Atlantes: A System of GPS Transformers for Global Maritime Behavior

Henry Herzog (Josh Hansen, Yawen Zhang, Patrick Beukema)

Ai2

Global Real-time Modeling of GPS Data at Sea

Core use case

Provide real-time activity classifications of vessel movement patterns to global under-resourced coastal state organizations

Processes ~5B GPS messages per day

- In real time (deployed in Skylight a maritime awareness tool used in over 70 countries)
- 5 T4 GPUs (cost-efficient for self-hosting)
- Production: 28 classifications per second

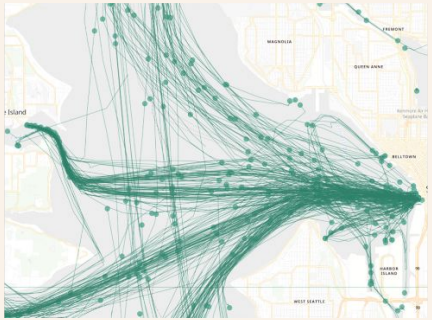
Fishing

Off the west coast of Africa

Anchored

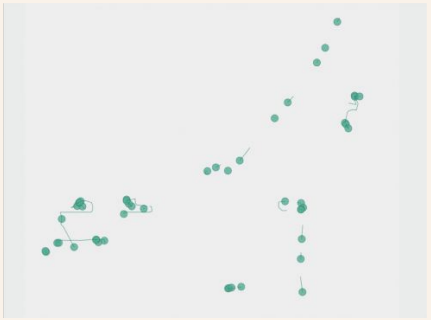
South Pacific, North of Australia

Transiting



In Seattle!

Buoy



In the middle of the Atlantic

Ai2

Outline

1. Description of the task
2. Data
3. Model
4. Evaluation

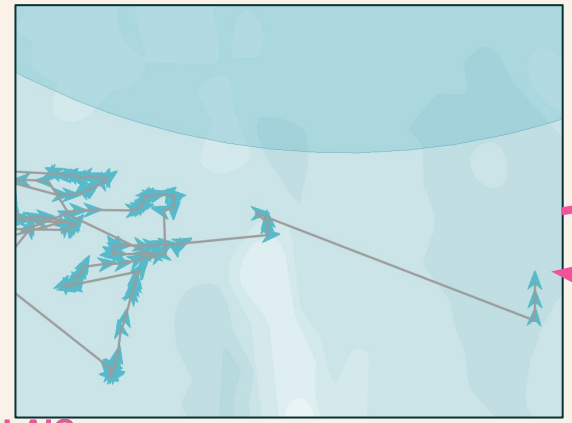
What mattered:

1. High quality dataset
2. Evaluating outputs from a user's perspective
3. ML adapted engineering conventions

Activity Classification Task: Seq2Class

Sequence of N messages where each message contains:

- (Latitude, Longitude)
- Speed over Ground
- Course over Ground
- Timestamp



What activity occurs in the Nth message?

Model

Activity Classification

- Fishing
- Transiting
- Moored
- Anchored

Ai2

Dataset Creation

Annotated over 15M+ GPS messages from 2022-2023.

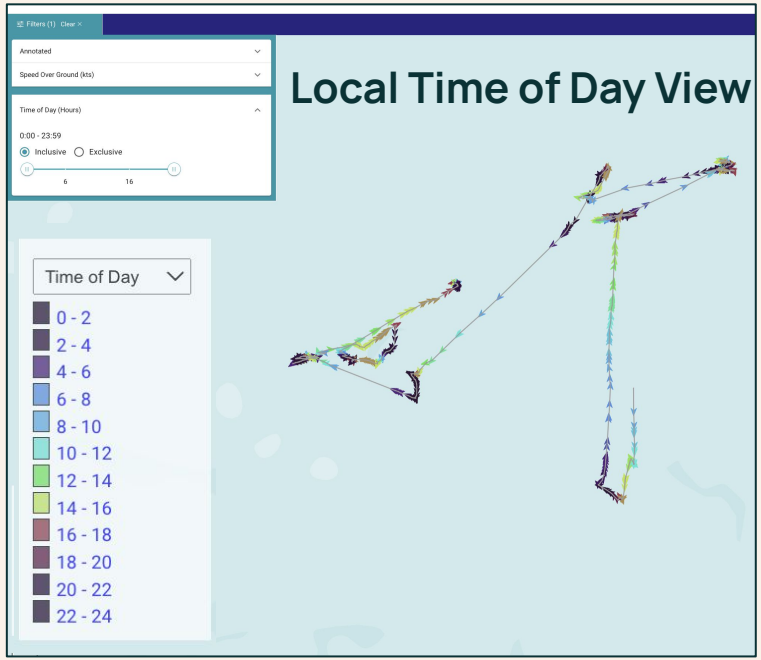
Required Expert Annotators

- Annotators were 20 globally distributed maritime experts

Stratified Sampling + Active Learning

- Across geography, season, and fishing type
- Second phase of samples chosen via model based criteria

Custom GPS Annotation Platform*

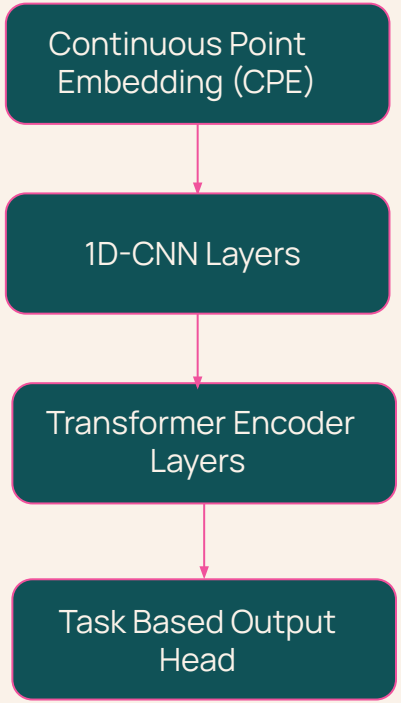


*no prior satisfactory GPS annotation platform

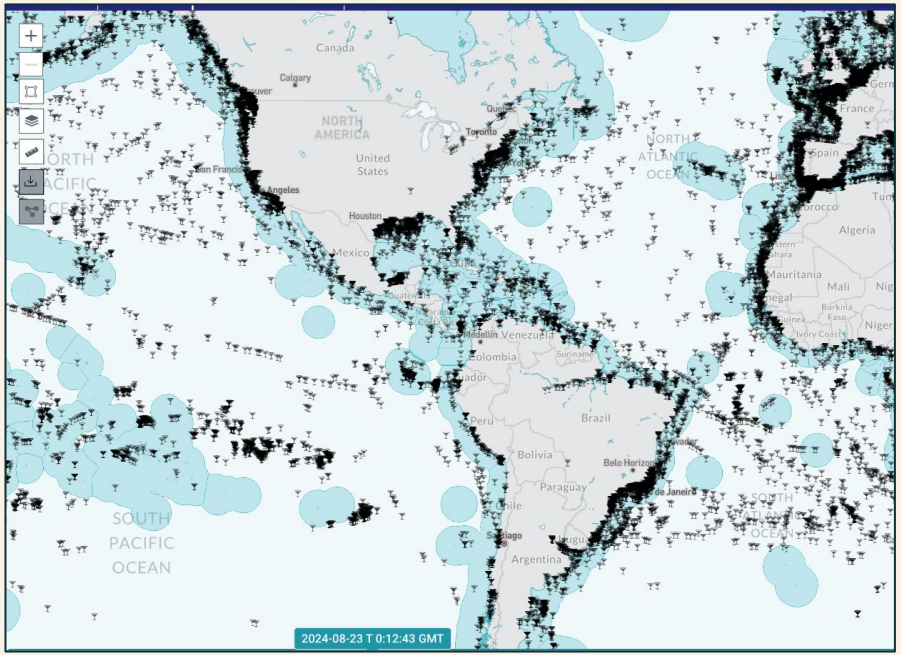
Model Architecture and Training

The architecture was sufficient and not a large focus of our development process

- 4.7 million parameters
- Class weighted cross entropy loss
- Supervised training took 6 hours on 4 H100 gpus with
  - 1.5 million instances (2000 message length, activity label)



Evaluate Outputs from Users Perspective



Ensures alignment to true definition of a “better model”

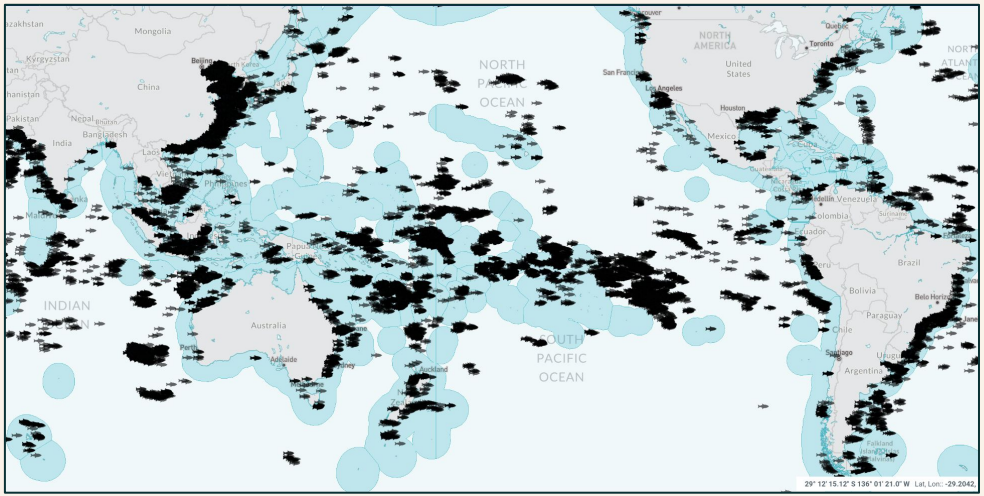
1. Measure outputs at the granularity users care about
   a. fishing events vs message level labels

2. Validate against more diverse false positives

3. Identify and eliminate “obvious” errors
   a. Obvious errors erode user trust

Improved Model Performance

- Only 2 expert evaluation loops were needed for Activity Evaluation where we improved from **44%** accuracy to **71%** accuracy
- Vessel changes behavior → satellite → downlink → inference on our GPUs within minutes

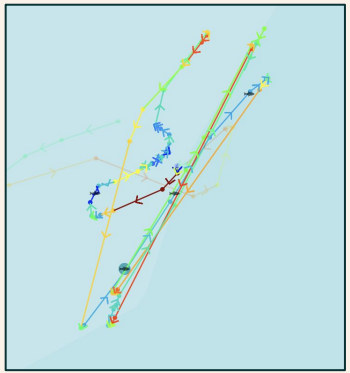


Real Time outputs are far more operationally relevant.

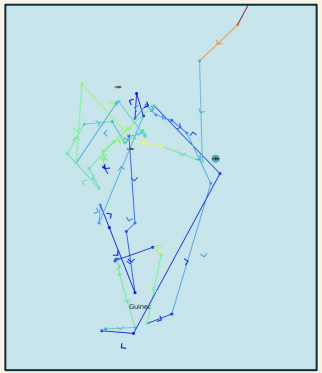
Ai2

Limitations and Areas of Improvement

- Performance is between 70-80% (human is 85%) –
  - Even humans struggle with this task
- Our goal is continual improvement by evaluating our mistakes, listening to feedback, striving to never repeat the same errors
- Examples of false positives from uncommon vessel types doing fishing-like things.



Fire Fighting



Tug Boat

Continuous Machine Learning

Machine Learning Specific Engineering Conventions

- Code for training and inference is identical –
  - There is no “productionisation” step on our team. The code is immediately “productionized” from the beginning of the project.
- Ability to recreate production model outputs from source data
  - Many components of the entire system can affect the outputs
- Ability to gather new targeted datasets from production

Industry Standing Software Engineering Best Practices

- CI/CD with Github Actions
- Extensive Unit, Integration, and Regression Testing
- Static typing with mypy

This was a massive team effort

Please reach out, we welcome all thoughts and feedback.

