The role of large language models in interdisciplinary research: opportunities, challenges, and ways forward

Christos Mammides[1, *] and Harris Papadopoulos[2]

1. Nature Conservation Unit, Frederick University
2. Computational Intelligence (COIN) Laboratory, Department of Electrical Engineering, Computer Engineering and Informatics, Frederick University

*Correspondence:

Christos Mammides;

cmammides@outlook.com

Running headline

LLMs in interdisciplinary research

Abstract

1. Large language models (LLMs) are gaining importance in research as they offer many benefits. One often overlooked benefit is their potential to facilitate and support interdisciplinary research, which is key to addressing current global challenges, such as the twin crises of biodiversity loss and climate change.
2. LMMs can help reduce the costs associated with knowledge transfer and bridge gaps between different fields of study. They can also be especially useful in helping ecologists understand and adopt significant techniques common in other fields.
3. However, using LLMs in research carries important risks, including the possibility of generating inaccurate information that can lead to false conclusions.
4. We recommend that researchers adhere to best practices when using LMMs for research by providing appropriate prompts and dividing complex tasks into smaller, more manageable tasks that facilitate learning and testing. Additionally, journals could implement policies to ensure that information and code generated using LMMs are properly validated. Moreover, academic programs should incorporate formal training in analytical and programming tools, which will equip researchers in ecology with the necessary skills to use LMMs effectively and responsibly.

Keywords

ChatGPT; ecological research; Gemini; generative artificial intelligence; large language models; machine learning

1. Introduction

Large language models (LLMs), such as Gemini and ChatGPT, play a growingly important role in scientific research. In their article, Cooper et al. (2024) explain many of the benefits of integrating LLMs into research activities in ecology and evolution. Researchers in other fields have also listed such benefits and have suggested guidelines for the ethical and responsible use of LLMs (Lubiana et al., 2023). Here, we highlight another potentially important contribution of LLMs, which is perhaps less discussed but could have a substantial impact. The potential of LMMs to facilitate interdisciplinary research.

2. LMMs in Interdisciplinary Research

It is now accepted that global challenges, such as the twin crises of biodiversity loss and climate change, require interdisciplinary solutions. For example, projects funded under the European Union's Horizon Europe programme, aimed at developing solutions to address biodiversity loss, often require the integration of multidisciplinary expertise. However, despite its importance, interdisciplinary research remains limited. A major obstacle is the absence of common language and understanding, which hinders effective communication and collaboration between researchers from different fields (Pellmar and Eisenberg, 2000). Consequently, key concepts and tools available in one field may not be adopted in another despite their potential to make a significant impact.

We have found that LLMs can facilitate interdisciplinary research by reducing costs associated with knowledge transfer and bridging gaps between fields. For example, in our project, BIOMON, we are exploring the use of machine-learning techniques to monitor bird communities using acoustic sensors. Our team consists of CM, a Conservation Biologist, and HP, a Machine Learning expert, focusing on Conformal Prediction, a framework for quantifying the uncertainty of machine learning predictions. While this framework has been employed successfully in many other fields (Papadopoulos and Haralambous, 2011), it has not yet been applied in ecology despite its potential to be used to identify uncertain predictions. During our project implementation, LLMs have been instrumental in promoting effective communication by offering, for example, customised "crash courses" in each other's areas of expertise, using a language familiar to each party. Importantly, LLMs can be consulted repeatedly at one’s convenience, providing prompt feedback (Cooper et al., 2024) at no burden to the other party. Another way LLM models have been pivotal in our collaboration, especially for CM, is for coding purposes, particularly for developing code for newly adopted methods and in new programming languages. For example, although many ecologists these days tend to have a good grasp of data analysis using the R programming language, many of the cutting-edge machine-learning techniques are being developed in Python. LLMs have the potential of making the adoption of new techniques from unfamiliar to the researcher programming languages considerably easier.

3. Challenges associated with LMMs

The use of LLMs does not come without issues. Researchers have identified multiple risks associated with using LMMs to generate code (Lubiana et al., 2023). A significant risk is that LMMs often “hallucinate”, producing inaccurate information and code that relies, for example, on non-existent functions or packages. Additionally, LMMs may produce "silent errors" that are difficult to detect since the code appears to be running correctly. Still, it is not performing the intended task, leading to false conclusions (Lubiana et al., 2023). As a solution, it has been suggested that the code be carefully checked and tested (Lubiana et al., 2023). However, this can be a daunting task for researchers with no formal training in coding and validation methods, which is often the case in ecology. The issue may be even more pronounced when LMMs are used in an interdisciplinary context and applied to less familiar techniques and programming languages. Despite these challenges, the use and usefulness of LMMs will continue to rise. Consequently, it is crucial we develop robust solutions that will allow us to benefit from the availability of LMMs while minimising the risks associated with their use.

