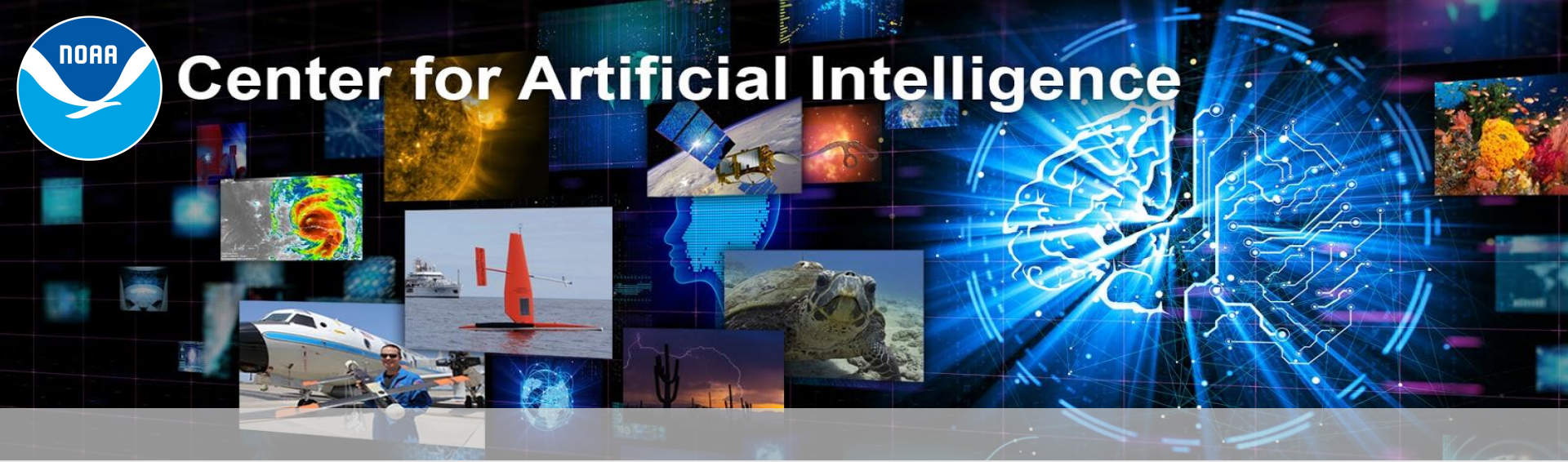


NOAA Center for Artificial Intelligence:
Getting our Agency AI-Ready Agency

Rob Redmon and the NCAI Team

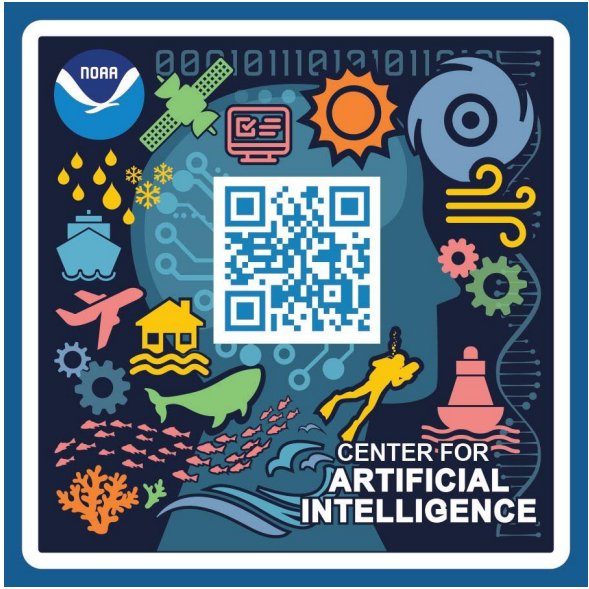
Unifying Innovations in Forecasting Capabilities Workshop

“Emerging Applications - Coastal, Fire Weather, Space Weather, Hydrology” - July 21, 2022

Agenda

- Motivating Challenges and Opportunities
- NOAA's Approach
  - National AI Initiative, Strategies and Plans
- NOAA Center for Artificial Intelligence
  - Who is NCAI?
  - Getting our feet wet with Pilot Projects and Initiatives
    - Engagement and Partnerships
    - Training the Workforce
    - Developing an "AI-ready" data standard
- Summary and Engagement

noaa.gov/ai



A place for publicly connecting to NOAA's 550+ member Community of Practice around AI for Earth system science to develop synergies and partnerships

NCAI Mailing List: tinyurl.com/y2ehvhfg

AI Call to Action

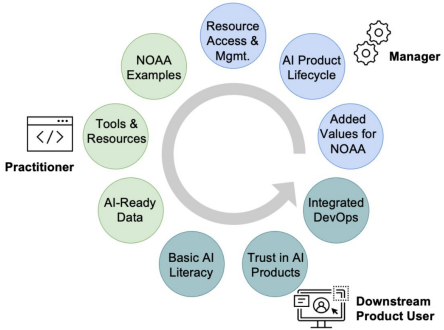
Key Needs for the Nation and the Public:

- Agencies trained on building Trustworthy & Equitable AI applications
  - For Digital Twin Earth systems fueling economic & equitable climate policies
  - Automating and augmenting stewardship of our national critical ecosystem resources and environmental observations
- AI/Data science-ready open environmental datasets
  - To power the Digital Blue Economy, Climate Ready Nation, Social Environmental Justice, and federated National AI Research Resources

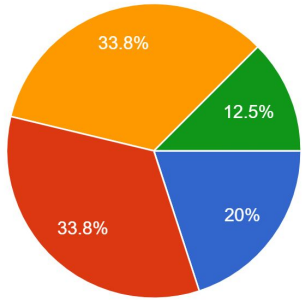
NCAI's Recommendations:

- Develop AI training standards centered around interactive training tools as "Learning Journeys" demonstrating Lessons Learned
  - This ensures AI applications are Trustworthy, maximizes technology transfer and directly benefits Research to Applications and Sustainment.
- Develop AI-ready data standards for open environmental data
  - This is the future of data stewardship, building upon Findable, Accessible, Interoperable and Reusable (FAIR principles) and Analysis Ready partnerships

Integrated AI Lifecycle

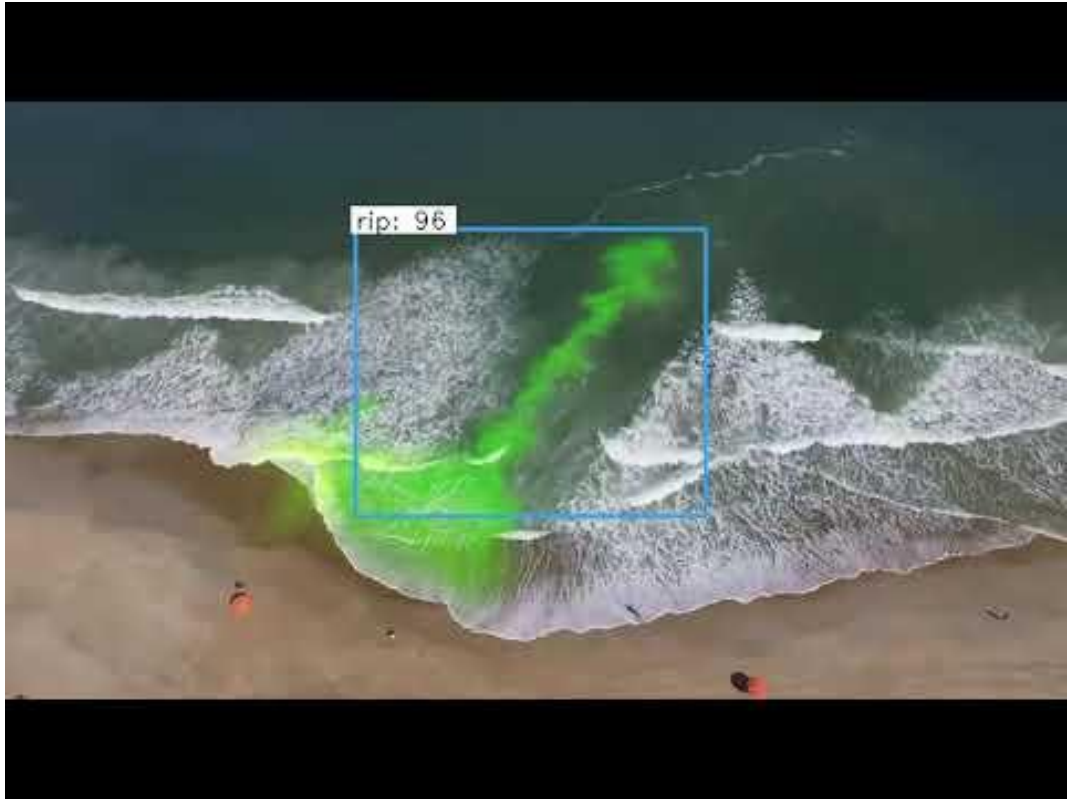


NOAA Survey: How much time do you spend to find, access and pre-process AI data?