Code on github

Paper on arxiv

Full (Ai2) talk on youtube

Data Provider:

spire

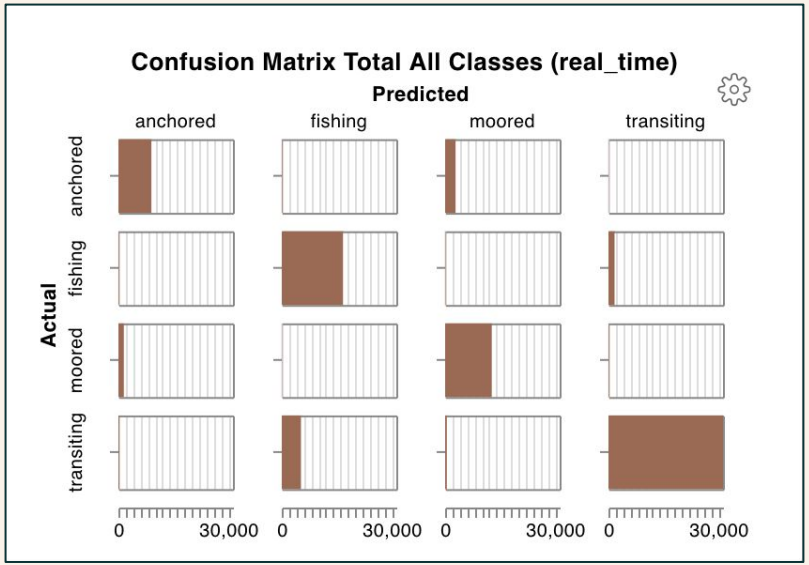
Ai2

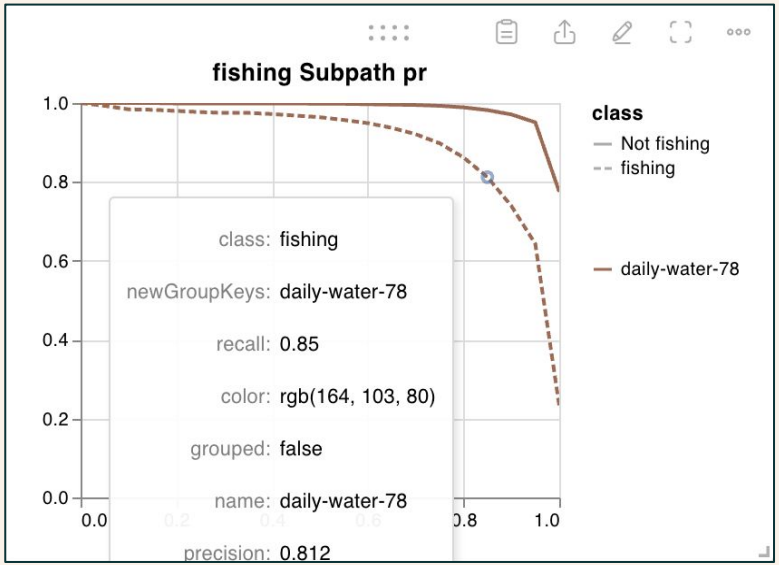
Thank you.

Appendix

Offline Results are Only a Proxy

- ~0.80 F1 on held out Human annotated data → Time for the Real World 🌎 ...





SKYLIGHT
A product of AI2

Add slide on CPD if people ask about it

Future Directions for GPS Modeling

- Regional Fine-tuning for VMS Data (VMS data is another type of GPS sequence)
  - (Indonesian government was just here asking for this)
- Identifying Human-Wildlife Conflict risks with Earth Ranger
- Additional AIS tasks for Skylight such as vessel type classification or one sided rendezvous (OSR)

EarthRanger

SKYLIGHT
A product of AI2

Ablations

Add brief information about that

Why Seq 2 Class?

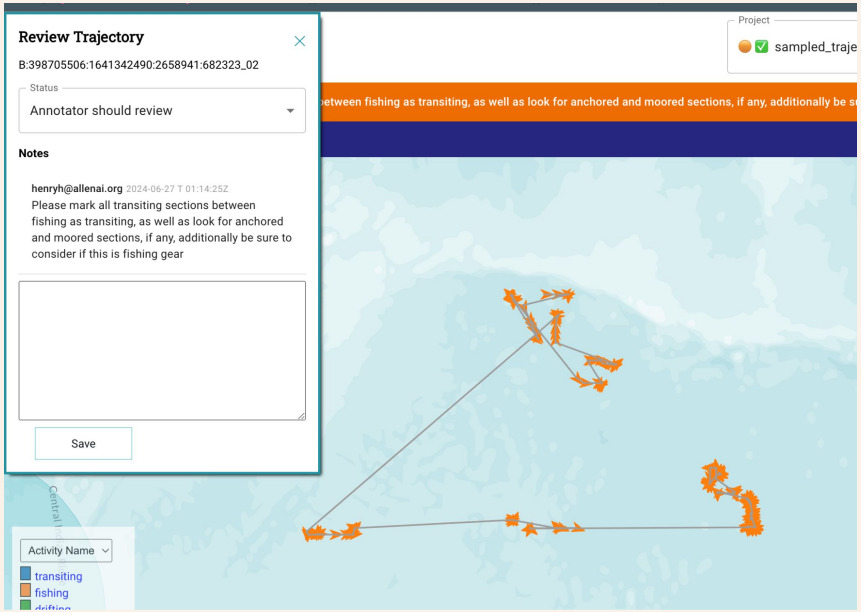
- **Simple Translation between Training and inference**
  - Simple Validation Loop from Expert User Feedback
  - Easier to incorporate Metadata for Post Processing
  - Single Record of History

Annotation Management

For Expert Level Tasks, it is difficult to balance high quality outputs and time to annotate.

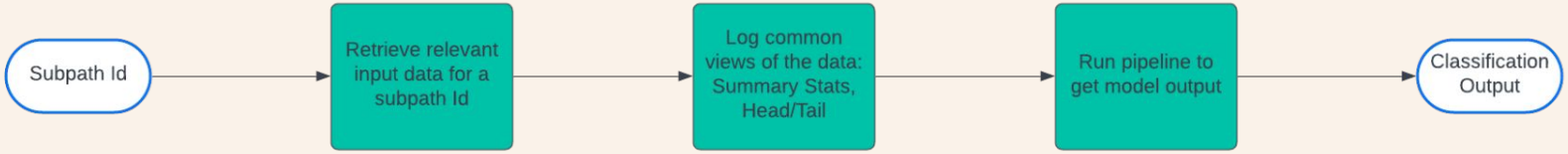
Key Learnings

- Iterative Improvement of Annotation Platform was a key driver of Success
- Regularly Identify Systemic Task Misunderstanding and provide ability to Kick back Annotations to Users



SKYLIGHT
A product of AI2

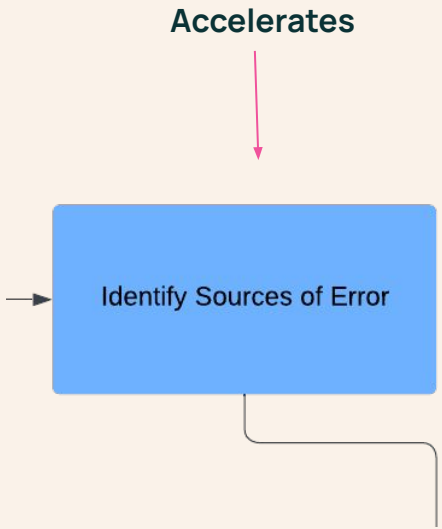
Tool Profile: Debug Endpoint



To determine if a bad output is caused by the model or some other part of the system we need to be able to

- Recover exact inputs for a given model output
- Rerun entire deployment pipeline of preprocessing, model inference and postprocessing offline

Note: this is tricky for AIS because the state of the available data is different at debug time vs inference time do to delayed downlinking

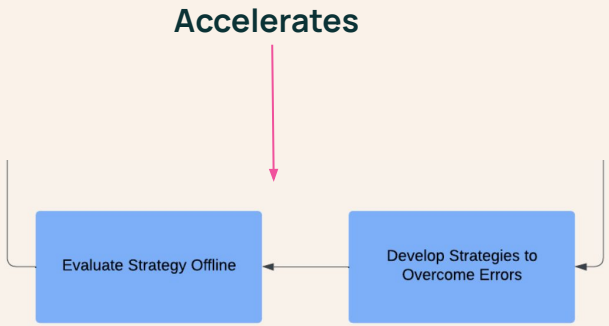


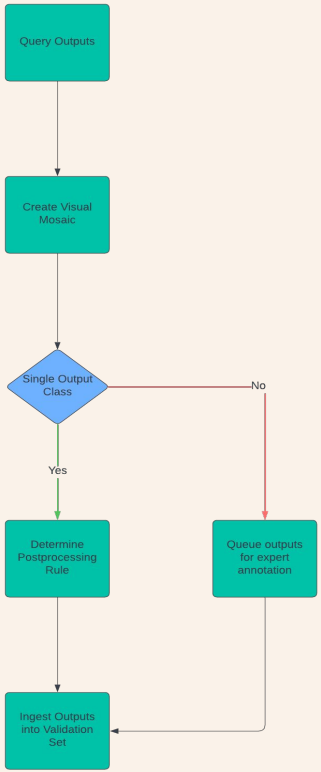
SKYLIGHT
A product of AI2

Tool: Queryable Validation Set Creator

To efficiently gather new validation sets to target specific errors we must be able to:

- Write queries for model outputs based on outputs, associated metadata, input data

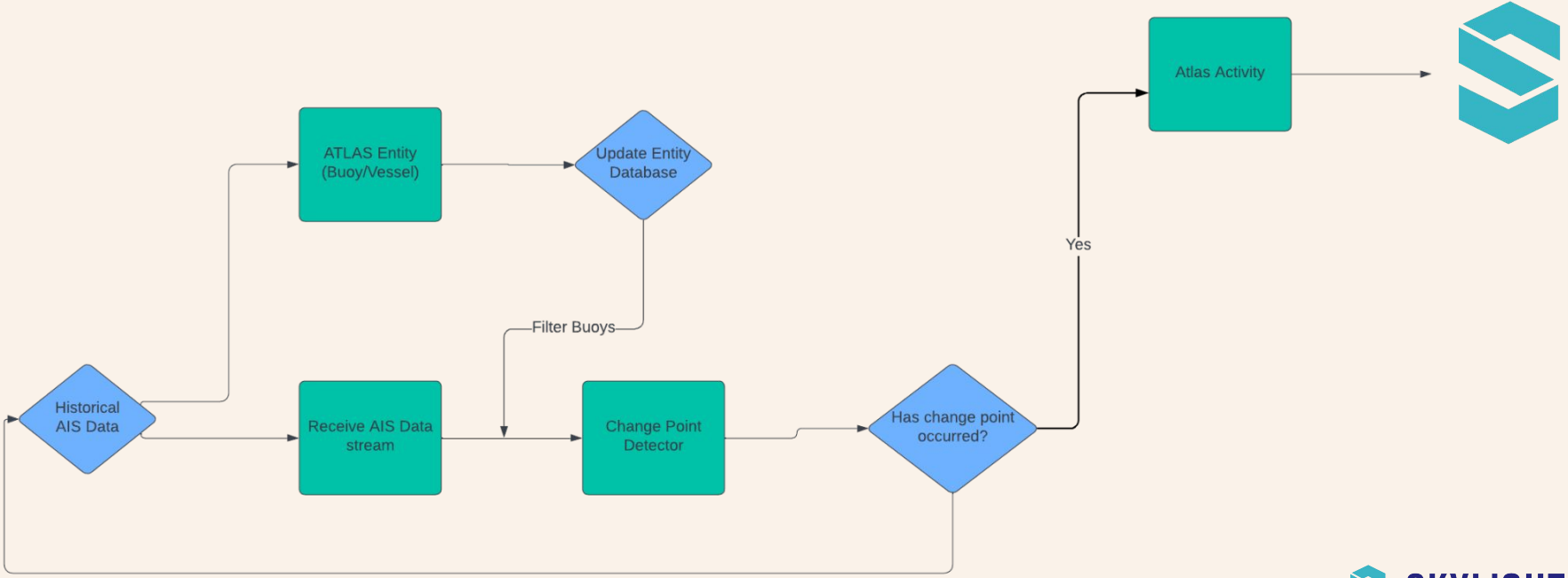




SKYLIGHT
A product of AI2

Where are the bottlenecks?

In the Activity Classification Pipeline!



