Ryan Gross
G47667332
CSCI 6917
Guided Research Grad I
Project Proposal

Project Title: Leveraging Large Language Models (LLM) for AIS Trajectory Prediction

What are you going to do?

Leverage state-of-the-art LLMs, such as GPT-3.5 and GPT-4, to predict the trajectory a vessel will take in the future. This will be accomplished by feeding the first half of a vessel trip into the system, and asking the system to plot the rest of the trip. This is very similar to asking ChatGPT a question and getting a response back. All LLMs are based on the transformer, which is a sequence-to-sequence (seq2seq) model. All sea2seq models take input and try to complete the output. The task of vessel trajectory prediction is very similar to sentence completion, but instead of language the system is completing an array of coordinates.

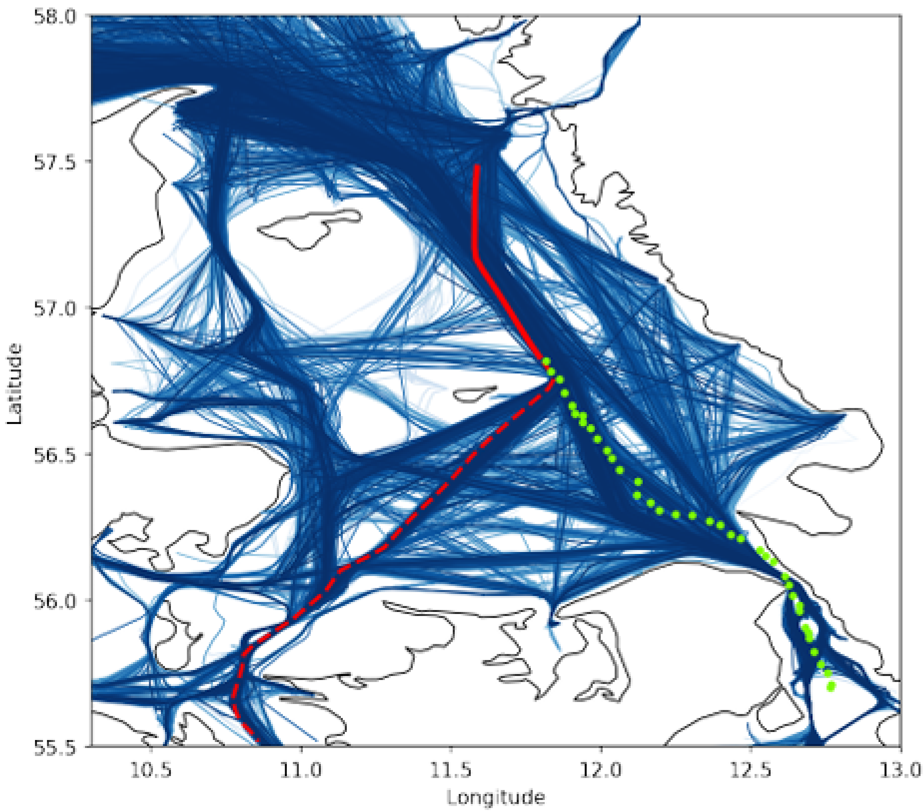

Figure 1. The solid red line represents the first half of the historical set of locations from a vessel's AIS receiver. The point where a solid red line turns into a dashed red line represents where the system starts predicting the future trajectory. The dashed red line was the actual trajectory, and the dotted green line was the prediction. In this example the prediction is incorrect.

How is it done today? Current Limitations?

In the past, other machine learning techniques have been employed such as Gausian distribution models, Bayesian reasoning, and older seq2seq methods such as Recurrent Neural Networks (RNN) and Long-Short Term Memory (LSTM) models. More recently researchers have utilized the transformer architecture. One paper [1] even utilized

GPT-2 to predict vessel tracks using AIS data. This paper utilized the open source code from Open AI and modified it to train the transformer from scratch specifically for the purposes of vessel trajectory prediction. All the aforementioned methods work well, but require substantial amounts of time to build and train.

What is your idea to do something better?

The goal of this project is to see if the amount of time required to build a system from scratch can be eliminated or reduced by leveraging pre-trained LLMs. LLMs are beating previous machine learning methods at many tasks and GPT in particular has very good zero and few-shot learning abilities. If a problem can be re-stated in a seq2seq, input-output format, then it is a candidate for leveraging pre-trained LLMs. Even if the LLM doesn't at first perform well, these models can be "fine-tuned" which allows for additional training on a customer specific dataset. This project will help to determine whether pre-trained LLMs are versatile enough to handle non-language specific tasks.

Who will benefit from your work? Why?

Other researchers wanting to know how versatile LLMs are would benefit from the findings of this project. If successful, this and many other findings from the Computer Science research community would start to build the narrative that LLMs are very useful across a wide spectrum of machine learning tasks. Specifically for vessel trajectory prediction, the US Coast Guard, US Navy, and other law enforcement agencies monitoring vessel traffic could benefit from a deployed system with these prediction capabilities.