Current and Potential Value for AI @ NOAA

Automated Rip Current Detection with Region based Convolutional Neural Networks



Flow-Based Rip Current Detection and Visualization (IEEE)

doi:10.1109/ACCESS.2022.3140340

Gregory Dusek (NOS) and UC Santa Cruz

Debra Hernandez, Southeast Coastal Ocean Observing Regional Association (SECOORA) Executive Director:

"Whether it's identifying a right whale or a rip current or shoreline erosion, we need faster analysis for more effective alerts to inform decision-makers."

secoora.org/noaa-launches-a-new-life-saving-rip-current-model

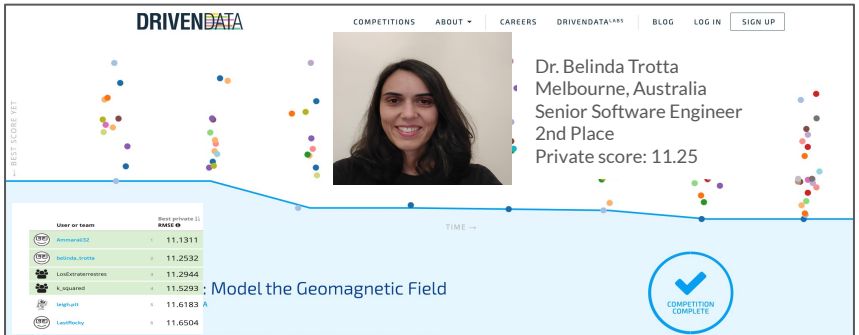
Video: https://arxiv.org/pdf/2102.02902.pdf

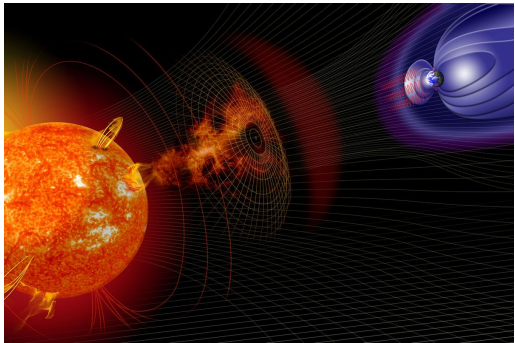
Partnership Development Through Competition

MagNet: Model the Geomagnetic Field - Internationally Open Machine Learning Competition

- *NCEI innovates* 2020 funded an open data-science competition to improve our ML model to predict *Dst* from solar-wind data, mitigating the impacts of space weather geomagnetic storm events on magnetic navigation systems
  - 600 participants and 1200 model submissions, with prize finalists achieving performance levels very close to the theoretical limit.
  - NCEI internally validated all four of the winning models
- R2O: Incorporate the winning model in the HDGM-RT (ongoing)
- Training: Documentation - Developed NCAI interactive *Learning Journeys* and supported the Trustworthy AI for Env. Science 2022 Summer School (June 27-30).
- Partnerships: Sponsored by NOAA, with support from NASA's Center of Excellence for Collab. Innovation (CoECI). Challenge was conducted by DrivenData and HeroX.
  - Long term collaboration with the 2nd place winner, Dr. Belinda Trotta
- Contacts: Manoj Nair (PI), Rob Redmon (Gov't POC)

Article in *Room Space* Magazine





NOAA Center for AI (NCAI)

National AI Initiative Act of 2020:
"The Administrator of NOAA [...] shall establish, a Center for Artificial Intelligence"

sciencecouncil.noaa.gov

Several Executive Orders, including:

- "Maintaining American Leadership in Artificial Intelligence"
- "Tackling the Climate Crisis at Home and Abroad"
- "Protecting Public Health and the Environment and Restoring Science To Tackle the Climate Crisis"

Related NOAA Strategic Plan Goals & Objectives

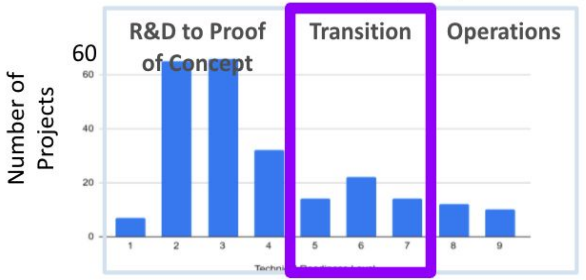
Foster an Information-Based Blue Economy:

NOAA will introduce innovation to data collection through various in-situ methods for species detection and explore AI/ML and data visualization technologies...

Ensure accessibility and enable an enterprise climate information framework to meet the needs of NOAA's users:

Analysis-ready datasets available (or percentage of existing satellite/other observational data made AI/ML ready on the cloud for climate, weather, oceans, etc. products and services)

2022 Data Call included 261 projects (188 in 2020)



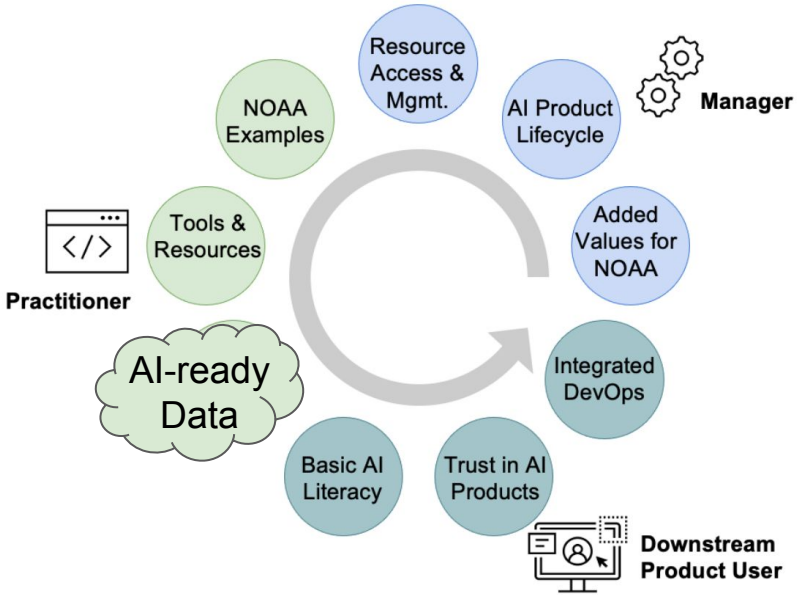
Initiative: Training the Workforce
Powering Discovery and Innovation

Training + AI-ready data → Trustworthy + Equitable Services

URGENT: Need **NOAA-specific training material** using NOAA data and computing resources to **remove common barriers** to the "Research to Operations, Applications, and Services" pipeline.

To address needs, resource creation should be prioritized to **convert NOAA AI success stories** into interactive training material in a **sandbox computing environment** that allows the workforce to apply learning outcomes to support NOAA's mission via the AI Strategic Plan.

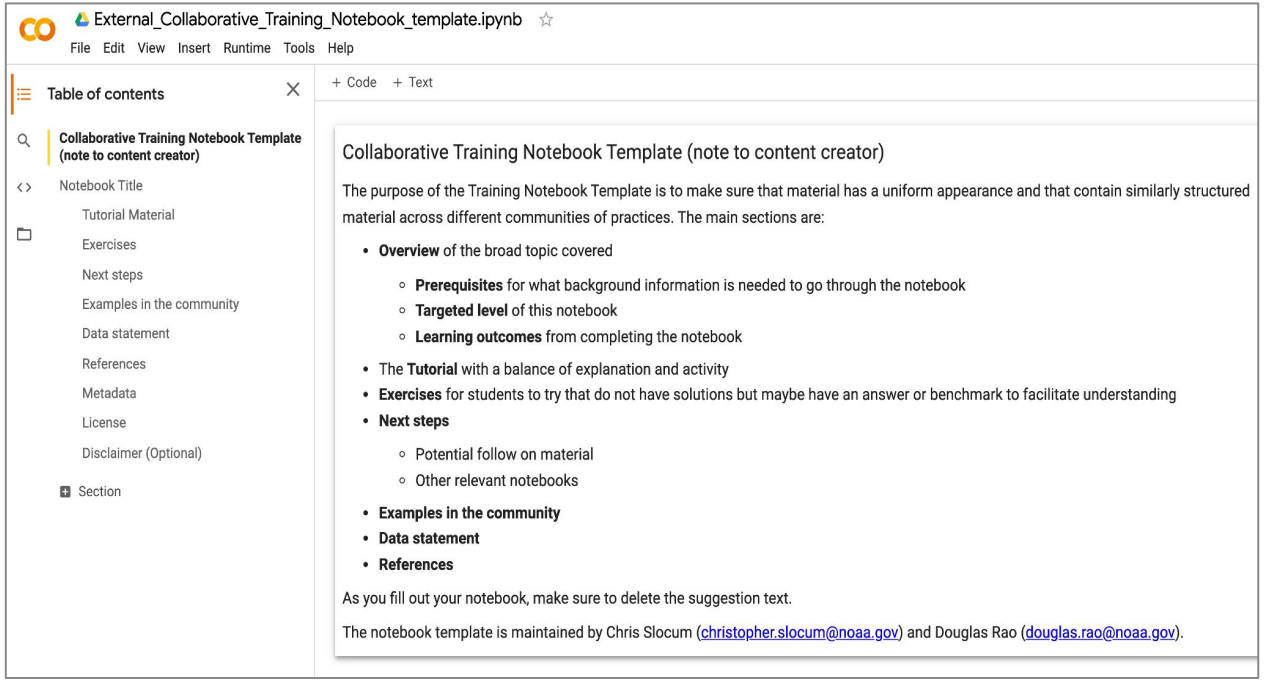
Factsheet: *noaa.gov/ai/training*



NOAA training action priority lifecycle highlighted by workforce role and relationship to AI.