SKYLIGHT
A product of AI2

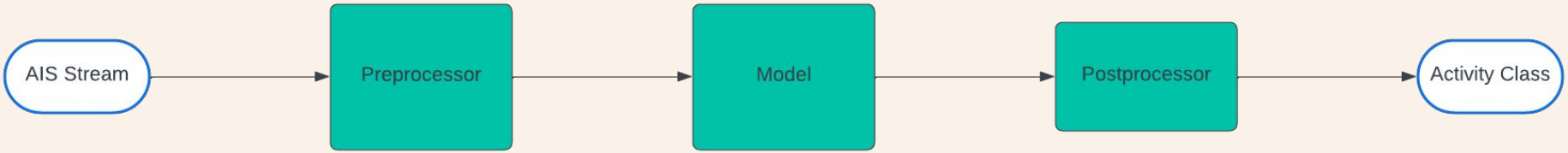
Activity Classification Pipeline

Activity Pipeline has 3 components

Preprocessor: Prepares data for model (data cleaning, normalization, etc)

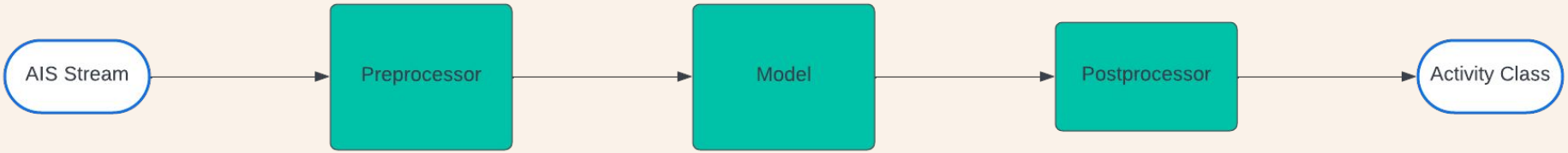
Model: Performs forward pass

Postprocessor: Performs thresholding and metadata based reconciliation to create final output



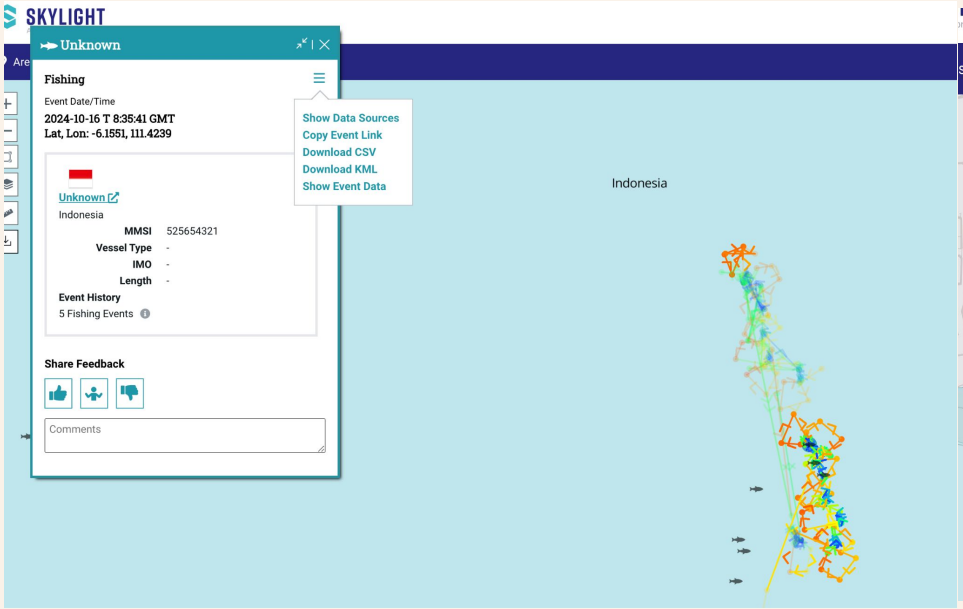
SKYLIGHT
A product of AI2

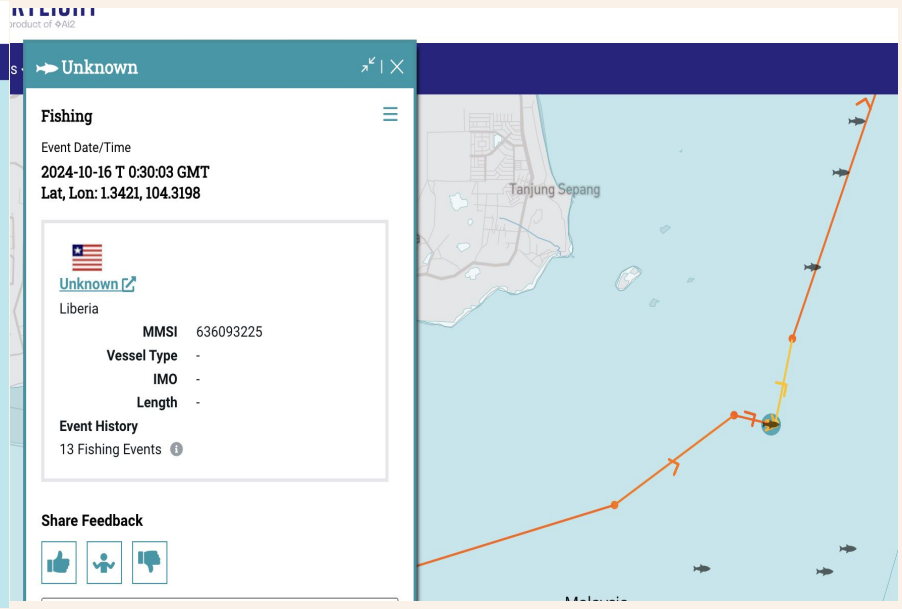
Activity Pipeline Deployment



SKYLIGHT
A product of AI2

User Feedback Enables Continuous Improvement

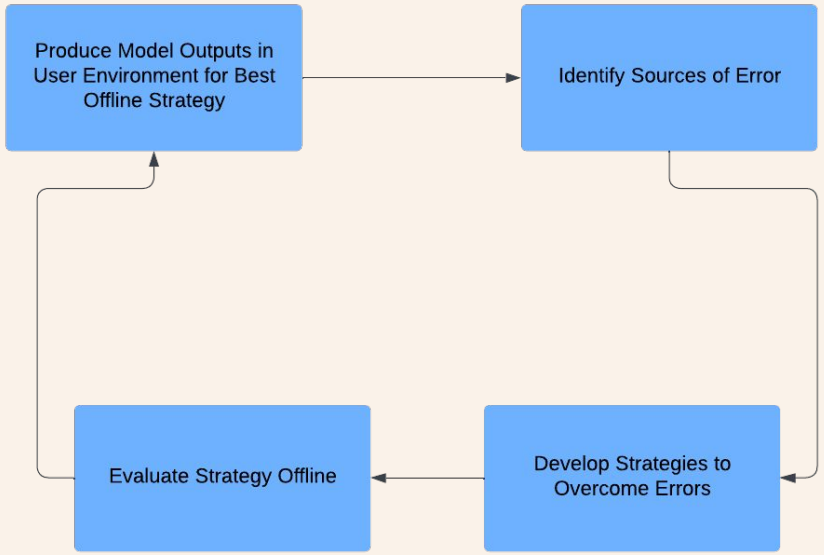




SKYLIGHT
A product of AI2

Why build continuous ML/Engineering?

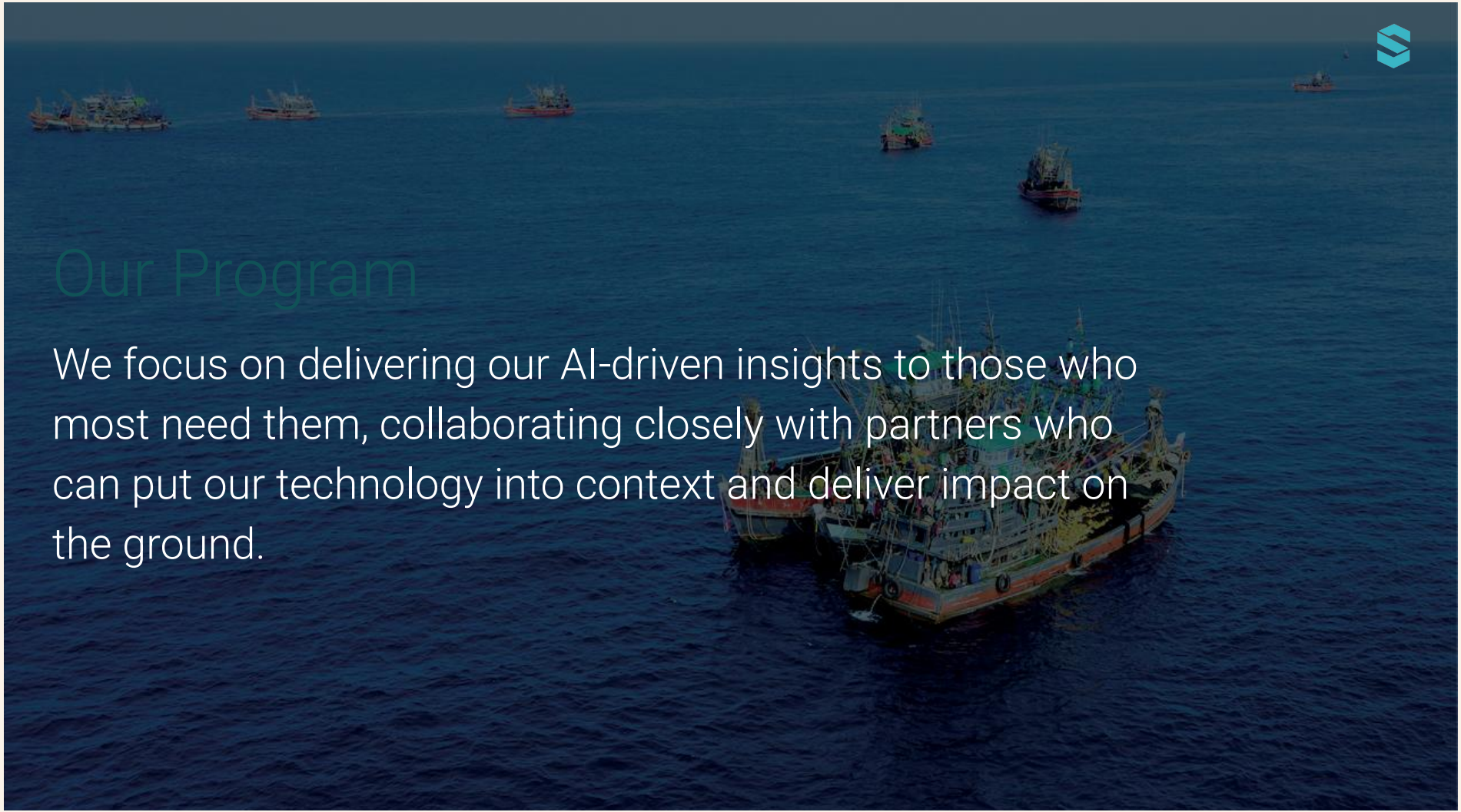
Cycle of Machine Learning Development



Advantages of Continuous ML

- Ensure End to End Alignment with User needs
- Faster Iteration due to early alignment of entire cycle
  - e.g overcome latency bottlenecks with a dummy model before there is a need to see model outputs at scale

SKYLIGHT
A product of AI2

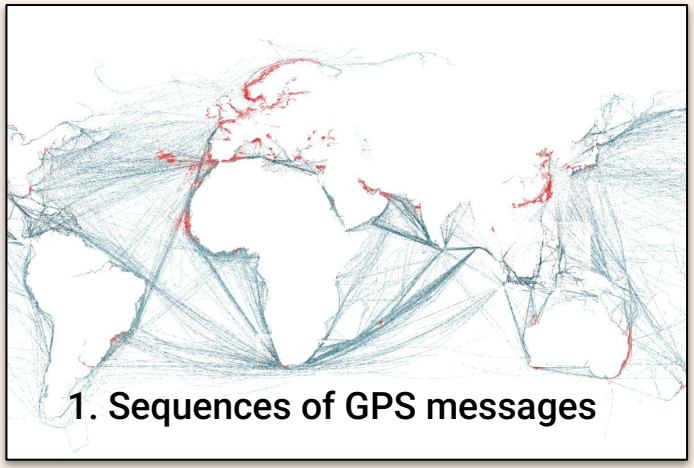


Our Program

We focus on delivering our AI-driven insights to those who most need them, collaborating closely with partners who can put our technology into context and deliver impact on the ground.

