EGU24-18493, updated on 13 Jul 2024
https://doi.org/10.5194/egusphere-egu24-18493
EGU General Assembly 2024

Leveraging recent advances in Large Language Models for the ocean science community

Redouane Lguensat
IPSL - IRD, Paris, France (redouane.lguensat@ipsl.fr)

Large Language Models (LLMs) have made significant strides in language understanding, including natural language processing, summarization, and translation, and they have the potential to be applied to a range of climate-related challenges. For instance, LLMs can be leveraged for data cleaning and transformation, and also assisting scientists/engineers in their daily work tasks.

For the machine learning community, the year 2023 was arguably the year of breakthroughts in LLM use in production. I present in this work the exciting potential for recent advances in LLMs to revolutionize how the ocean science community can interact with computer code, information gathering, dataset finding, etc. Specifically, I will present simple applications of how these advancements in Natural Language Processing (NLP) can assist the NEMO ocean model community. Examples range from using question answering systems for browsing efficiently NEMO documentation to creating conversational agents or chatbots that can assist not only new members wanting to learn about the NEMO model but also confirmed users.

An important aspect of this work is relying only on open source LLMs, evaluating the performances of several models and discussing the ethical implications of these tools. I also discuss the question of whether using these LLMs blindly without domain knowledge is a good idea, as an important chunk of this work can arguably be easily done by anyone with good computer science skills thanks to the democratization of data science tools and learning materials.