Tool Development to Empower the Community

Jupyter notebook template with guidebook



github.com/ESIPFed/earth-science-community-ML-tutorials/tree/main/tutorial_template

Notebook readability assessment

```
Processing: ENSO_Seasonal_Forecasting.ipynb

Notebook metadata
   Language:Python 3
   Number Markdown Cells: 35
      Number of URLs: 7
      Number of valid URLs: 7
   Number Code Cells: 22
      Percent Text Output: 59%
      Percent Display Output: 50%

Readability metrics:
   Flesch reading ease:
      Score: 46.7
      Text interpretation:Difficult
      Grade level:College
   Flesch-Kincaid Grade Level: 12.8

Reading time estimates:
   Estimated text reading time:                13 to 25 min
   Estimated code reading time:                10 to 20 min
      Estimated code annotation reading time:  4 to 7 min
   Estimated total reading time:               23 to 45 min
```

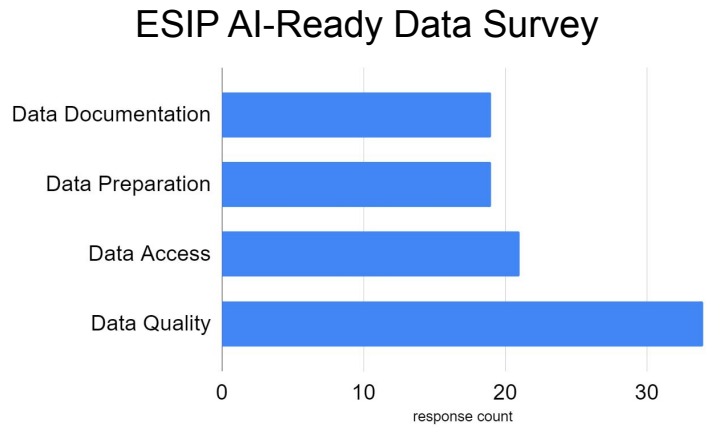
RATCHET - Readability Assessment Tool for Code that Helps with Effective Training

AI-Ready Data Initiative

An Earth Science Information Partnership (ESIP) survey was conducted in January -- Thanks for telling the Data Readiness Cluster about your data needs!
https://wiki.esipfed.org/Data_Readiness

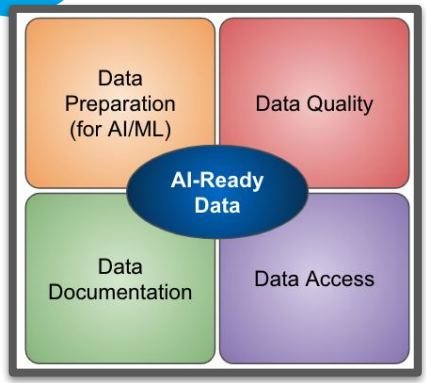
You're Not AI-ready Until Your Data Is

"The biggest roadblock to implementing a proof of concept for machine learning or deep learning is sourcing, organizing, and feeding the right kind of data into your model." – Intel.com



AI-Ready Data

Why?, What? and How?



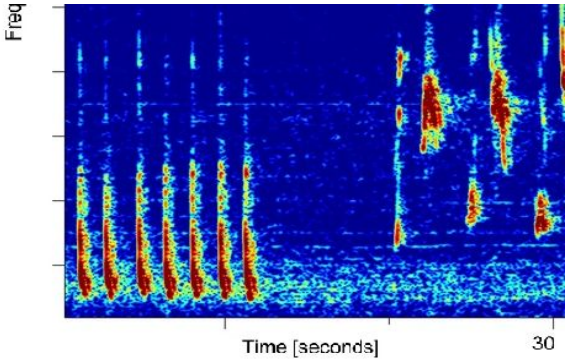
Goal: users spend less time data wrangling, more time on AI / ML

How can users find data that is easy to use in AI/ML?
How can data providers assess and improve usability?

What's needed:

- Community-driven definition of AI-readiness requirements
- Assessment tools for data providers and stewards
- Representation of readiness level conditioned on use / application
- Feedback and iteration to improve the standard
- Ideally, a formal standard (or set of standards)

NOAA leads a collaboration under the Earth Science Information Partners (ESIP) to develop a standard. Membership includes: US Federal agencies, academia, NGOs, private sector, and international (wiki.esipfed.org/Data_Readiness)



NOAA acoustic data used in deep learning to identify whale songs

fisheries.noaa.gov/science-blog/ok-google-find-humpback-whales

OK Google: Find the Humpback Whales

Pilot Project: Valuing NOAA's Data with Publications

Challenge: Can machine learning help value NOAA's data by connecting research articles and the data referenced in those articles?

Description and Expected Outcomes: Validate Natural Language Processing (NLP) model results to improve training datasets of data products and publications, while working with external partners. This funds NOAA's in-kind effort for the proposed CRADA (ETA June, 2022) with the Coleridge Initiative. Outcomes include a Lessons Learned presentation describing the model validation process and results, and quarterly status/updates.

Partners: Coleridge Initiative, NASA, USDA, NSF, Texas Supercomputing Center, Elsevier, and others. Strong interest from NOAA CDO, Commerce Department, and other federal agencies.

Goals:

- Help researchers find data used in their research topic
- Improved understanding our data users and ROI

How? NLP is a Machine Learning technique where algorithms identify patterns and context of words to find meaning in unstructured text documents

TRL: 3 (start), 5 (end)

Pilot Project: Platform Agnostic QC Development and Evaluation

Challenge

Increased adoption and use of uncrewed observing platforms add flexibility to observing systems, but also add to the diversity of observations. Automated data quality assessment capabilities is essential to plan for tsunami-wave of data. Build on promising results from PMEL Innovation Grant funded development.

Expected Outcomes

- A Machine Learning enabled QC implementation for near-realtime streamed data.
- Project will show relevance to:
  - Saildrone and gliders (year 1)
  - Satellite data (year 2)
- Develop a Python Notebook to be shared using NCAI template.
- Whitepaper on AI readiness of data delivered by Saildrone and data from the IOOS Glider DAC.

TRL: 2 (start), 4 (end)

Benefits

Rapid quality assessment of in situ observation measurements provides research and operational users insights to the quality of the data before data are applied, providing added utility and data value. This method will be tested against remotely observed (satellite) data.
NCAI benefits: AI-ready data assessment, interactive notebook for NCAI's Learning Journey library, potential technology reuse at NCEI.

A Saildrone in the Bering Sea with NOAA Ship Oscar Dyson. (Source)

Updated: June 14, 2022

Pilot Project:
Cloud Optimized Data Lake of Archived Water Column Sonar

Challenge

AI-ready Water Column Sonar (WCS) data are needed to extract the full potential of the ocean soundscape.

Description and Expected Outcomes

Develop a data lake of archived WCS data translated into cloud-friendly formats to improve interoperability and scaled processing.

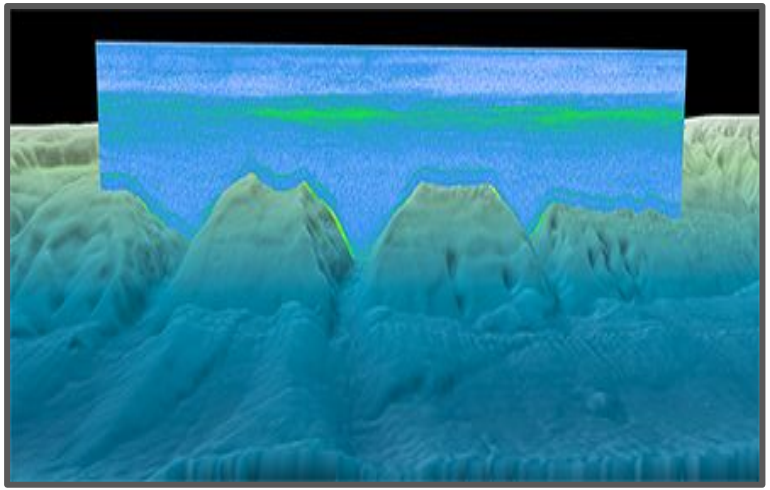
- The complex, binary, and priority formatted data are hosted on AWS currently through the NODD and are not AI-ready.
- Data will be transformed and made AI-ready and an interactive notebook will be developed for others to learn from.

Benefits

This effort will lower the learning curve to analyze the data more easily, especially in a cloud environment.

NCAI Benefits

AI-ready data, lessons learned as interactive notebook for NCAI's Learning Journey library.



Water column sonar data collected on the NOAA Okeanos Explorer in the North Atlantic Ocean. Sonar data are overlaid onto coastal relief model bathymetry. (Source)

Updated: June 14, 2022

Pilot Project: Create Tropical Cyclone Model Training Dataset

Challenge: AI-ready and accessible benchmark satellite datasets are needed to drive the future of tropical cyclone trajectory, intensity and coastal impact prediction (e.g. coastal flooding and other infrastructure damage).

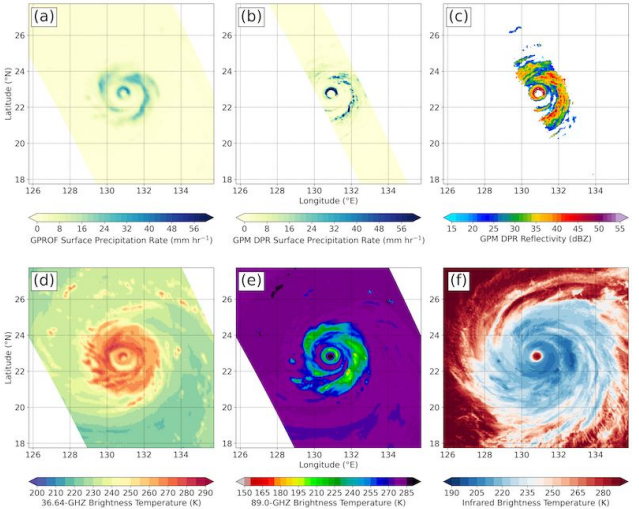
Description and Expected Outcomes: Evaluate a new dataset's AI-readiness against NCAI draft standards and make necessary changes to brand it as AI-ready. The new dataset, Tropical Cyclone PRecipitation, Infrared, Microwave, and Environmental Dataset (TC PRIMED), collocates and subsets LEO/GEO satellite imagery with ancillary model information to create a 22-yr dataset of TC-centric scenes. This dataset will supersede NCEI's HURSAT.

TRL: 6 (start), 8 (end)

Updates: Project funds routing to start execution.

NCAI Benefits: AI-ready standard maturation; Lesson Learned via interactive Python notebook; increased collaboration with NOAA's NODD (previously BDP).

POCs: Chris Slocum (NESDIS/STAR)



A sampling of TC PRIMED products from Typhoon Maria (2018) at 10:13 UTC on 9 July 2018 in the western Pacific, where a) is GPROF, b) is GPM DPR precipitation rate, c) GPM DPR reflectivity, d) 36.6 GHz, e) 89 GHZ, and IR from Himawari-8.

