Timeline:
Project Pitch: June 28, 9am PT
Projects: June 29 - Aug 28
Demo Day: August 28

The NewAtlantis Labs Fellowship Program:

This program focuses on teaching through hands-on projects. Mentors will provide support and guidance throughout the process. By the end of the program, researchers will have contributed new open source tools to the marine data science ecosystem while gaining practical experience in modern software engineering practices. They will be well-prepared to apply their mathematical talents to make an impact in machine learning and beyond.

The ocean is the world's largest ecosystem. It covers more than 70% of the earth's surface, is home to 80% of all life, produces 50% of the earth's oxygen, plays a critical role in the fight against climate change and yet much of this ecosystem remains a mystery. By enabling better access to ocean data through open source software developed in this program, NewAtlantis Labs seeks to support researchers working to understand and regenerate ocean health.

Mentor:
Stanley Bishop is NAL's Chief Architect. His career in scientific technology leadership started 15 years ago when he served as principal machine learning scientist for language localization for Google doing pioneering work deploying deep learning at cluster scale in the cloud with linguistic and scientific datasets. Stanley specializes in creating functional data systems to support scientific research at scale and has architected production systems used for precision medicine, marine biodiversity, and drug discovery. Stanley is currently affiliated with the Snyder Lab at Stanford (Stanford Genetics) where he served as a machine learning architect for Research to the People.

Stanley will kick off the project with you and then be available for weekly office hours in addition to being available via Slack during the off-hours.

Project One: Machine Learning Models for Real-Time Fishery Population and Health Forecasting

Leverage machine learning and diverse environmental datasets (eg: temperature, seamounts, seasonality) to develop robust predictive models for determining what variables most heavily impact the existence of chlorophyll which in turn impacts fishery population dynamics and ecosystem health. You will work with your mentor to define the scope of this project which could include:

- Develop Deep Learning learning models (e.g. transformers and CNNs) to predict fishery population metrics from the Fishbase database using satellite-derived environmental variables, chlorophyll levels, and biogeochemical data as features.

- Create an integrated modeling framework that combines environmental, biogeochemical, climactic, chlorophyll and fishery data to develop validated scales to provide a holistic view of and prediction model for fishery health.

- Explore enterprise quality software patterns building translational and comparative data structures to support feature generation and machine learning operations.

Fellows will use cloud computing platforms for processing large environmental datasets and training complex models. They will work with tools like Python, TensorFlow, and Google Earth Engine to build scalable prediction pipelines.

Technologies: Python, TensorFlow, Google Earth Engine, SQL, Docker, Kubernetes

Why this is important: Fisheries managers and policymakers need accurate, timely information on fish populations and ecosystem health to make informed decisions. Climate change is rapidly altering marine environments, creating an urgent need for adaptive management approaches. By leveraging diverse datasets and advanced machine learning, this project aims to provide robust real-time forecasting capabilities to support sustainable fisheries management in a changing climate. The resulting models and systems will enhance food security by enabling proactive measures to maintain healthy fish stocks. Additionally, tracking ecosystem changes will provide crucial data for climate change mitigation and adaptation efforts in marine environments.

Data elements for the project:

Ocean Color dataset from NASA

- https://oceancolor.gsfc.nasa.gov/

Fishbase Dataset:

- https://fishbase.se/

World Oceans Database:

- https://www.ncei.noaa.gov/products/world-ocean-database

Believe these three datasets we can almost def train some pretty good models, but will take a bit of poking to work out specifics.

Project Two: AI-driven MPA object monitors

Leverage the power of edge AI and computer vision to develop automated analysis pipelines of underwater video that can be easily deployed by marine biologists with minimal technical expertise.

You will work with your mentor to define the scope of this project which could be:

- Develop advanced image segmentation and classification algorithms to automatically identify and quantify plankton species from PlanktoScope images.

- Implement unsupervised learning techniques, such as clustering for anomaly detection, to discover novel or rare plankton species in PlanktoScope images. Useful for identifying potential indicator species or detect shifts in plankton community structure that may serve as early warning signals of ecosystem disturbances or regime shifts

- Integrate PlanktoScope image analysis with environmental sensor data, such as temperature, salinity, and nutrient levels, to explore the relationships between plankton community structure and environmental conditions.

- Create interactive visualization tools that allow researchers to explore PlanktoScope image datasets, visualize plankton community composition and size structure, and compare metrics across different spatial and temporal scales

Fellows will use hardware like NVIDIA Jetson modules and Raspberry Pi with neural accelerators to create robust deep learning models for tasks like object detection, segmentation and tracking of marine life. Technologies: Python, PyTorch, TensorFlow, OpenCV, embedded AI accelerators, DeepLabCut

Why this is important: Marine researchers need high-quality, scalable tools to efficiently gather and analyze large underwater video datasets for bio-monitoring of key ocean health indicators

like fish populations, coral reefs, and marine mammals. For example, computer vision is instrumental in assessing at scale key drivers impacting coral reef health measured through metrics such as coral cover. However, much of this video annotation work is still done manually by students, creating a major bottleneck. The ultimate goal is to accelerate ocean science and conservation by enabling the seamless.

Your Subject Mentor: Fabio Favoretto is a numerical ecologist with extensive expertise in marine conservation and data science. Holding a PhD in Marine Ecology and Remote Sensing, Fabio currently serves as a postdoctoral researcher at Scripps Institution of Oceanography in the Aburto Lab.

—----------

About NewAtlantis Labs:
NewAtlantis Labs is a first of its kind marine data platform providing long term diversified revenue potential to Marine Protected Areas and the governments, communities and NGOs that support them. By combining the latest in omics-led science, AI/LLM engineering tools and novel financial products NAL's work generates key insights into ecosystem health, potential discovery of novel genes for natural product sourcing as well as long term predictive modeling.