# Poster Summary: Configuring Large Language Models for Regional Ocean Model Development


Aidan Janney
CGD Oceanography Section
NSF NCAR
Boulder, CO, United States
aidan.janney@colorado.edu

Giovanni Seijo-Ellis
CGD Oceanography Section
NSF NCAR
Boulder, CO, United States
gesijo@ucar.edu

Dan Amrhein
CGD Oceanography Section
NSF NCAR
Boulder, CO, United States
damrhein@colorado.edu


## POSTER SUMMARY

Ocean models support our understanding of long-term variations and sensitivities in atmospheric and marine systems around the Earth, and regional ocean models give researchers the ability to resolve more complex processes that affect weather, climate, and ecosystems on a local level [4].

New modeling software is regularly updated and released to provide access to regional models. However, once a model can compile and run, researchers are faced with a difficult and sparsely documented process: model tuning. Tuning a model involves changing different features of computation and parameterization (e.g. viscosity, mixing, input fields, boundary conditions) to nudge the behavior of a model towards observational results or experimental goals (e.g. change temperature and current strength in a specific region to reflect observations). There are many degrees of freedom in this process, and researchers must use a combination of scientific and institutional expertise when changing parameters and adjusting the model. Tuning is essential for producing effective models of the ocean, but it often walks a line between improvisation and iteration, as unexpected combinations of changes might lead to the largest improvement [1]. To compound this challenge, researchers often record only the final settings of a model that were successful, and the process by which they reached that point is left without documentation. The work in this poster presents a procedure for tailoring LLMs to bridge this gap in experience and resources by providing guidance in development and tuning.

Tailoring LLMs towards a specific purpose involves a fine-tuning framework and curated data. This study uses open-source models (Llama 3), proprietary models (ChatGPT 4) with parameter-efficient fine-tuning (PEFT) fine-tuning and context-tuning methods to steer model output with a custom dataset of interviews from expert regional ocean modelers.

Fine-tuning LLMs is most effective when working with a large, high-quality dataset, but the target application of model tuning lacks almost any data. We document the creation of a small, but extremely high-quality dataset consisting of reformatted interviews with regional ocean modelers. The expert ocean modelers were prompted with questions about their experiences tuning models and how they approach solving different types of technical and abstract issues with models. The transcripts from six interviews are reformatted to resemble an interaction between an LLM and a user querying the LLM about model tuning.

Fine-tuning is the process of taking a pre-trained LLM and adjusting the model weights to be more effective for a specific task [2]. Traditionally, fine-tuning involves modifying the entire parameter space of the LLM, which can be expensive and slow, but there are parameter-efficient techniques that only modify a subset of parameters while maintaining comparable performance improvements. The technique used in this poster is Low-Rank Adaptation (LoRA) [2]. The reformatted conversations interviews are processed and tokenized to be fed to the LLM, and the model undergoes several training steps and sets of training steps to refine its output based on these data. At the highest level, the LLM is given a prompt, it forms a response, and then it is given the expected response to score against; these scores are then optimized to produce the "best" response (i.e. more aligned with the data). LoRA outputs a set of weights that sits on top of the base weights of an LLM; these weights are generated from a set of lower rank Context-tuning is a much cheaper and less intensive process where we provide the LLM with information that it uses to inform its response. This can be highly effective, acting like memory, but LLMs also have a finite context window that can quickly become overwhelmed. There are additional ways to make context-tuning more dynamic and specific, like retrieval-augmented generation (RAG) [3]. LoRA was used to fine-tune an open-source Llama 3.2-Instruct model, and context-tuning was used to tailor a custom ChatGPT-4O model from OpenAI.

This poster documents some comparisons between the behavior of these models, before and after fine-tuning and context-tuning. The Llama 3.2 model, while much smaller than ChatGPT-4O, produced positive responses after fine-tuning that were significantly influenced by the training data. Promisingly, all of the models generated relevant guidance for model tuning.

## REFERENCES


1. Frédéric Hourdin, Thorsten Mauritsen, Andrew Gettelman, et al. 2017. The Art and Science of Climate Model Tuning. .

2. Edward J. Hu, Yelong Shen, Phillip Wallis, et al. 2021. LoRA: Low-Rank Adaptation of Large Language Models. Retrieved August 25, 2025 from http://arxiv.org/abs/2106.09685.

3. Zhuowan Li, Cheng Li, Mingyang Zhang, Qiaozhu Mei, and Michael Bendersky. 2024. Retrieval Augmented Generation or Long-Context LLMs? A Comprehensive Study and Hybrid Approach. Retrieved August 25, 2025 from http://arxiv.org/abs/2407.16833.

4. Jeff Polton, James Harle, Jason Holt, et al. 2023. Reproducible and relocatable regional ocean modelling: fundamentals and practices. *Geoscientific Model Development* 16, 5: 1481–1510.