Pilot Project: Towards Fusing Humidity and Socioeconomic Data

Challenge

The influence of humidity on human heat stress is an understudied mechanism. This is due, in part, to the lack of a homogenized humidity dataset at the spatial (US County) and temporal (daily) resolutions necessary for coordinated analysis with public health data.

Description and Expected Outcomes

This project will work towards the creation of a spatially complete surface humidity dataset by blending remotely-sensed and in-situ surface humidity data using AI methods. The primary expected outcome is a beta version of a dataset that will meet requirements to align with public health data and associated socioeconomic metrics, along with a demonstration of a heat health application.

TRL: RL 2 (start), RL 4 (end)

Updates: Project to kick off July 2022

NCAI Benefits

It will demonstrate the utility of AI methods to create AI-ready datasets leveraging the advantages of both in-situ and remotely-sensed observations. AI-ready data, lessons learned as interactive notebook for NCAI's Learning Journey library.

Rising temperatures coupled with high humidity creates dangerous conditions for outdoor workers.
Source: Cyrus McCrimmon/Denver Post/Getty Images

POC: Jessica Matthews (NOAA/NESDIS/NCEI), jessica.matthews@noaa.gov

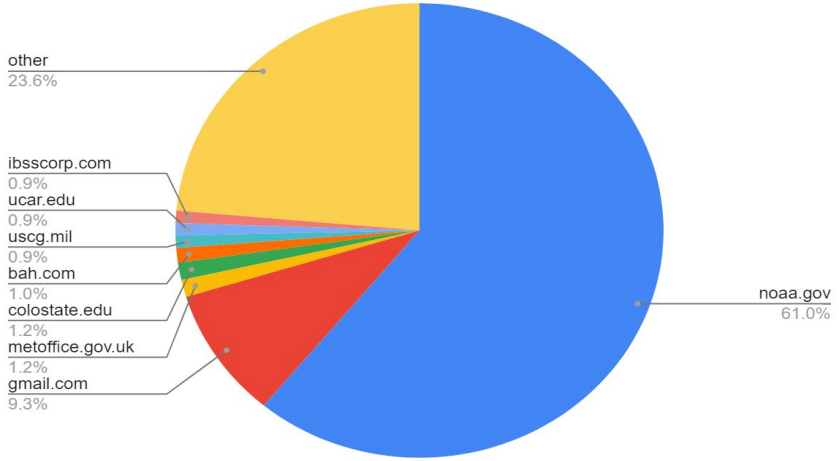
NCAI Community of Practice

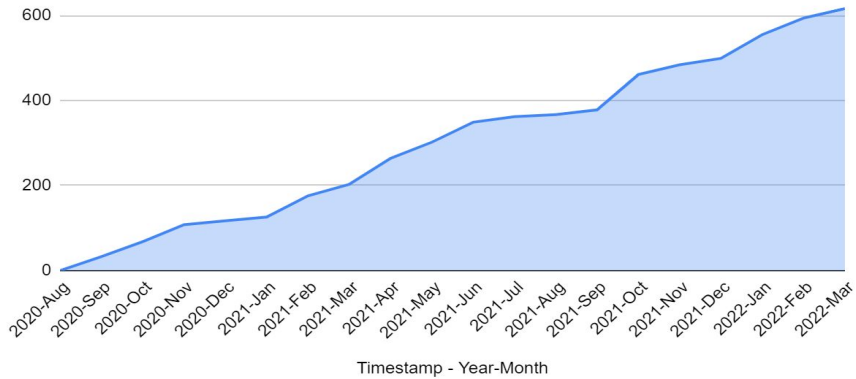
Join the mailing List: tinyurl.com/y2ehvhfg

608

Members in NOAA's AI Community of Practice from USG, Academia and Industry are looking to NCAI to facilitate conversations around infusing AI into Climate, Wx, Ecosystems and Environmental Justice.

- 608 (*as of May 2022*): more than half are NOAA
- 109 Organizations
  - NASA, USGS, USCG, WYO, USDA, US Navy, NREL
  - University of Colorado and Colorado State, Hawaii, North Carolina State, Albany, Alaska, Montana, Massachusetts, Montana, UC San Diego, Texas, Exeter and many more
  - AECOM, AccuWeather, BAH, IBSScorp, Riverside, tomorrow.io, Raytheon, and many more





Partnership Development

Through NOAA's 4th AI Workshop

Info and Registration: noaa.gov/ai

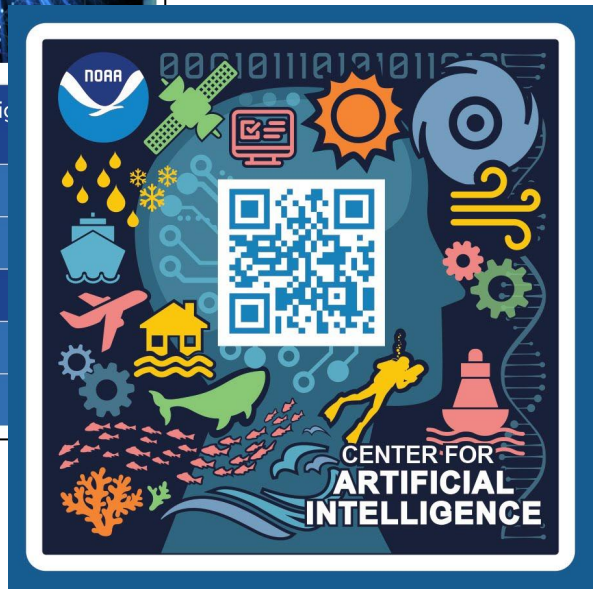
Jupyter notebook hacking with NOAA data for Fire Weather

Ocean application capabilities / tech transfer

Interoperable Digital Twins will leverage the best global abilities, connecting physical, social and policy science and application, (e.g. for vulnerability and mitigation).

BACKUP MATERIALS

Connect With NCAI



noaa.gov/ai

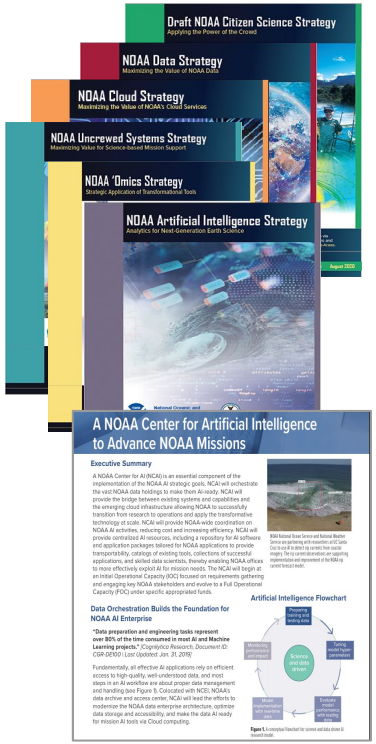
A place for publicly connecting to NOAA's 550+ member Community of Practice around AI for Earth system science to develop synergies and partnerships

NCAI Mailing List: tinyurl.com/y2ehvhfg

NOAA's AI Strategy and Plans for a NOAA Center for AI

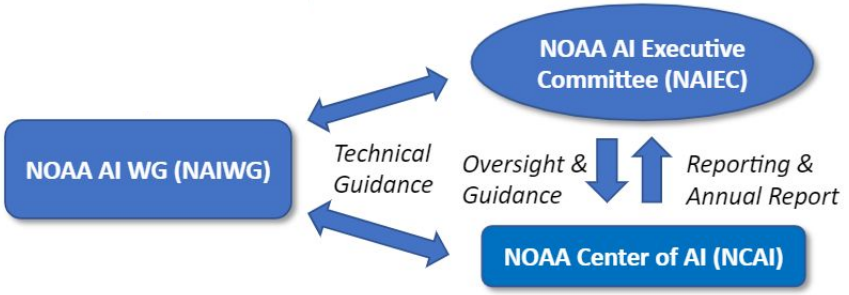
sciencecouncil.noaa.gov/NOAA-Science-Technology-Focus-Areas

- **Goal 1: Organization & Process - *Develop Congressionally Authorized NCAI***
  - Program Office coordinated with Public and Private Partners;
  - AI expertise embedded in each LO supporting Mission Scientists;
- **Goal 2: Advance AI Research and Innovation in Support of NOAA's Mission**
  - Stimulation of AI outcomes across all mission areas with long-term impacts via Grants and Partnerships
- **Goal 3: Accelerate the Transition of AI Research to Applications (R2X)**
  - Bridging the R2X "valley of death" with a fully curated repository of AI software, apps, and policies on ethics, mission validation metrics, ops reqs and an AI App Handbook;
- **Goal 4: Strengthen & Expand Partnerships**
  - A robust and fully realized AI partnership program to leverage capabilities from commercial, academic and government partners.
- **Goal 5: AI Proficiency**
  - AI fully capable workforce established through widespread benchmark AI-ready data, Learning Journeys library, multiple developmental sandboxes, and professional training



NOAA AI Working Group (NAIWG)

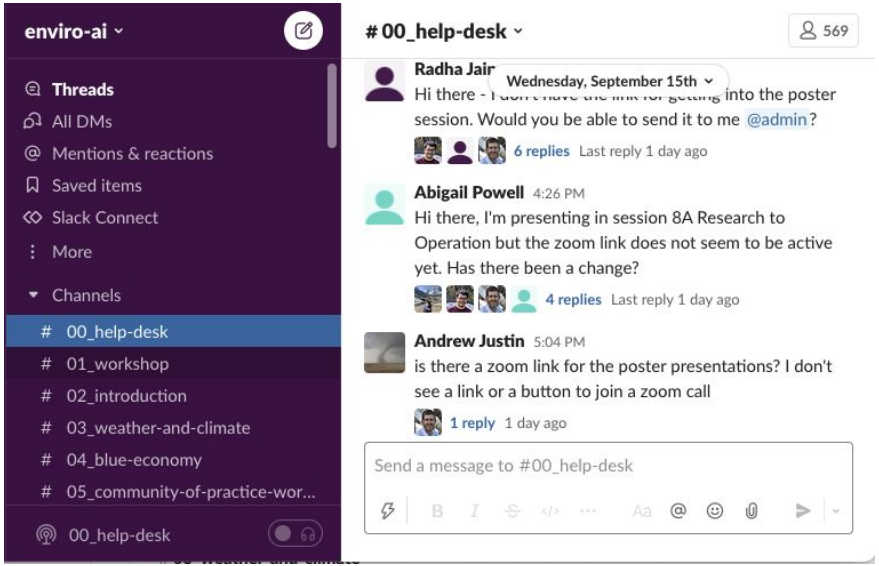
- Support NCAI to infuse AI across NOAA's Line Offices and mission areas
- Fulfill the role of distributed team members
- Lead activities that support the Center-level Implementation Plan elements which directly support NOAA's AI Strategic Plan
- Provide technical guidance in the definition, development and execution of the Center
- Support the NAIEC and Center on executing, messaging and optimizing AI S&T actions
- Support the alignment of dependent and co-productive actions and activities of other S&T Strategies