What risks do you anticipate?

These models tend to be very verbose, making it difficult to get concise and structured responses back. With the code not being open source, it can be harder to modify this behavior. The going-in mitigation strategy is to use fine-tuning to provide examples of structured inputs and outputs to teach the system how to provide structured output for this specific task.

Out of pocket costs? Complete within 11 weeks?

Open AI has a typical cloud API pricing model where thousands of API calls cost a few cents. Microsoft Azure also hosts Open AI services for similar costs. I am willing to pay these out-of-pocket costs as I really want to get my hands on these APIs to see what the system is capable of. In terms of schedule, there is low risk of not completing the experiment in 11 weeks. This project requires code to format data, interact with the APIs, and format output results, but the amount of coding required presents a low risk to the schedule.

Midterm Results?

By the midterm period, some basic prediction should be working even if the accuracy is poor. This requires the basic shell of the code to format input data, call the APIs, and format the output data. This may not necessarily include fine-tuning of the model. Results will be measured against performance measures done in the GPT-2 based paper [1], to graph prediction error over time.

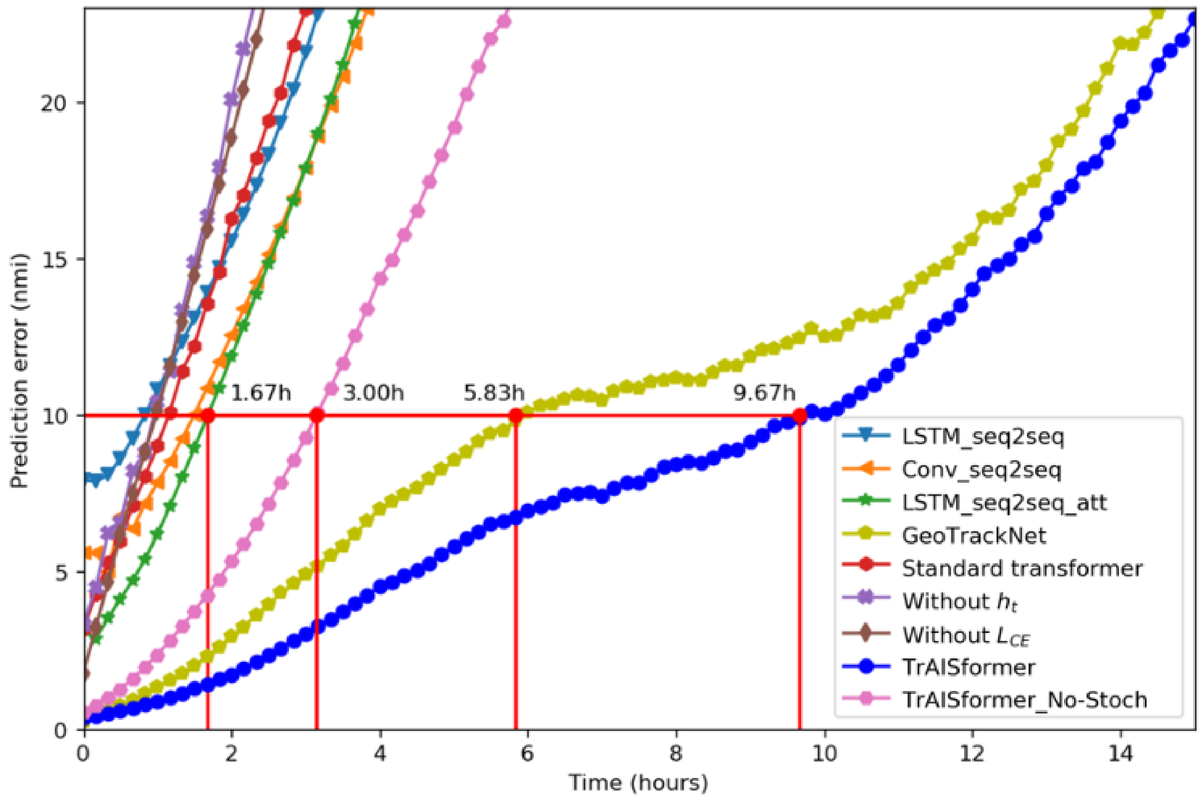

Figure 2. Many machine learning techniques attempting to predict vessel trajectories based on the same AIS training data. Results are measured in prediction error over time.

Final Demonstration?

The final demonstration will be the same as the mid-term, but hopefully with improved prediction error over time. This provides a direct comparison across many methods with the exact same dataset. Prediction error below 10nmi after 10 hours would be an incredible result. The period between the mid-term and final results will be where most of the fine-tuning experimentation will be done, as well as trying out different models provided by Open AI

References

[1] Nguyen, Duong and Ronan Fablet. “TrAISformer-A generative transformer for AIS trajectory prediction.” ArXiv abs/2109.03958 (2021): n. pag..