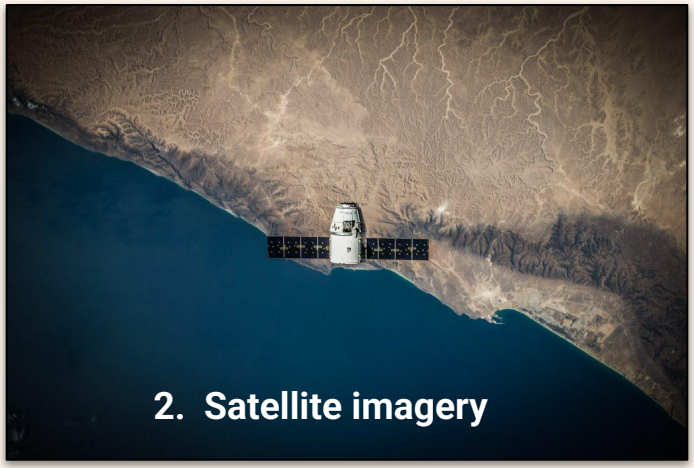
How Does Skylight Use ML?

What behavior occurs?



GPS Behavior Classification

What Vessels are present?

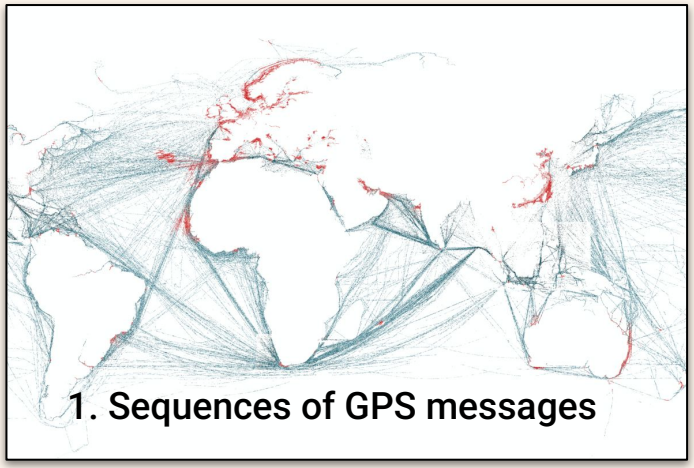


Vessel Detection

SKYLIGHT
A product of AI2

Why is this information important?

What behavior occurs?

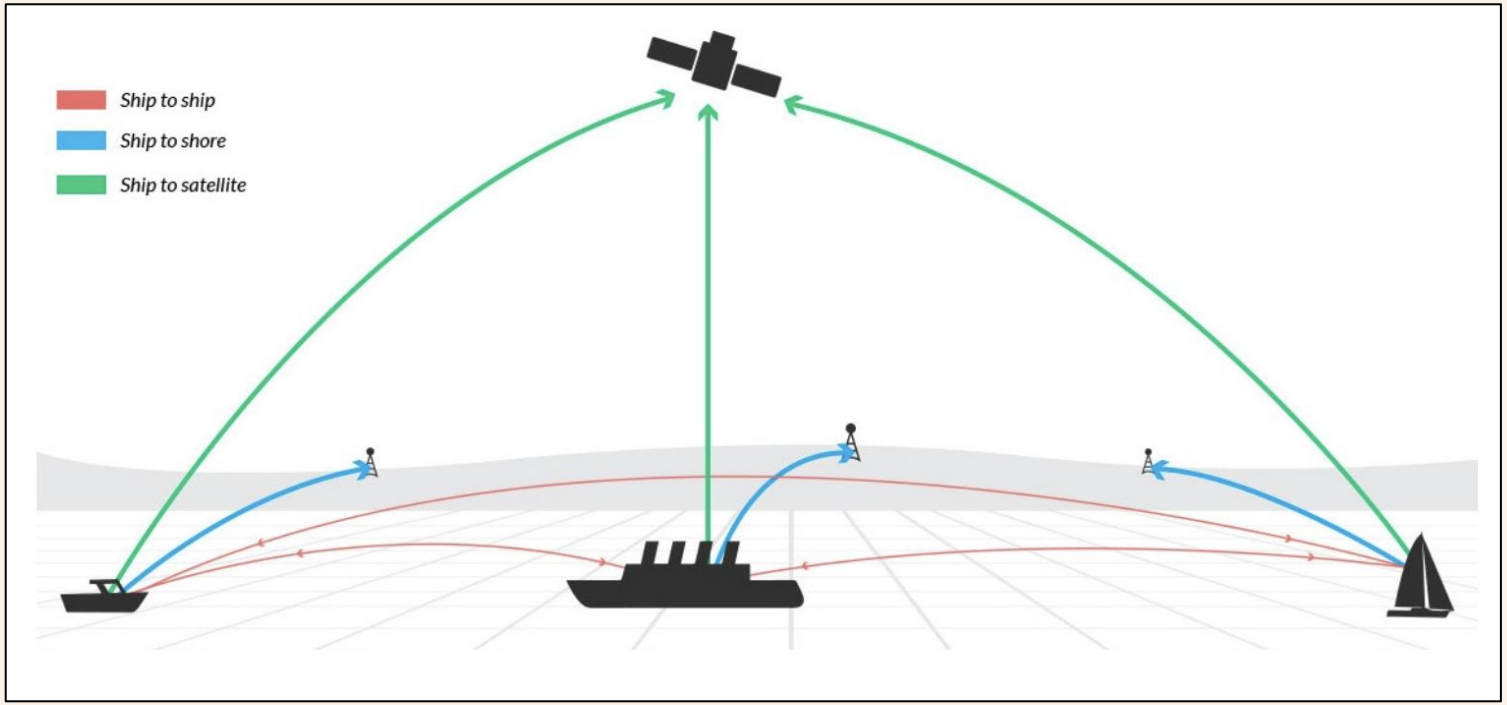


GPS Behavior Classification

SKYLIGHT
A product of AI2

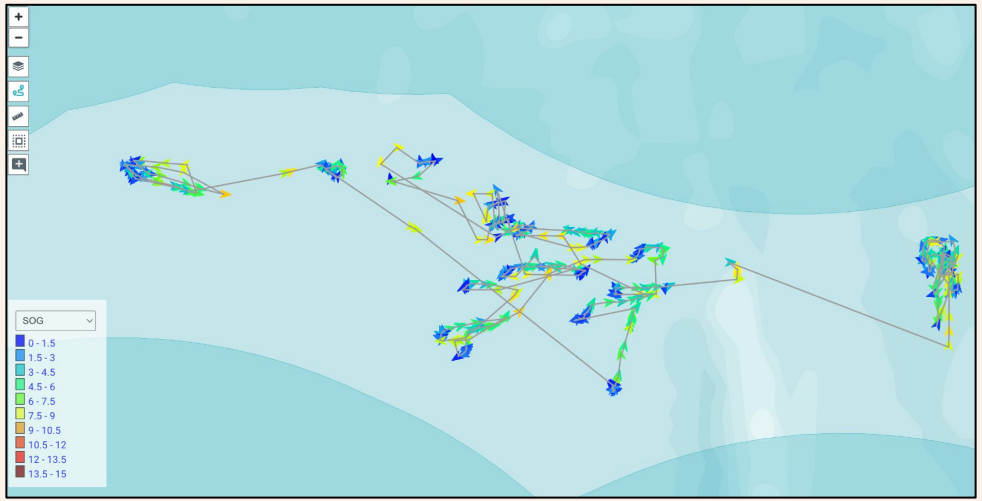
What Is AIS?

- Automatic Identification System (AIS) is a GPS system at sea



SKYLIGHT
A product of AI2

What Makes AIS Data Unique?

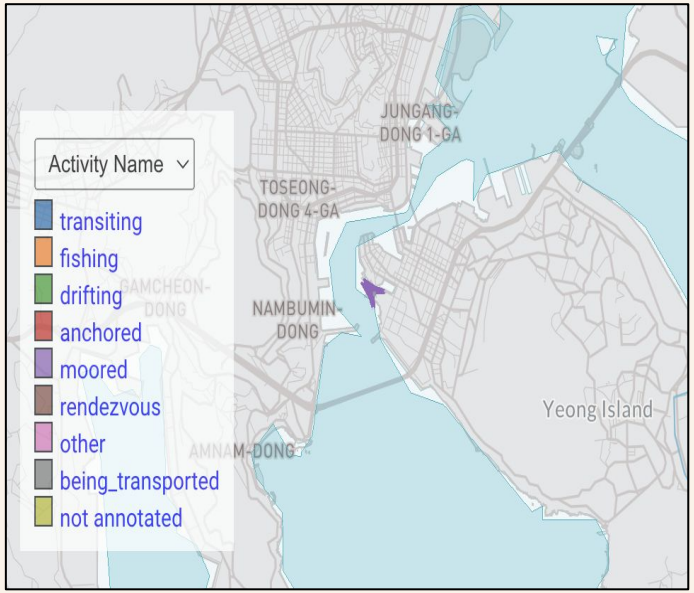


- Irregular Transmissions
  - AIS messages are broadcasted at irregular spatial and temporal frequencies, with gaps of varying lengths between transmissions.
- Metadata Availability
  - AIS messages do not always contain metadata indicating attributes like vessel type or vessel names.

SKYLIGHT
A product of AI2

What Does Vessel Behavior Look Like?

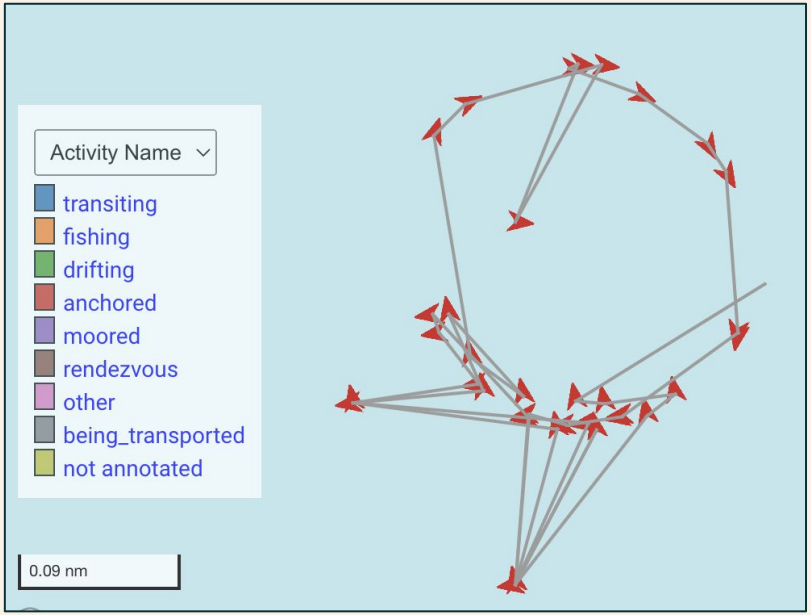
Moored activity occurs when a vessel stays in place by tying up to a dock or another boat.



SKYLIGHT
A product of AI2

What Does Vessel Behavior Look Like?

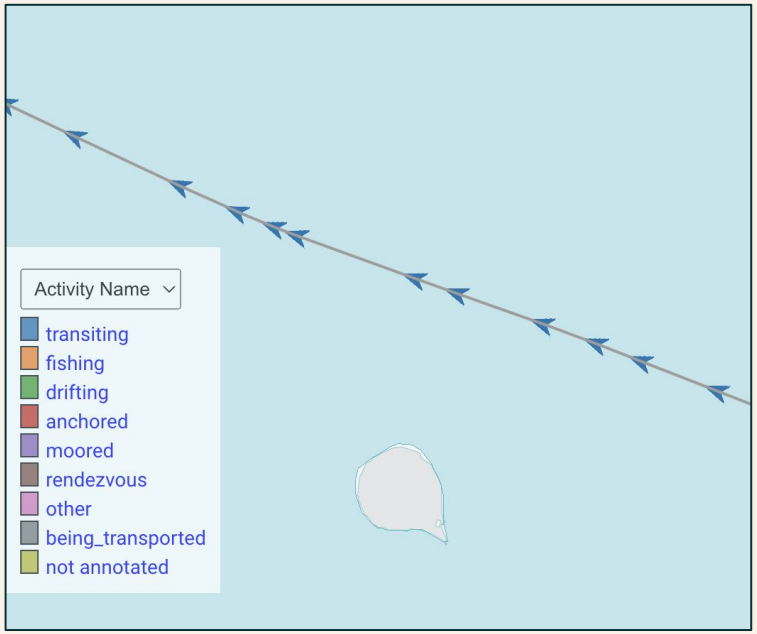
Anchored activity occurs when a vessel drops anchor and drifts around the anchor in a circular pattern.