Partnership Development
Through Workshop Engagement

580 ***Slack Workspace Members***

3381 ***Messages in Slack during the workshop***



193 messages
#00_help-desk
Asking for help for workshop

118 messages
#02_introduction
Creating networking opportunities

435 discussions
#03_weather-and-climate
Active discussion for AI research in weather and climate

156 discussions
#10_hackathon
Active engagement during hackathon events

Partnership Development

Through Workshops and Hackathons

Erin Moreland
NMFS

"This is the most engaging virtual workshop I've attended"

Yagmur Derin
Hackathon participant

"You are doing an amazing job! I really am enjoying this hackathon and learning a lot."

Amy McGovern
AI2ES

"The workshop was fabulous! And the posters were the best online poster halls I have experienced."

Julie Demuth (NCAR Scientist, and Risk communication researcher)

"I wanted to express a hearty congratulations to you and everyone who was on the organizing committee, especially the logistic organizing committee. All of the technical and logistical aspects were impressively and incredibly well done. In the past 18 months, I've had many not-so-great virtual conference experiences and only a couple of experiences that were actually well done. This workshop knocked it out of the park! I know it was the result of a LOT of hard work by you all. Kudos!"

Ann Bostrom (University of Washington, Professor in Env Policy)

"You all did a fantastic job of running the workshop, and coordinating all the technologies so that they integrated well and improved the workshop experience. It was a pleasure to participate"

Weiming Hu
Tutorial participant

"I attended the workshop tutorial and it was a very informative one. I learned a lot. Thank you for organizing."

Philippe Tissot
Texas A&M University - Corpus Christi

"What a job! This was great including the diversity of presentations, speakers, topics, panels .... Impactful contribution to our fast growing field."

V. Ramaswamy (Director, OAR/GFDL)

"The talks not only related to the frontiers in this field through their technical details, but also provoked interesting ideas and perspectives. This should impart more momentum to advancements in the various topics. I found several of the questions to be probing. They drew thoughtful responses, pointing to directions where the field can be taken to as part of the future developments. All in all, an excellent and satisfying Conference. Kudos to everyone."

NOAA's 4th AI Workshop Ideation

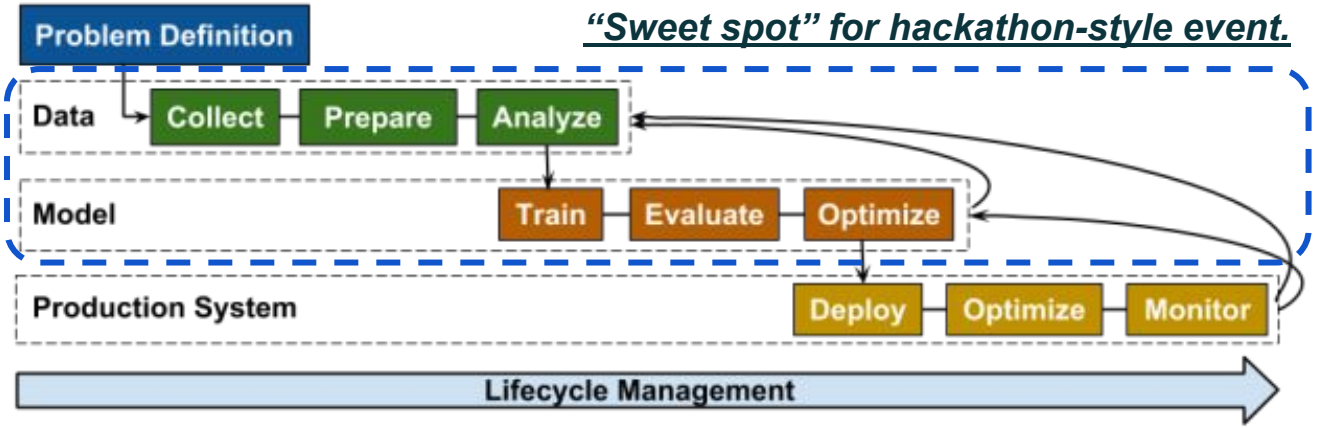
Bridging the Gap Between Awareness and Actions

In 2022, the 4th NOAA AI Workshop will be a Hackathon approach.

Infusing ***NOAA priority areas*** in problem definition process by ***engaging with diverse stakeholders***.

Improving ***AI-readiness*** for NOAA data

Addressing NOAA priorities through ***co-development***



Facilitate ***workforce training*** & ***partnership*** to accelerate AI adoption.

U.S. 2021 Billion-Dollar Weather and Climate Disasters

Drought/Heat Wave | Flooding | Hail | Hurricane | Tornado Outbreak | Severe Weather | Wildfire | Winter Storm/Cold Wave

Western Wildfires 2021

Central Severe Weather July 8–11

Midwest Derecho and Tornado Outbreak December 15

North Central Severe Weather August 10–13

Central Severe Weather June 24–26

Ohio Valley Hail Storms June 17–18

Southeast, Central Tornado Outbreak December 10

Eastern Severe Weather March 27–28

Southeast Tornadoes and Severe Weather

California Flooding and Severe Weather January 24–29

Western Drought and Heat Wave 2021

oma
ther
–28

tern
ave
–19

Texas Hail Storms April 12–15

Hurricane Nicholas September 14–18

Louisiana Flooding May 17–18

Souther
Southea
May 2–

Hurricane Ida
August 29–S

location for each of the **20 separate billion-dollar weather and climate dis**

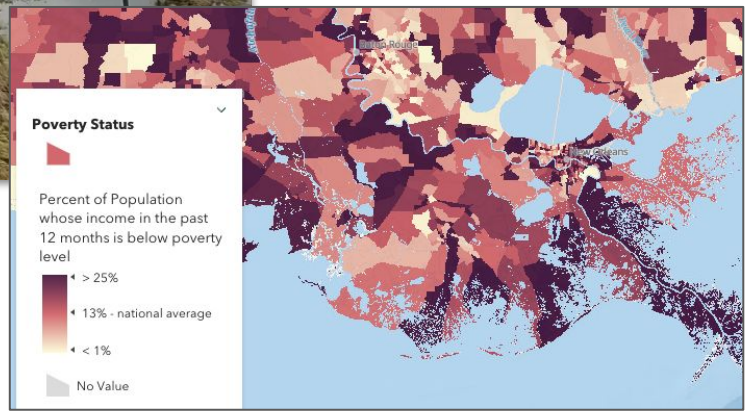
ncdc.noaa.gov/billions

Hurricane Ida and Social Inequity

Map: 2021 Heat Anomalies showing areas that are hotter or cooler than the average temperature for that same area. Red, orange and yellow indicate warmer than average temperatures. (Sources: The Trust for Public Land, USGS)

In the past, storms like Hurricane Ida could have simply been seen as a natural disaster, affecting both economically advantaged and disadvantaged alike.

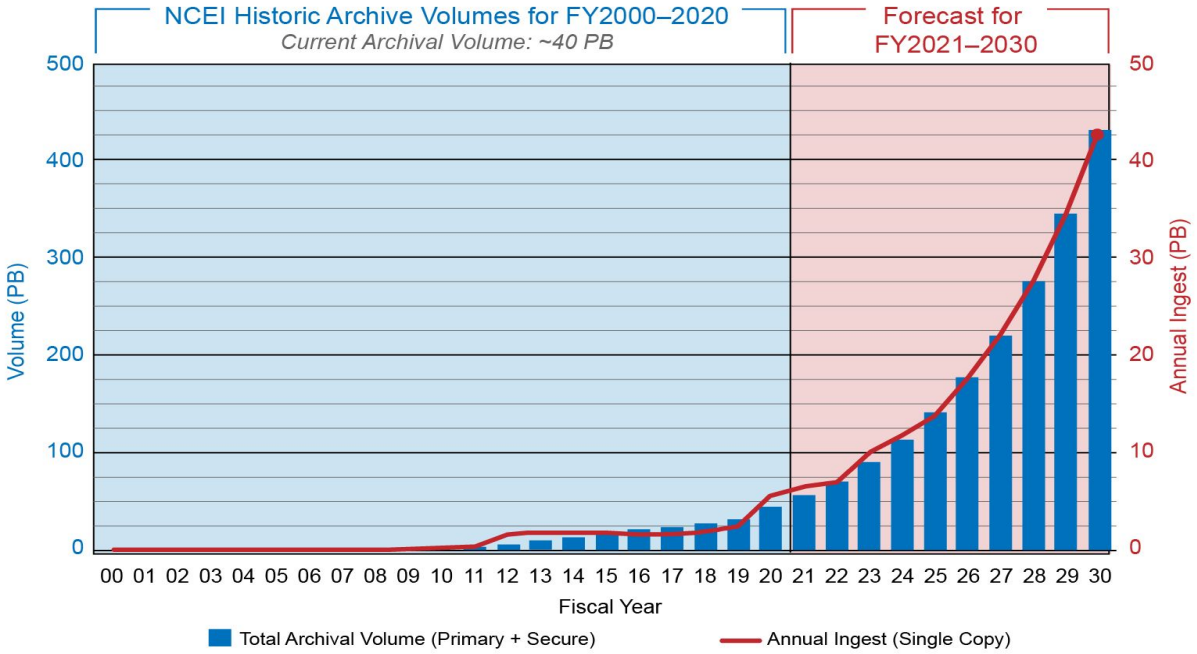
But it's becoming readily apparent that low-income communities suffer more damage and are at greater risk from extreme events. Research shows they are less prepared for the effects of extreme weather events.

