” *Biodiversity and Conservation* 3 (5): 390–405.

471 Mittelstadt, Brent, Sandra Wachter, and Chris Russell. 2023. “To Protect Science, We Must  
472 Use LLMs as Zero-Shot Translators.” *Nature Human Behaviour* 7 (11): 1830–32.

473 Mohit, Behrang. 2014. “Named Entity Recognition.” In *Natural Language Processing of*  
474 *Semitic Languages*, 221–45. Berlin, Heidelberg: Springer Berlin Heidelberg.

475 Peeperkorn, Max, Tom Kouwenhoven, Dan Brown, and Anna Jordanous. 2024. “Is  
476 Temperature the Creativity Parameter of Large Language Models?” *arXiv [cs.CL]*. arXiv.  
477 [https://computationalcreativity.net/iccc24/papers/ICCC24\\_paper\\_70.pdf](https://computationalcreativity.net/iccc24/papers/ICCC24_paper_70.pdf).

478 Proudfoot, Beatrice, Patrick L. Thompson, Krista Bohlen, Mike Collyer, Jennifer  
479 Yakimishyn, and Clifford L. K. Robinson. 2023. “Eelgrass Meadow Edge Habitat  
480 Heterogeneity Enhances Fish Diversity on the Pacific Coast of Canada.” *Estuaries and*  
481 *Coasts: Journal of the Estuarine Research Federation* 46 (5): 1326–44.

482 Qi, Man, Matthew Gadd, De Daniele Martini, Katrina Davis, Biao Xiong, Alice Rosen, Nick  
483 Hawes, and Rob Salguero-Gomez. 2025. “Biodiversity Research Requires More Rotors  
484 and Wheels on and above Ground, as Well as below Water.”  
485 <https://ecoevorxiv.org/repository/view/8354/>.

486 Qureshi, Riaz, Daniel Shaughnessy, Kayden A. R. Gill, Karen A. Robinson, Tianjing Li, and  
487 Eitan Agai. 2023. “Are ChatGPT and Large Language Models ‘the Answer’ to Bringing Us  
488 Closer to Systematic Review Automation?” *Systematic Reviews* 12 (1): 72.

489 Ren, Shaolei, Bill Tomlinson, Rebecca W. Black, and Andrew W. Torrance. 2024.  
490 “Reconciling the Contrasting Narratives on the Environmental Impact of Large Language  
491 Models.” *Scientific Reports* 14 (1): 26310.

492 Reynolds, Sam A., Sara Beery, Neil Burgess, Mark Burgman, Stuart H. M. Butchart,  
493 Steven J. Cooke, David Coomes, et al. 2024. “The Potential for AI to Revolutionize



494 Conservation: A Horizon Scan.” *Trends in Ecology & Evolution*, December.  
495 <https://doi.org/10.1016/j.tree.2024.11.013>.

496 Sacoransky, Ethan, Benjamin Y. M. Kwan, and Donald Soboleski. 2024. “ChatGPT and  
497 Assistive AI in Structured Radiology Reporting: A Systematic Review.” *Current Problems in*  
498 *Diagnostic Radiology* 53 (6): 728–37.

499 Torres, Leigh G., Dawn R. Barlow, Todd E. Chandler, and Jonathan D. Burnett. 2020.  
500 “Insight into the Kinematics of Blue Whale Surface Foraging through Drone Observations  
501 and Prey Data.” *PeerJ* 8 (e8906): e8906.

502 Tsertsvadze, Alexander, Yen-Fu Chen, David Moher, Paul Sutcliffe, and Noel McCarthy.  
503 2015. “How to Conduct Systematic Reviews More Expeditiously?” *Systematic Reviews* 4  
504 (1): 160.

505 Vasiliev, Yuli. 2020. *Natural Language Processing with Python and spaCy: A Practical*  
506 *Introduction*. San Francisco, CA: No Starch Press.

507 Whitfield, Sharon, and Melissa A. Hofmann. 2023. “Elicit: AI Literature Review Research  
508 Assistant.” *Public Services Quarterly* 19 (3): 201–7.

509 Yang, Bo, Juying Jiao, Xiaowu Ma, Wenting Zhao, Qi Ling, Xinhua Zhang, Jianqiao Han,  
510 Pengfei Du, Yin Chen, and Hao Chen. 2023. “Distribution and Formation of Soil Balls under  
511 Heavy Rainstorm Conditions in the Northern Loess Plateau.” *Journal of Hydrology* 625  
512 (130103): 130103.