SKYLIGHT
A product of AI2

What Does Vessel Behavior Look Like?

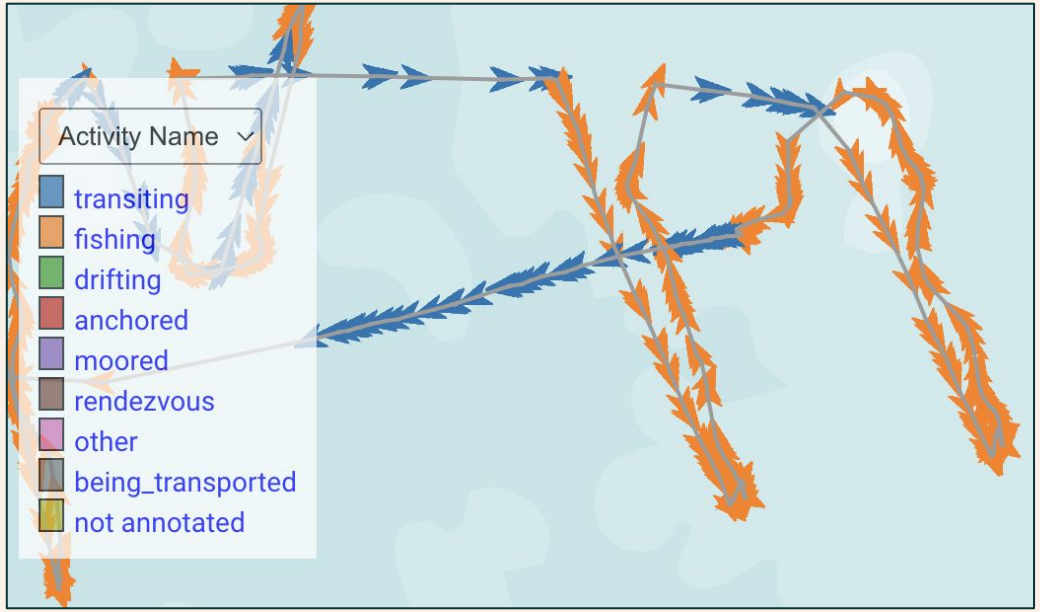
Transiting activity occurs when a vessel is going straight.



SKYLIGHT
A product of AI2

What Does Vessel Behavior Look Like?

Fishing activity occurs when a vessel doubles back on its path.

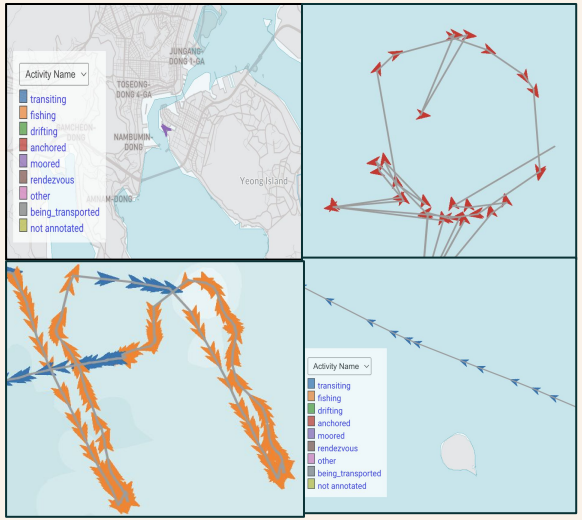


SKYLIGHT
A product of AI2

So AIS Modeling Is Really Easy…

It seems like vessel are either:

- Stopped
- Going in a circle
- Going straight
- Fishing



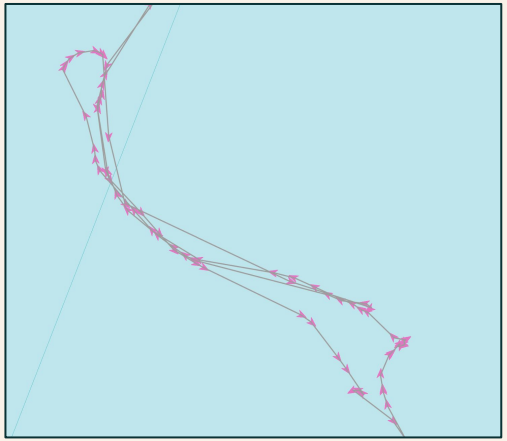
But does this match the real world??????

SKYLIGHT
A product of AI2

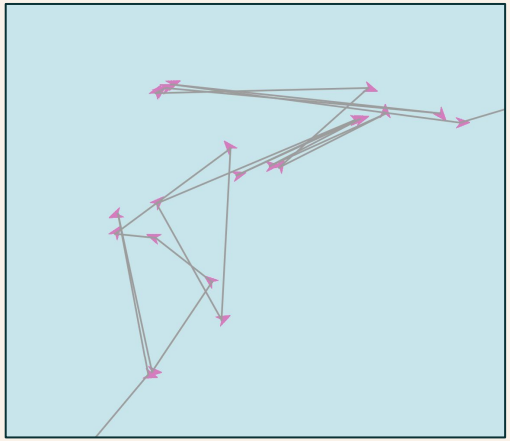
AIS Behavior is Incredibly Complex

- Vessels can move in all sorts of ways for all sorts of reasons, and the AIS we actually get can be sparse.

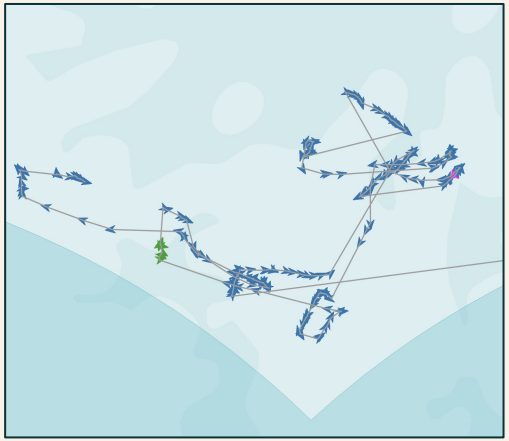
Transiting Behavior



Anchored Behavior



Transiting Behavior



SKYLIGHT
A product of AI2

Global View of GPS at Sea

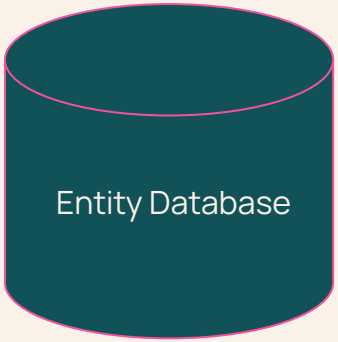
20+B GPS messages/year, 5+TB of data. 50 M+ messages/day

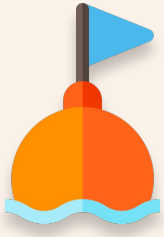
AI2/skylight (2023)

What Do Skylight Users Want From AIS?

What Entity is transmitting the gps messages?

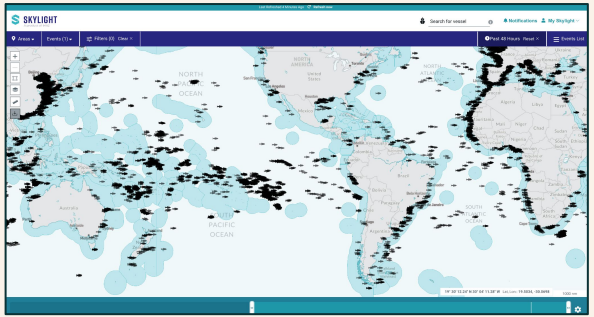
Goal: Stateful understanding of if an AIS transmission comes from a buoy or a vessel





What Activity is a vessel doing at a given point in time?

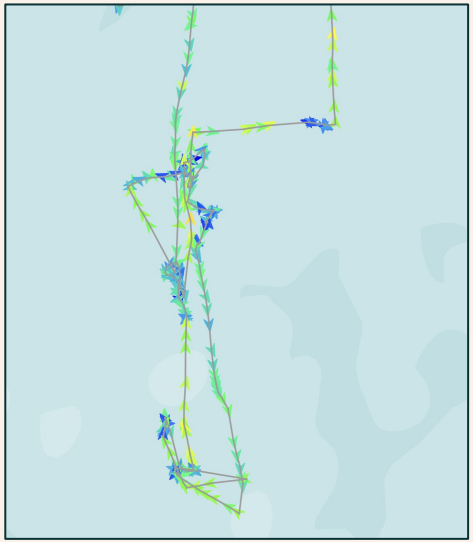
Goal: Provide Real Time Fishing Events for Skylight Users



SKYLIGHT
A product of AI2

Entity Classification Task

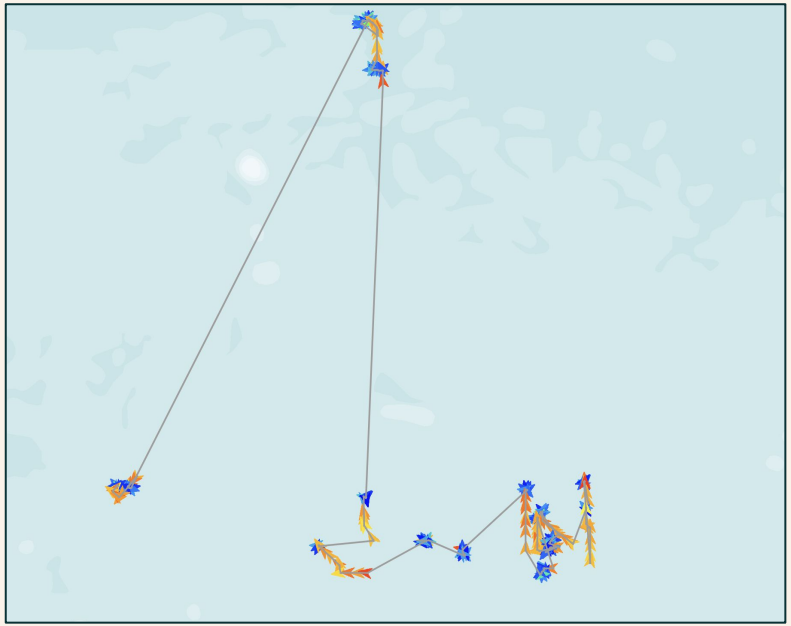
- Given a Historical Sequence of N messages, predict if the messages came from a buoy or a vessel.



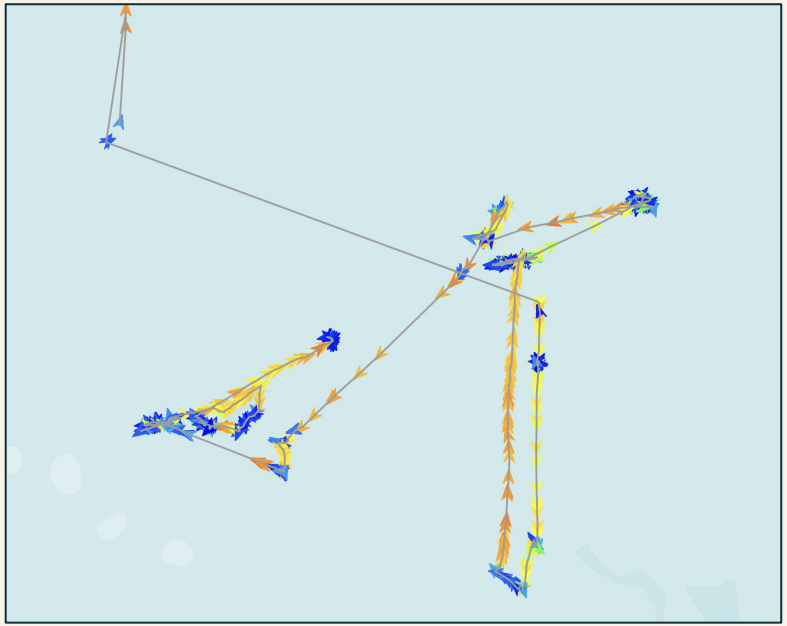
SKYLIGHT
A product of AI2

What Trajectory Is From a Buoy?