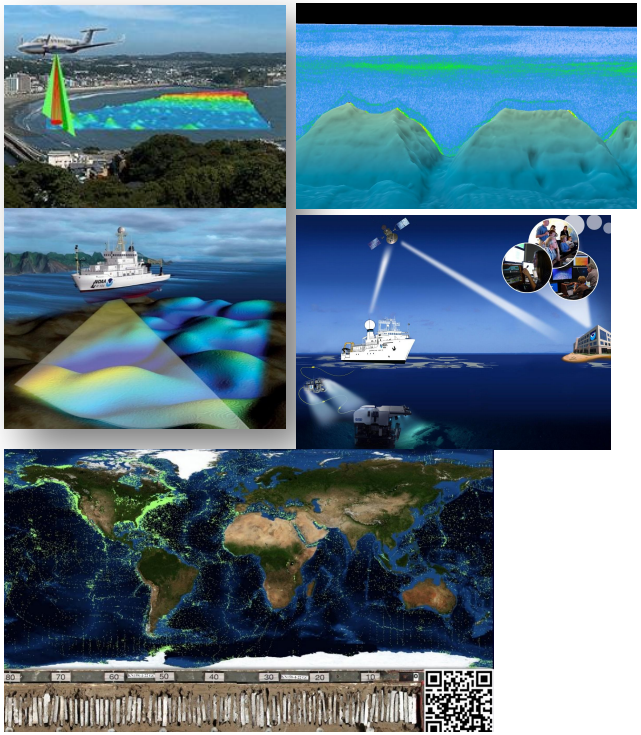


www.nesdis.noaa.gov/news/harms-way-hurricane-idas-impact-socially-vulnerable-communities

NCEI Archival Volume History and Forecast

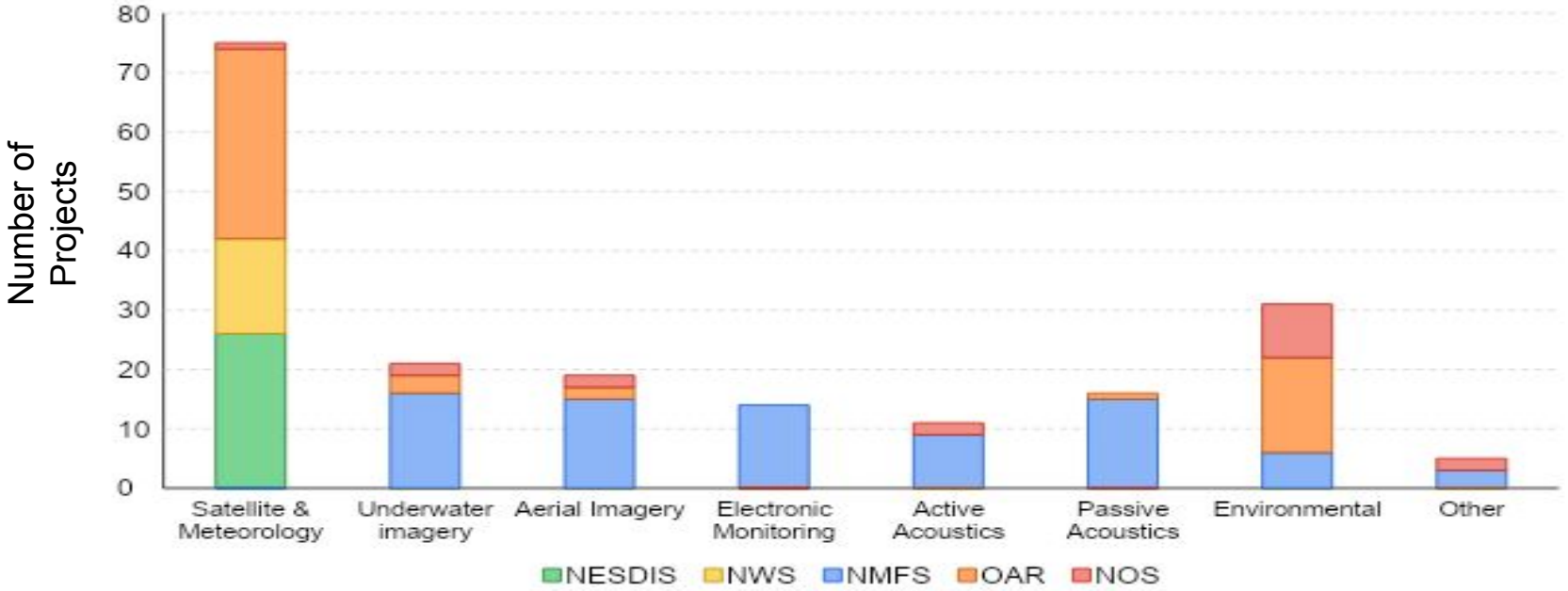
Increasing Data Volumes from Station, Model, Radar, UxS, Acoustics, 'Omics, and Satellite Sources





NOAA AI Data Call 2020 – Data Types

NOAA has applied AI-ML to a variety of environmental data demonstrating its interdisciplinary research and operational capabilities in support of its cross-functional mission requirements.



AI Representation and Access for Everyone

Envisioning a National Artificial Intelligence Research Resource (NAIRR)

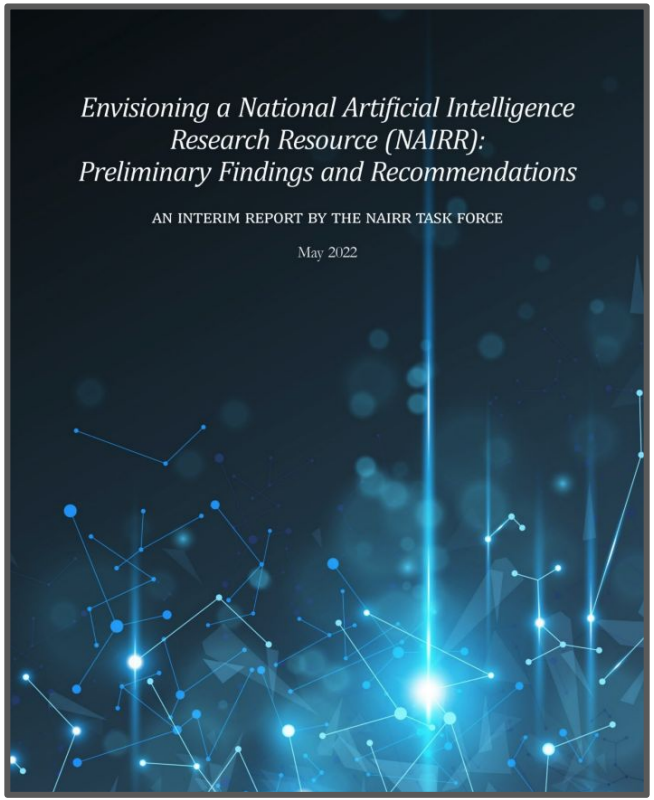
May 2022

"Much of today's cutting-edge AI research relies on access to large volumes of data and advanced computational power, which are often unavailable to many researchers beyond those at large technology companies and well-resourced universities.

This lack of access can lead to communities—particularly those that have long been underrepresented or underserved—being left out of the AI research and development process.

In turn, this lack of diversity in the design of novel AI systems limits the breadth of ideas incorporated into AI innovations and contributes to biases and other systemic inequalities."

Source: OSTP
ai.gov/wp-content/uploads/2022/05/NAIRR-TF-Interim-Report-2022.pdf



How can AI go wrong?

Ways in which AI can go wrong for environmental sciences

Issues related to training data:

1. Non-representative training data, including lack of geo-diversity
2. Training labels are biased or faulty
3. Data is affected by adversaries

Issues related to AI models:

1. Model training choices
2. Algorithm learns faulty strategies
3. AI learns to fake something plausible
4. AI model used in inappropriate situations
5. Non-trustworthy AI model deployed
6. Lack of robustness in the AI model

Other issues related to workforce and society:

1. Globally applicable AI approaches may stymie burgeoning efforts in developing countries.
2. Lack of input or consent on data collection and model training
3. Scientists might feel disenfranchised.
4. Increase of CO_2 emissions due to computing

Ex. of Non Representative Data:
"For example, the national radar network has coverage gaps which can inadvertently under-represent some populations, as indicated… "

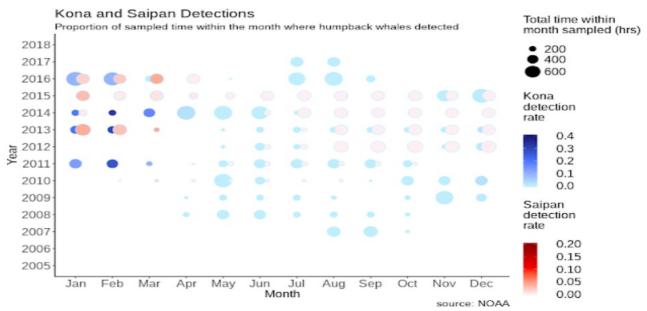
"... we can put AI to good use by developing AI algorithms to estimate sensor values at the missing location"

Ex. of Biased Labeled Data for Training:
If we trained on only human reports → bias toward areas with higher population → "This might bias the AI model to over-predict urban hail/tornadoes and under-predict rural hail/tornadoes."