513

# Supplementary Section

514

515

516

## **Harnessing Large Language Models for Ecological Literature**

517

### **Reviews: A Practical Pipeline**

518

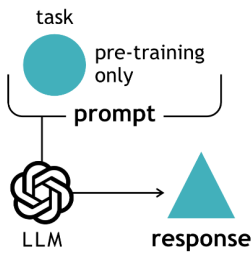
Sruthi M. Krishna Moorthy<sup>1,2,3\*</sup>, Man Qi<sup>2</sup>, Alice Rosen<sup>2</sup>, Yadvinder Malhi<sup>3,4</sup>, Roberto

519

Salguero-Gomez<sup>1,2</sup>

# 1. Different prompting styles with examples

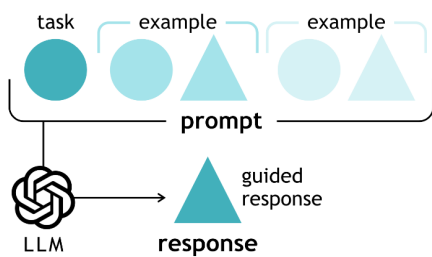
## a. Zero-shot prompting



Identify the study organisms monitored by drones in the abstract. The abstract is within triple back ticks.  
 Abstract:  
 ```Eelgrass (Zostera marina) meadows are important...```

The study organisms monitored by drones in the abstract are fish species that inhabit eelgrass meadows in British Columbia. The research focused on understanding how environmental variables, particularly those derived from unmanned aerial vehicles (UAVs), influence fish diversity in these meadows.

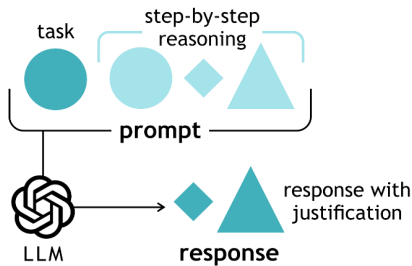
## b. Few-shot prompting



Identify the study organisms monitored by drones in the abstract. The abstract is within triple back ticks.  
 Example: ```Paddy field farms are being converted to provide...```  
 Answer: No organisms monitored by drones. Only ditch feature is monitored.  
 Abstract:  
 ```Eelgrass (Zostera marina) meadows are important...```

The study organisms monitored by drones in the second abstract are **fish in eelgrass meadows**. The research focused on understanding how seascape patterns and environmental variables influence fish diversity in these habitats.