Buoy



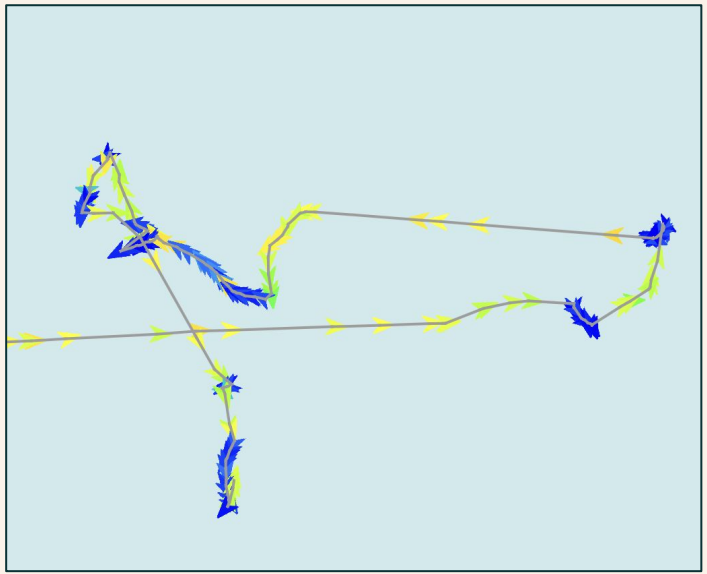
Vessel



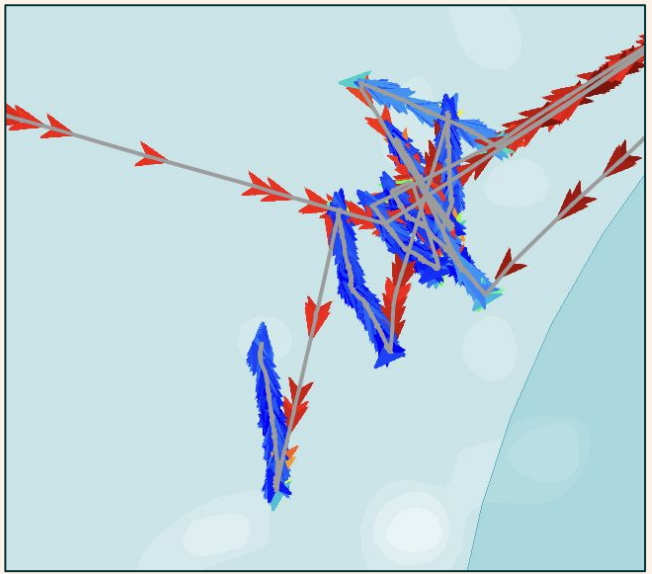
SKYLIGHT
A product of AI2

Which Activity Is Fishing Behavior?

Fishing



Not Fishing

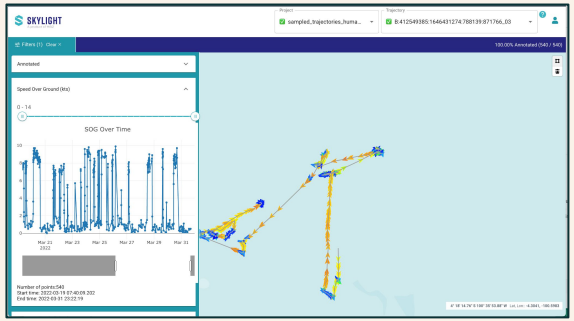


SKYLIGHT
A product of AI2

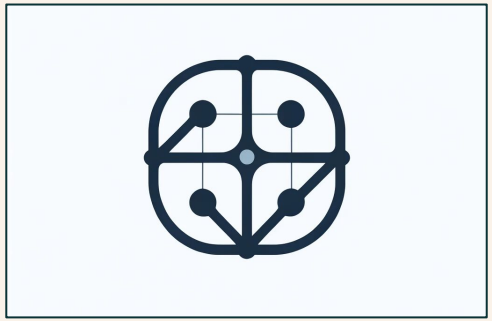
Introducing Atlantes

Atlantes combines ATLAS Entity and ATLAS Activity into a system for AIS based insights.

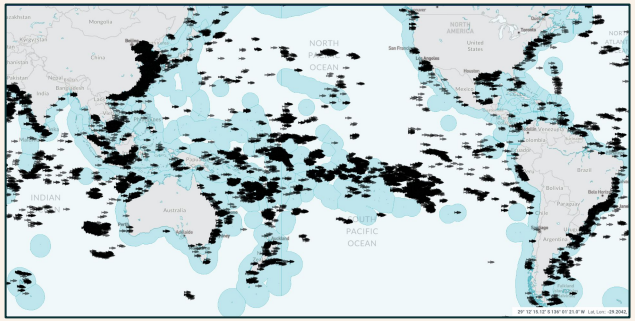
1. Building AIS Training Datasets



2. Modeling AIS Trajectories



3. Optimizing for High Performance



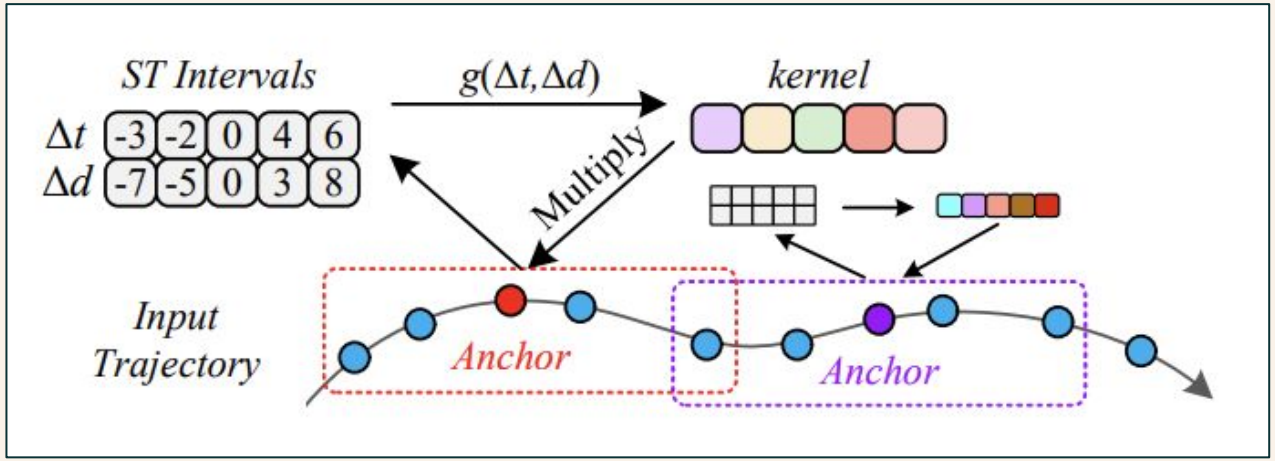
ATLAS: AIS Transformer Learned for Active Subpaths (Architecture)

SKYLIGHT
A product of AI2

Modeling AIS Data

Continuous Point Embedding

- Learn a continuous representation of the space time deltas around every message.



(Liang et. al, 2022)

SKYLIGHT
A product of AI2

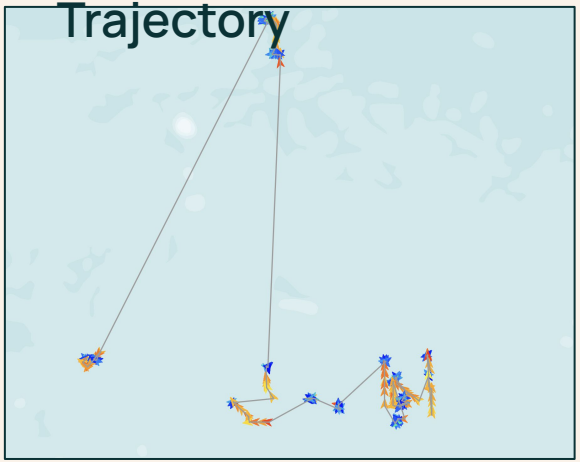
Building AIS Training Datasets

Building Entity Dataset

Entity Labels are constructed from available static metadata

- ~300K Buoy labels
- ~ 10M Vessel Labels

Historical Trajectory



Class Label

0

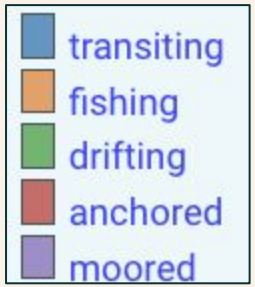
1

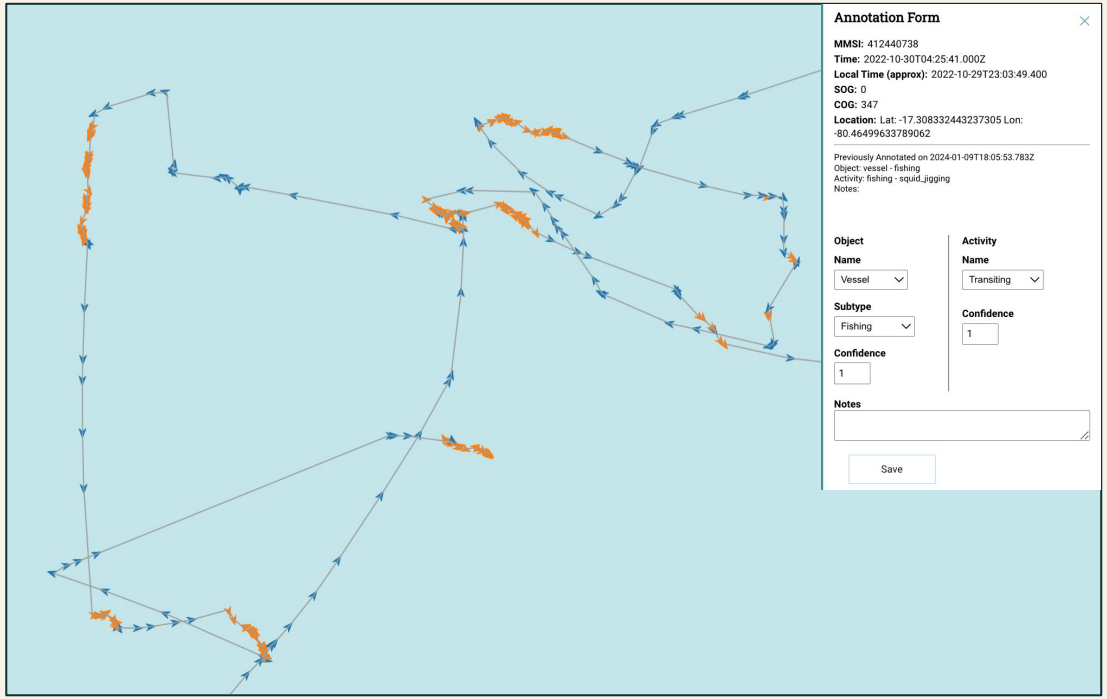
SKYLIGHT
A product of AI2

Message Labels for Real Time Inference

Message labels allow the model to learn the difference between:

- Is fishing right now?
- Was fishing in the recent past?