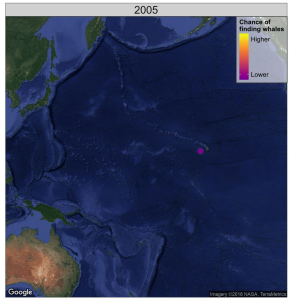
From McGovern et al., 2022, "Why we need to focus on developing ethical, responsible, and trustworthy artificial intelligence approaches for environmental science".

Example: OK Google, Find the humpback whales

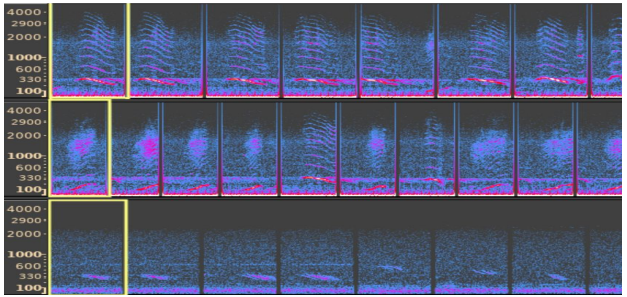
- PIFSC deployed bottom mounted HARPs since 2005
  - 13 Pacific Island sites
  - 200+ TB
- **AI for Social Good**
  - Identify humpbacks
  - Decimated to 9.2 TB
    - 200 kHz -> 10 kHz
  - Classifier for 75 sec segments attains 90% precision at 90% recall for supervised learning
  - Unsupervised learning retrieves similar song units



Time density of presence for Kona and Saipan



Relative call abundance identified by the classifier



Manually chosen calls (boxed) and nearest neighbors

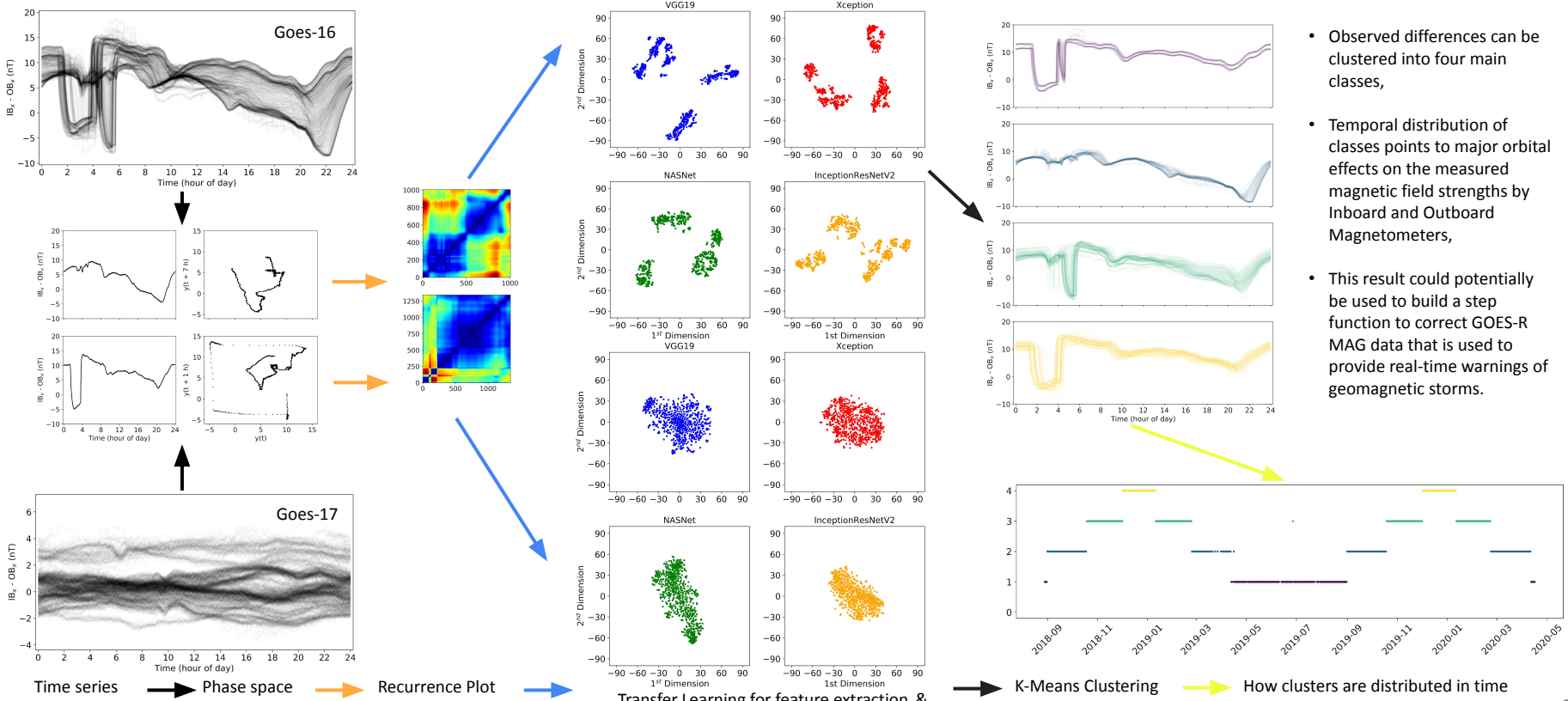
PIFSC: Ann Allen & Karlina Merkens

Google: Matt Harvey, Jiayang Liu, Julie Cattiau, Aren Jansen, Rif Saurous, & Lauren Harrell

Google AI

Advancing AI Research through Exploring Critical Problems

GOES-R Magnetometer Challenges - Classification of Differences



- Observed differences can be clustered into four main classes,
- Temporal distribution of classes points to major orbital effects on the measured magnetic field strengths by Inboard and Outboard Magnetometers,
- This result could potentially be used to build a step function to correct GOES-R MAG data that is used to provide real-time warnings of geomagnetic storms.

Advancing AI Research through Life Saving R2O

Detecting Rip Currents in Coastal Webcam Imagery

National Ocean Service

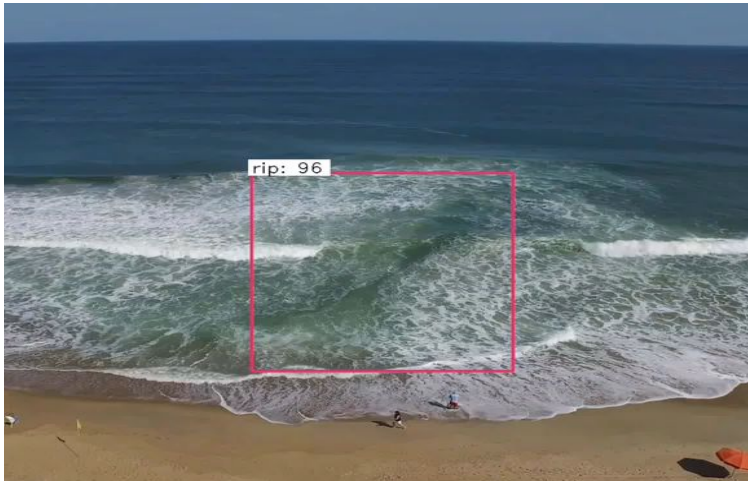
Objective: Detect rip currents in coastal webcam imagery using regional convolutional neural networks

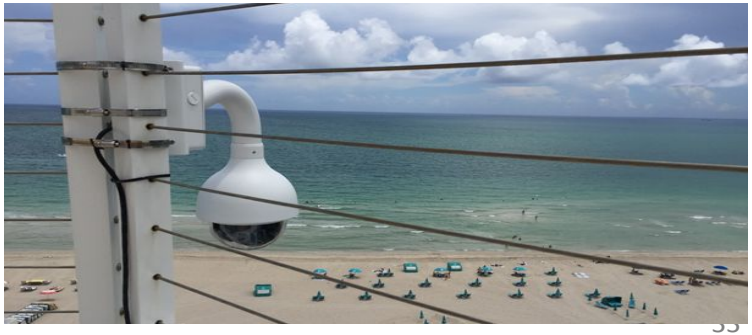
Benefits: The algorithm will be applied to webcam imagery collected by the NOAA-funded WebCOOS camera network in the Southeastern US. The results will support implementation and improvement of the NOAA rip current forecast model

Partnerships: UC Santa Cruz; UNC Wilmington; U of South Carolina; SECOORA; USGS

De Silva, A., I. Mori, G. Dusek, J. Davis and A. Pang (2021), Automated rip current detection with region based convolutional neural networks, Coastal Engineering, https://doi.org/10.1016/j.coastaleng.2021.103859

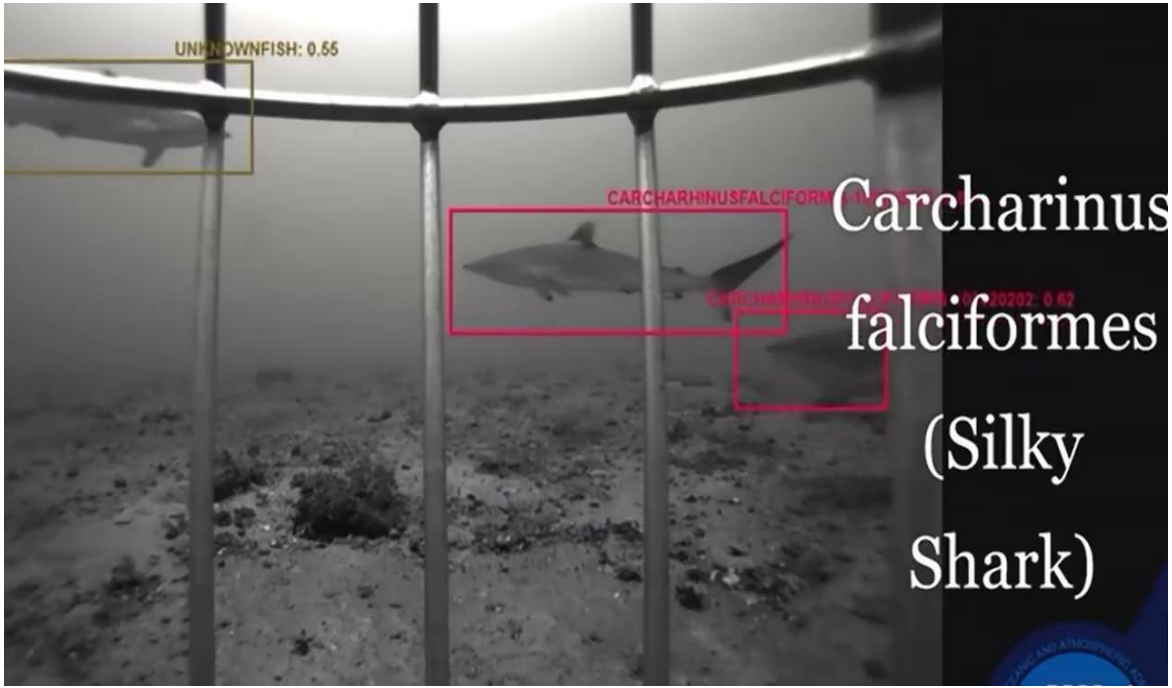
Principal investigators: Gregory Dusek (gregory.dusek@noaa.gov), Alex Pang (pang@soe.ucsc.edu)