4. Recommendations and Conclusions

First, researchers should take steps to ensure LMMs are used as responsibly and accurately as possible. For instance, one approach when using LMMs to develop code is to do it incrementally instead of feeding large and complex tasks to LLMs, which could be misinterpreted even if subtly and result in errors. Breaking down coding into smaller tasks has several advantages, such as providing better control over the coding process, allowing each added code segment to be reviewed more effectively, and facilitating the learning process by making it easier to identify which line of code produces each desired output.

Second, academic journals must acknowledge and address the multifaceted use of LLMs in research. While several measures have been taken already, the rate by which LLMs develop emphasises the need to expedite these efforts. Many journals already require authors to acknowledge the use of LLMs in their research methods, but additional steps should be taken to ensure transparency and accuracy. Journals must continue promoting open research and encourage authors to share the code used in analyses (Fernández-Juricic, 2021). Many of the barriers to sharing code can now be overcome using LLMs, which can help researchers optimise, comment, and test code (Cooper et al., 2024). Journals must also consider who is reviewing and validating the code. Oftentimes, the code is written exclusively by one of the authors, e.g., the junior researcher, while the rest of the co-authors, reviewers, and editors have no access to or are unable to review it. As more code will inevitably be generated using LLMs in the future, journals could consider appointing code reviewers for papers that use sophisticated code for their analyses.

Third, we recommend that academic programs include formal training in data analysis and computer programming. Unfortunately, there are still too many programs in Ecology and Evolution, including in our home country, that fail to provide students with the necessary skills to meet the current realities and needs of the field. Given the growing availability of sophisticated methods for data collection and analysis, including machine learning methods, and the increased recognition of their potential in ecology and conservation, students must have a basic understanding of how to use them properly. This includes knowing how to write and test code efficiently (Cooper, 2017). Having this knowledge will equip ecologists with the skills needed to use LMMs efficiently but responsibly.

In conclusion, we consider ourselves fortunate to be conducting research in this exciting era and look forward to future advancements in LLMs. LLMs can positively impact research, including interdisciplinary research. However, like any other disruptive tool, we must take steps to ensure that they are employed efficiently and effectively.

Acknowledgements

We thank Natalie Cooper, Aaron Ellison, and Adam Clark for organizing and delivering the workshop “Coding with Chat-GPT”, which CM attended during the British Ecological Society's annual meeting in 2023. BIOMON is funded by the European Union’s Horizon Europe programme, ERA Talents, under grant agreement 101090273. The manuscript's title is a modified version of a title recommended by Chat-GPT v3.5. Yes, we made sure to include the word “please” when asking for title suggestions, just in case.

Conflicts of interest

The authors have no conflicts of interest to declare.

Authors' Contributions

CM and HP conceived the ideas and outlined the manuscript; CM led the writing of the manuscript. All authors contributed critically to the drafts and gave final approval for publication.

References

Cooper, N. (2017). A guide to reproducible code in ecology and evolution. British Ecological Society. Available at: http://hdl.handle.net/10141/622618

Cooper, N., Clark, A. T., Lecomte, N., Qiao, H., and Ellison, A. M. (2024). Harnessing large language models for coding, teaching and inclusion to empower research in ecology and evolution. *Methods in Ecology and Evolution*. doi: 10.1111/2041-210X.14325

Fernández-Juricic, E. (2021). Why sharing data and code during peer review can enhance behavioral ecology research. *Behav Ecol Sociobiol* 75, 103. doi: 10.1007/s00265-021-03036-x

Lubiana, T., Lopes, R., Medeiros, P., Silva, J. C., Goncalves, A. N. A., Maracaja-Coutinho, V., et al. (2023). Ten quick tips for harnessing the power of ChatGPT in computational biology. *PLoS Comput Biol* 19, e1011319. doi: 10.1371/journal.pcbi.1011319

Papadopoulos, H., and Haralambous, H. (2011). Reliable prediction intervals with regression neural networks. *Neural Networks* 24, 842–851. doi: 10.1016/j.neunet.2011.05.008

Pellmar, T. C., and Eisenberg, L. (2000). "Barriers to interdisciplinary research and training," in *Bridging disciplines in the brain, behavioral, and clinical sciences*, (National Academies Press (US)).