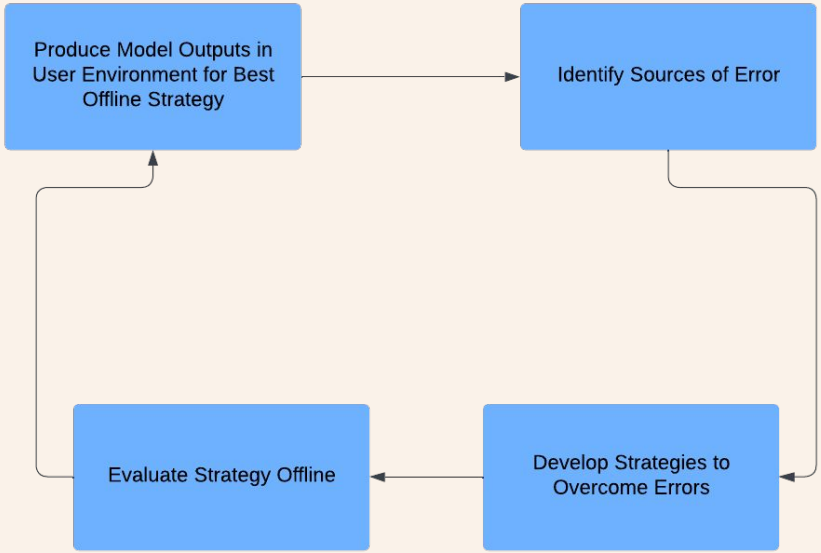




Single months of data were annotated containing about ~1k-6k

SKYLIGHT
A product of AI2

Challenges for Effective Iteration



Key Points of Friction

- Determining if a bad output is due to the model or a bug in another part of the system (e.g data pipeline)
- Gathering and constructing targeted Validation Sets
- Partial Regressions
- Annotator time to build new validation set
- Getting new model into production

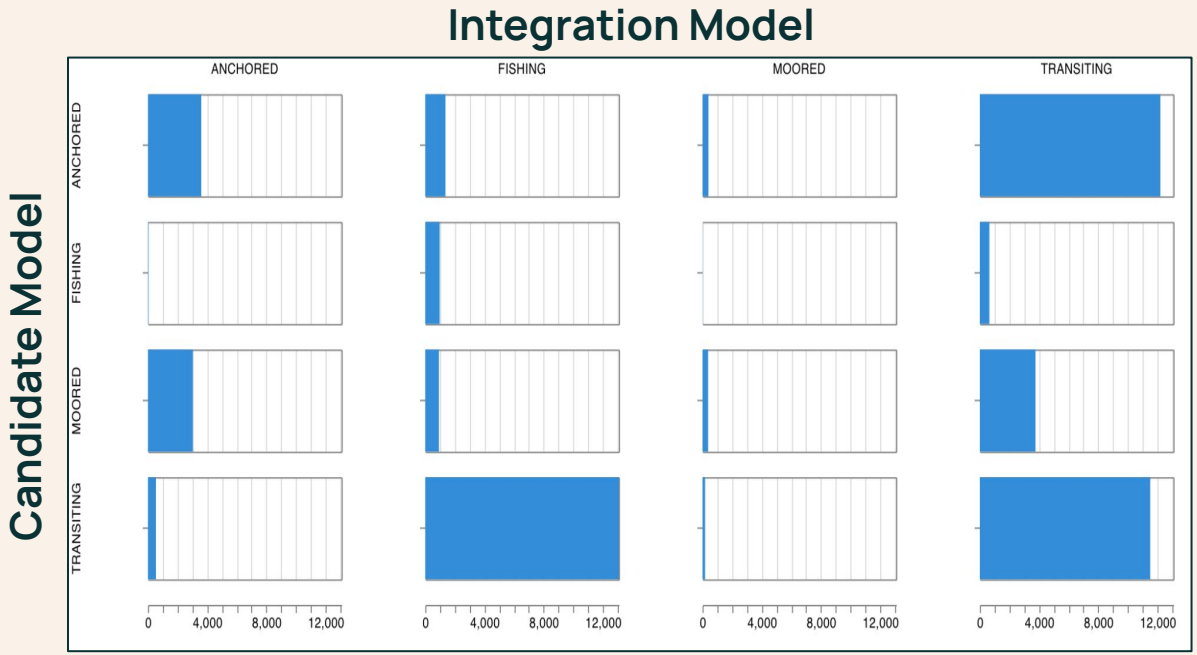
How Do We Iterate Quickly?

We ensure fast iteration by focusing on:

1. **Building Tools!**
   a. **Debug Endpoint**
      i. Determining if a bad output is due to the model or a bug in another part of the system
   b. **Queryable Validation Set Creator**
      i. Gathering and Constructing Targeted Validation Sets
   c. **Distribution Delta Analyzer**
      i. Preventing Partial Regressions
      ii. Annotator time to build new Validation Set
2. **Machine Learning Specific CI/CD**
   a. **Guard against Regressions**
   b. **Easily ship new models into Integration/Staging Environment**

Tool: Distribution Delta Analyzer

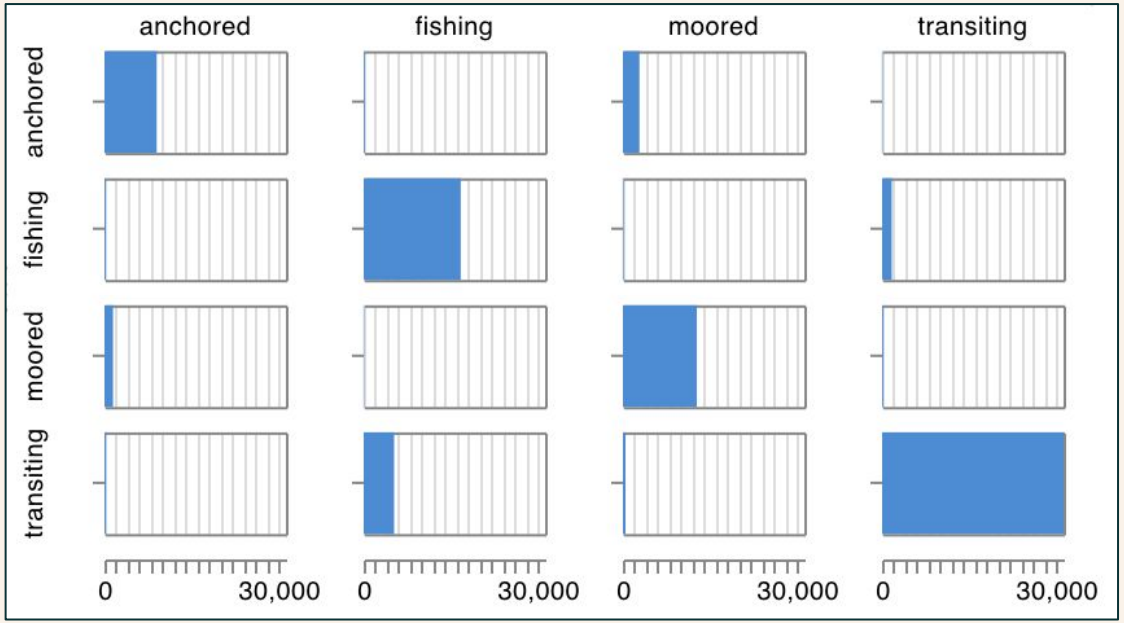
Runs direct comparisons between outputs of a new candidate model and the outputs of the integration model over a chosen time period



Example where offline change would have drastically reduced global recall despite promising validation results.

SKYLIGHT
A product of AI2

Tool: Distribution Delta Analyzer



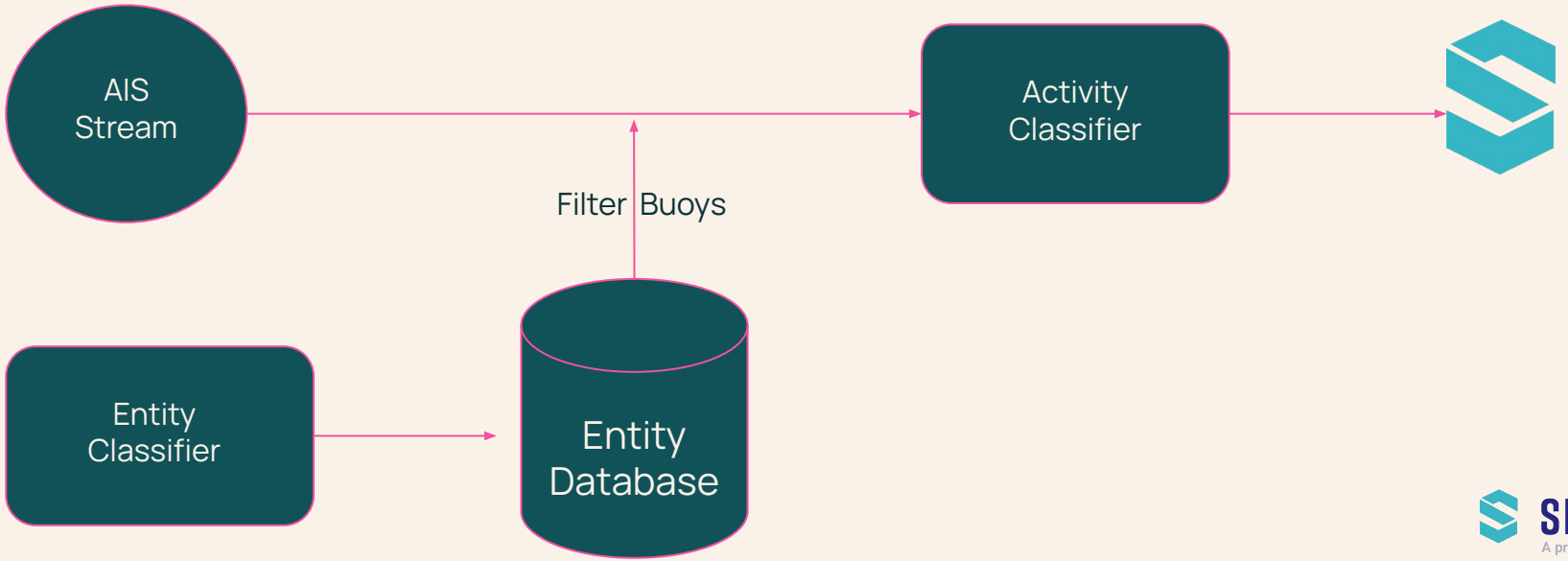
If there is a small difference between the two models, we can sample from the disagreement to get another point of validation before shipping a model to integration

Example where delta is quite small so we only need to do a small validation on the deltas to confirm offline performance.

SKYLIGHT
A product of AI2

Atlantes Architecture

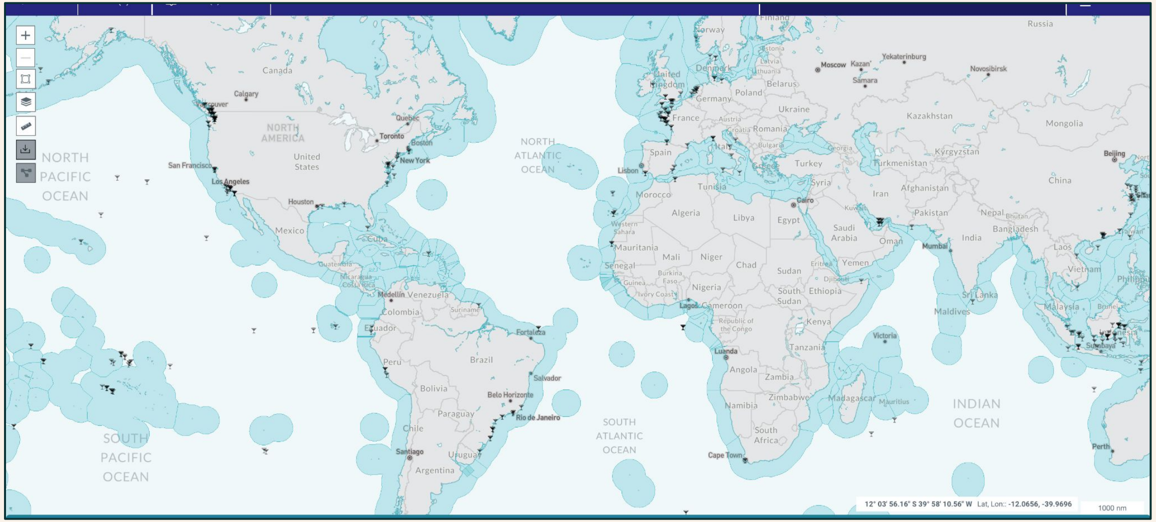
- Processes ~5B AIS messages per day in real time,
  - Context: 2K messages
  - Production: 28 classifications per second



SKYLIGHT
A product of AI2

Get Model Into Real World ASAP!