## c. Chain-of-thought prompting



Identify the study organisms monitored by drones in the abstract. For instance, consider a study with main objective of understanding how the bat population is affected by forest structure measured by drones. Since the drone is deployed to measure the forest structure directly and not the bats, the correct response for this question is plants and not bats. The abstract is within triple backticks.  
 Abstract:  
 ```Eelgrass (Zostera marina) meadows are important...```

In the abstract provided, the study organisms monitored by drones are the **seascape variables** such as meadow area and edge habitat heterogeneity, which influence the **fish diversity** in eelgrass meadows. The drones are used to monitor the environmental variables and not the fish. Therefore the correct response to the study organisms monitored by drones is **seascape variables**.

523 **Figure S1.** Illustration of different prompting styles in action for identifying the organisms  
 524 monitored by drones in an ecosystem from the abstract of the study (Proudfoot et al. 2023).

## 525 2.Final prompts

### 526 2.1. Case Study 1: Uncovering Research Trends in 527 Autonomous Biodiversity Monitoring

528

#### 529 System prompt

530 You are a research assistant specialized in ecology. You carefully read the text you are given,  
531 extract the requested information from the text and you only respond in JSON format as  
532 instructed. You never make up data that is not present in the text.

#### 533 User prompt

534 Extract information from the abstract on the application of drones or robots or UAVs to study  
535 any organisms and/or the environment. Your task is to perform the following actions:

536 1. Read the content of the abstract in its entirety

537 2. Specifically, extract the structured information on the following entities from the abstract:

538 Entity 1 - Country: Identify and return the standardized ISO3 names of the countries where the  
539 study was conducted. Only include countries where the study locations or water bodies studied  
540 are located. Do not include countries mentioned in the text that are not directly related to the  
541 study location.

542 Entity 2 - Biomes: Identify the biomes of the study. Check if the biomes belong to the following  
543 categories: Forests, Shrublands, Grasslands, Savannas, Mangroves, Wetlands, Coastal, Coral  
544 reef, Marine. Wetlands refers to all terrestrial water-logged ecosystems including peats,  
545 swamps, lakes, rivers and streams. Arid deserts should be classified as "NA" and semi-arid  
546 desert shrublands should be classified as "Shrublands". Coastal refers to the coastal  
547 ecosystems like delta, intertidal zones, coastal aquaculture, sea grass communities, estuarine,  
548 tidal marshes, lagoons, rocky shores. RETURN "NA" FOR BIOMES OTHER THAN Forests or

549 Shrublands or Grasslands or Savannas or Mangroves or Wetlands or Coastal or Coral reef or  
550 Marine.

551 Entity 3 - Kingdom: Identify the study organisms monitored by drones in the abstract. Return  
552 the kingdoms the organisms belong to. The values for this key should be one of those within the  
553 square brackets: ["Animalia", "Plantae", "Fungi", "Protista", "Archaea", "Bacteria", "NA"].

554 RETURN ONLY THE BIOLOGICAL KINGDOM AS VALUES. For instance, consider a study  
555 with the main objective of understanding how the bat population is affected by forest structure,  
556 measured by drones. Since the drone is deployed to measure the forest structure directly and  
557 not the bats, the correct response for this question is Plantae and not Animalia.

558 3. Do not interpret what the robots could technically measure or monitor. First find the relevant  
559 information in the abstract and then generate the response for a specific entity based on the  
560 information. If the information is not present in the text, use "NA" as the value.

561 4. Output your response as a JSON object as in the following format.

562 Country: List of countries

563 Biomes: RETURN ONLY ITEMS FROM THE LIST ["Forests", "Shrublands", "Grasslands",  
564 "Savannas", "Mangroves", "Wetlands", "Coastal", "Coral reef", "Marine", "NA"]

565 Kingdom: RETURN ONLY ITEMS FROM THE LIST ["Animalia", "Plantae", "Fungi", "Protista",  
566 "Archaea", "Bacteria", "NA"]

567 Comment: Justify your responses for all the entities

568 If more than one value is present or identified for an entity, return all the values in a comma-  
569 separated form.

570 Do not return values that are not listed within square brackets.

571 The abstract text is delimited with triple backticks.

572 ""{abstract}""

573

574                   2.2.     *Case Study 2: Evidence Synthesis for Terrestrial*  
575                                    *Protected Areas*

576    *Filtering relevant literature*

577  
578    System prompt

579    You are a research assistant specialized in ecology. You carefully read the text you are given  
580    and compare it to the specified inclusion criteria. You only respond in JSON format as  
581    instructed. You never make up data that is not present in the text.

582    User prompt

583    Below is an excerpt from a scientific study regarding protected areas. Determine whether it  
584    meets all of these criteria for inclusion in a meta-analysis:

- 585    1. The study focuses on terrestrial protected areas.  
586    2. It provides a quantitative comparison of an ecologically similar non-protected area versus the  
587    protected area(s).  
588    3. The study measures species richness or abundance as an outcome.

589    Please return one of two possible decisions in the field "overall\_inclusion\_decision":  
590    "yes" if the study meets all the above criteria.  
591    "no" if the study fails one or more criteria.

592    Along with the final decision, provide a short explanation for why you made that decision in a  
593    field called "justification".

594    The text to analyze is delimited with triple backticks.

595    IMPORTANT: Return your answer in JSON format only, using the structure below.

596    Example JSON Output:

597    "overall\_inclusion\_decision": "yes or no",  
598    "justification": "Short explanation of which criteria are met or not met."  
599    ""{abstract}""



601 Extracting effect sizes for meta-analysis

602

603 System prompt

604 You are a research assistant specialized in ecology. Your task is to carefully read the text you  
605 are given from a study's Results section and identify:

606 1. The effect size value(s) reported. Effect size is a standardized measure that  
607 compares groups or conditions (e.g., difference in species richness)

608 2. The effect size metric used (e.g., Cohen's d, Hedge's g, log response ratio, etc.).

609 3. The reported measure of uncertainty (e.g., standard error, confidence interval) and  
610 its numeric values if available.

611 4. Any relevant notes if the effect size or its uncertainty are not clearly stated.

612 You must output only in JSON format, without additional commentary. You never make up data  
613 that is not present in the text.

614 User prompt

615 Below is a segment of text from a study on terrestrial protected areas and biodiversity.

616 Task:

617 1. Identify any genuine standardized effect sizes that compares PAs with non PAS (e.g.,  
618 Cohen's d, Hedge's g, log response ratio, correlation coefficient, etc.) if they are explicitly  
619 reported.

620 2. Identify which metric is used to represent the effect size.

621 3. Identify the statistical method used to report the metric

622 4. Capture the measure of uncertainty (standard deviation, confidence interval, standard error)  
623 and its value(s).

624 5. State what the effect size specifically represents with the mention of the taxonomic group if  
625 applicable (e.g., species richness, species abundance, etc.).

626 6. Summarize if the final outcome of the effects of protected areas to be positive or negative



627 7. If multiple effect sizes are reported, list them all.

628 Important:

629 If the text does not contain any standardized effect size but only mentions interaction terms,  
630 reflect that in the JSON (e.g., an empty effect\_sizes array).

631 Example JSON Output:

632 "metric": "e.g., Cohen's d, Hedge's g, log response ratio",

633 "method": "e.g., Linear mixed model,",

634 "effect-size": "numeric value or 'not reported'",

635 "uncertainty\_type": "e.g., 95% CI, standard error, standard deviation",

636 "uncertainty\_values": "numeric range or 'not reported'",

637 "represents": "what does the effect size quantify (e.g., bird species richness, overall species  
638 abundance)?",

639 "notes": "additional context or 'none'"

640 The text to analyze is delimited with triple backticks.

641 ""{abstract}""

642

643

644

645

646