Current and Potential Value for AI @ NOAA

Marine Life Speciation using Video Image Analytics for the Marine Environment (VIAME)



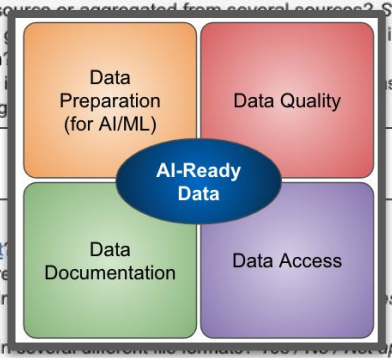
VIAME helps automate the detection and identification of fish species captured by video

Video adapted from: https://videos.fisheries.noaa.gov/detail/videos/science-technology/video/6255809190001/video-image-analytics-for-the-marine-environment

Initiative: AI-Ready Standard Development

Data Quality
- Have measures been taken to ensure completeness? *Yes / No / Not applicable*
- Are there automated processes to monitor consistency? *Yes / No / Not applicable*
- Have measures been taken to reduce bias? *Yes / No / Not applicable*
- What is the timeliness of the data? *Near real-time, 1 week, 1 month, 1 year, more than 1 year*

Data Preparation
- Have null values/gaps been filled? *Yes / No / Not applicable*
- Have outliers been identified? *Yes, tagged / Yes, removed / No / Not applicable*
- Is the data single-source or aggregated from several sources? *Single / Aggregated*
- Has the data been … it in the originally sampled resolution?
- Have targets been … s a training dataset for supervised learning



Data access
- What is the file format?
  - Is it machine-re…
  - Is it available in … s / No / Not applicable
  - Is it available in … applicable
- Data delivery:
  - Direct file download or ordering? Yes / No / Not applicable

Data documentation
- Does the dataset have metadata? *Yes / No / Not applicable*
  - Is the dataset metadata standardized? *Yes / No / Not applicable*
  - Is the dataset metadata machine-readable? *Yes / No / Not applicable*
  - Does it include details on the spatial and temporal extent? *Yes / No / Not applicable*
- Is there a comprehensive data dictionary/codebook to describe parameters? *Yes / No / Not applicable*
  - Is the data dictionary standardized? *Yes / No / Not applicable*
  - Is the data dictionary machine-readable? *Yes / No / Not applicable*

Progress and Future Steps

- Dec 2021: ESIP Community Survey on AI-ready data needs broadly distributed
- Jan 2022: session at ESIP meeting
- Mar 2022: survey results → Draft standard for AI-Ready Open Data
- Summer / Fall 2022
  - Assess sample open datasets (AOP)
  - **ESIP July Session** **(join us / register)**
  - Assign readiness level & develop improvement plans
  - Reality check with key AI/ML data users
  - Use feedback to improve the draft standard
  - Include the standard in at least 1 call for proposals or funded opportunity (AOP)

ESIP Data Readiness Cluster (2022): Checklist to Examine AI-readiness for Open Environmental Datasets. doi.org/10.6084/m9.figshare.19983722.v1

Initiative: AI-Ready Data - Join us in Pittsburgh (or virtually)

Earth Science Information Partners (ESIP)

[July 19-22 Meeting]

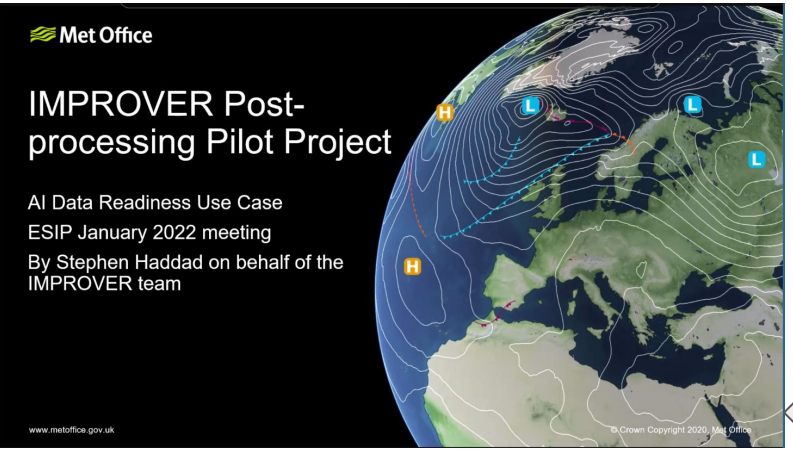
Hands-On Session

Session Title: Enabling AI Application for Climate: Developing A Collection of AI-ready Open Climate Data – Data-A-Thon.

Session Purpose: Initiate a community collaboration on the development of a pilot thematic AI-ready catalog of open climate datasets.

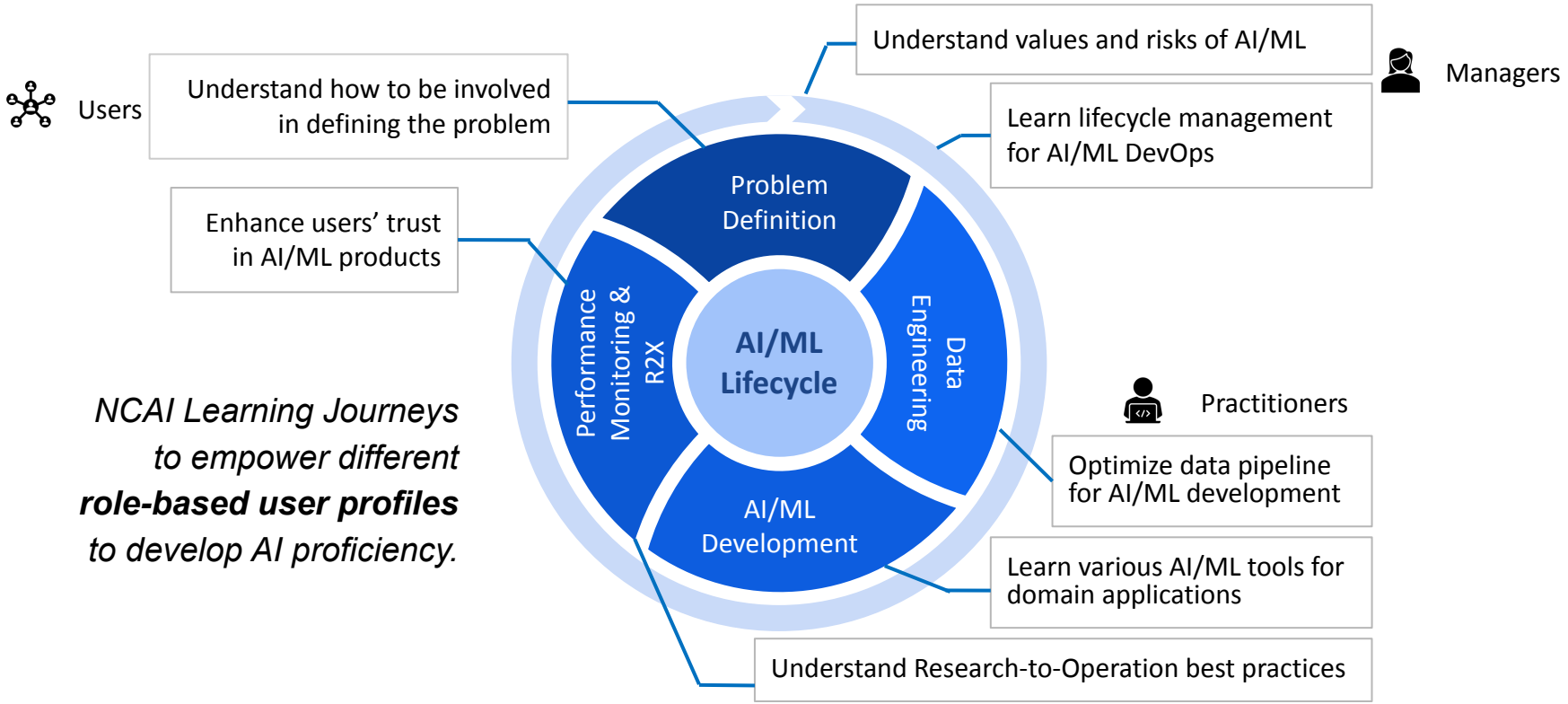
Outcomes/Goals:

1. Build a group of active contributors to develop a pilot thematic AI-ready open climate datasets;
2. Assess the readiness of a selection of open climate datasets for AI applications;
3. Design a catalog framework for representing AI-ready data collections;



Collaboration example from the prior ESIP January Meeting:
AI Data Readiness Use Case from Stephen Haddad (UK Met). Cloud access to Zarr.

Learning Journeys to Empower Diverse Learners



NCAI Learning Journeys to empower different ***role-based user profiles*** *to develop AI proficiency.*

37