Early Outputs in Skylight Integration (staging environment)

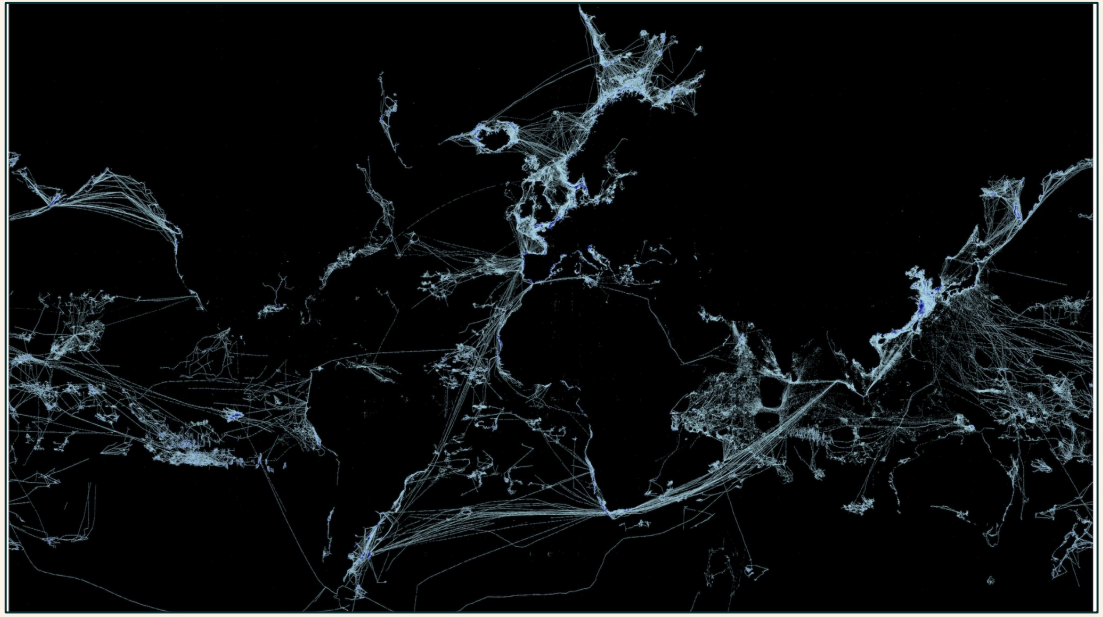


SKYLIGHT
A product of AI2

1. This talk is about how we built a system to provide global real-time activity classification to our users from gps sequences of vessels movements.

Sampling Strategy: Phase 1

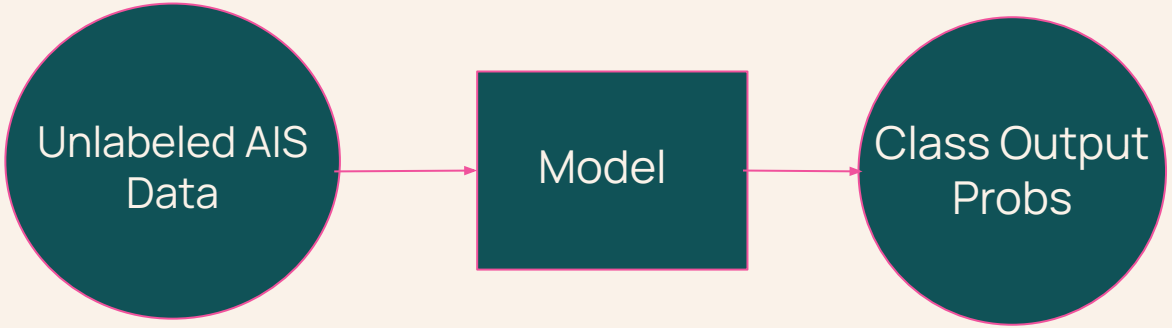
- Created global high resolution monthly AIS maps to have experts identify fishing areas to sample from.
- 85% Stratified Monthly Sample per Area + 15% Outside these Areas.
- These areas effectively covered gear type diversity as well.



SKYLIGHT
A product of AI2

Sampling Strategy: Phase 2

- After training on first 4k track months, we applied a model based sampling approach across the predefined strata (location, time of year).



1. Select examples where fishing is top class but class probability is low.
2. Select examples where fishing is second class and this probability is high.

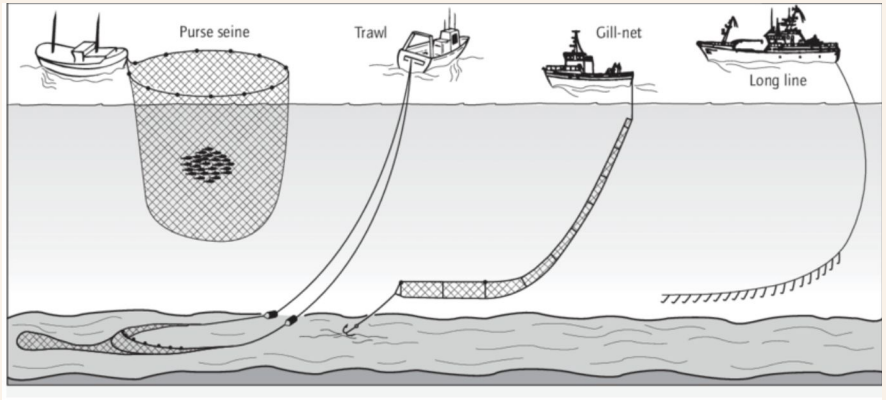
SKYLIGHT
A product of AI2

What needs to be in the dataset?

- Stratified sampling across 3 different stratas + active learning
  - Each sample contains a month of GPS data for a single vessel (1 track month)

Location

Fishing Gear Types



Time

Building a High Quality Dataset

There was no prior global scale vessel activity dataset so we had to build it.

Key challenges:

- No satisfactory GPS annotation platform → Custom annotation interface
- Annotation requires expert analysis → 20 globally based maritime experts

Previous Annotation Strategy 😨😨😨

(OS Geospatial Visualization Software)

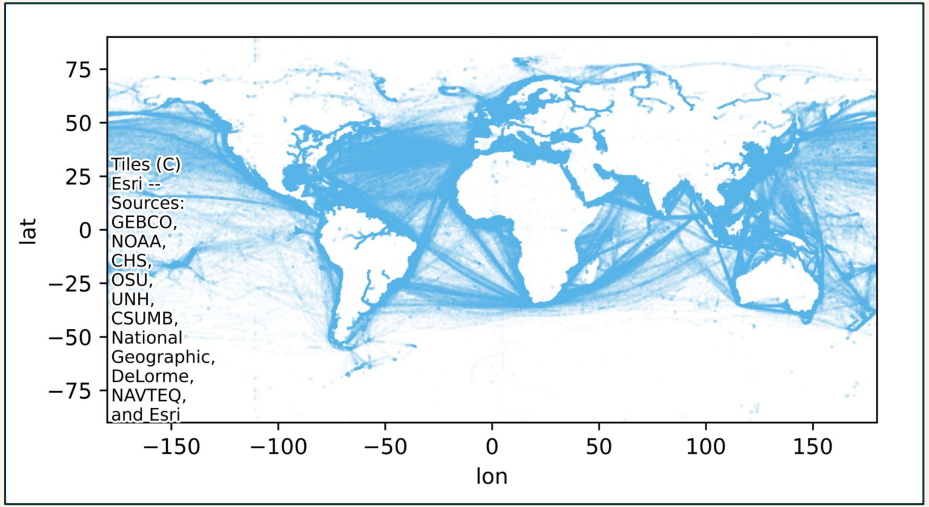
Ai2

Final Dataset

In total, we annotated over 7.5K track months of GPS data across both 2022 and 2023, totaling over 15M+ messages.

Label Counts

- Fishing: 558,347
- Transiting: 502,490
- Moored: 123,696
- Anchored: 105,299

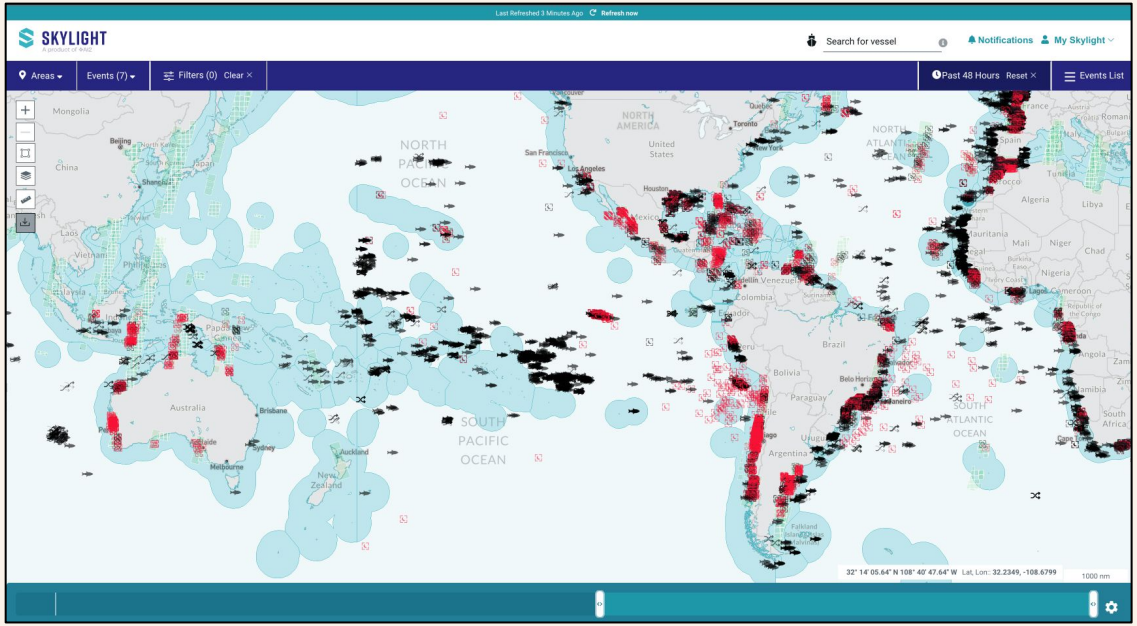


Note: Labels were downsampled to remove obvious redundancies i.e. messages close in time doing the same activity

Ai2

Atlantes is Deployed in Skylight

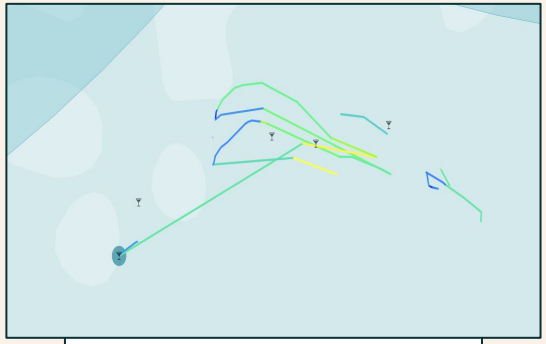
- Skylight is a platform for Maritime Intelligence that uses AI systems for global monitoring of fishing (ecosystems



Ai2

Identifying Sources of Error

1. Expert Event Based Annotation



Model Name

ATLAS-Activity-Real-Time_no_git_hash_2024-08-02-00-47-39_epoch3.pt

Model Version

Confidence

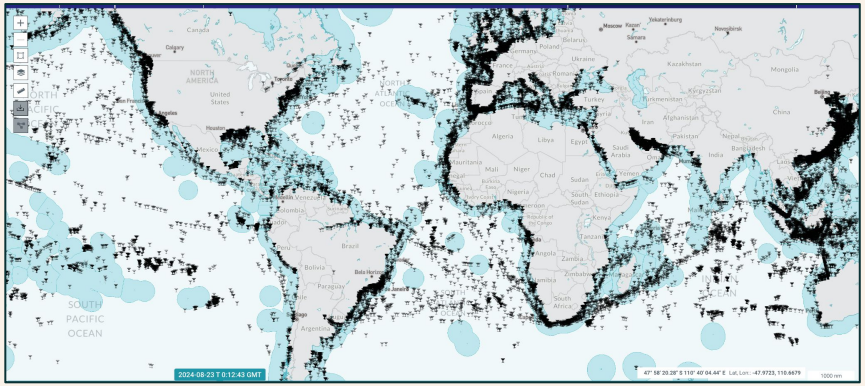
0.39264747500419617

Predictions

Share Feedback

Comments

2. Emergent patterns from global view
   a. Can we identify known shipping lanes?
   b. How do the outputs look near ports?
   c. Do we have events in known fishing Areas?



Ai2