Glossary: Artificial Intelligence and Machine Learning for Electronic Monitoring of Fisheries

In January 2023, The Pew Charitable Trusts hosted its first *Global Artificial Intelligence in Fisheries Monitoring Summit*, which brought together experts from across the globe to discuss how artificial intelligence and machine learning (AI/ML) can help secure greater transparency and accountability throughout the world's fisheries, particularly when used in conjunction with electronic monitoring (EM) technologies. During the meeting, participants agreed it would be beneficial to develop an AI/ML glossary to promote the use of common language between the many stakeholders who work in this space. After the meeting, a small group of attendees refined the following list of terms and definitions.

Recognizing the rapid advancements in automation, electronic monitoring, and their respective policies, these terms and definitions will likely evolve and necessitate future review. Further, these terms may be understood and defined differently in different sectors, situations, and/or specific circumstances. As such, these terms should be read and used with the understanding that they are dynamic and subject to modification in the future.

We'd like to thank all Summit participants for their initial input, and recognize Amanda Barney, Matt Dawkins, Jimmy Freese, Jamie Gibbon, Raiana McKinney, Hilario Murua, Marlon Roman, and Michael Stanley for their additional efforts in crafting this glossary.

AI/ML Terms

Term	Definition
Algorithm	A set of instructions or rules that a computer follows to perform a task. Machine learning algorithms are designed to learn from data and improve their performance over time on specific tasks, predictions, or decisions.
Artificial Intelligence (AI)	A science and engineering approach to solve problems using a digital computer performing tasks that are generally carried out by intelligent beings.
Audit	An assessment of a machine learning system and its performance, accuracy and compliance with procedures or standards. Audits should consider the wider system in which the AI algorithm is integrated, quantify any bias that may be present, and inform stakeholders of shortcomings or limitations.
Bias	A systematic directional error in a machine learning model that results in incorrect predictions or decisions. Bias can occur when a model is trained on a dataset that is not representative of the population or system modeled.
Bootstrapping	A technique used to estimate the variability of a statistical measure or to create multiple datasets with slight variations for training and evaluation purposes. Bootstrapping involves repeatedly choosing different sets of data points from an original dataset to create new samples that are similar but not identical to the original data. These new samples are used to train and test machine learning models.
Classification	Assigning data to one or more predefined categories or classes, which are then used to train a machine learning algorithm. Example: Assigning an image to the 'fish' category if the image contains fish and assigning an image to the 'no fish' category if no fish are present.

Convolutional Neural Network (CNN)	A type of deep learning algorithm that is often used to recognize, analyze, and process image data and separate images into distinct categories using multiple filters. For example, to detect a fish in an image a CNN would first use a filter that finds the eye of a fish and then a filter that extracts textures to find scales.
Cross-validation	A technique used to evaluate a model's performance by dividing the dataset into training and validation sets and training the model on different combinations of the data. Cross-validation can help reduce overfitting when splitting the data into a single training set would inadequately represent the overall distribution of the data.
Data Retention	Preserving data for a specified period of time after it has been used for its original intended purpose. Data is often retained to meet regulatory requirements and allows systems that were developed from the data to be audited.
Deep Learning	A subset of machine learning that uses algorithms with multiple layers to extract patterns and features, which allow the model to determine which features are the most important when classifying data. Example: A deep learning model could determine which features (i.e., fin or head shape) are most important when classifying an image of a fish as either a trout or a salmon.
Edge Processing	The application of machine learning algorithms by a device close to where the data was gathered, like onboard a fishing vessel.
Hyperparameter	A parameter that must be set before training a machine learning model, as it cannot be learned from the data. “Hyperparameters sweeps” involve training many versions of a model using different combinations of hyperparameters in order to evaluate the model’s performance.
Inference	The process of using a trained machine learning model to make predictions or decisions about new data.
Machine Learning	A subset of AI that refers to the use and development of computer systems that are able to learn and adapt without following explicit instructions, by using algorithms and statistical models to analyze and draw inferences from patterns in data.

Model	A mathematical representation of a system or process that is used to make predictions or decisions.
Model Architecture	The specific structure of a machine learning model. The architecture of a model defines the design, configurations, and connections of the elements of the model and determines how it processes input data and produces outputs, predictions, or inferences. Some popular examples of model architectures include feedforward neural networks, convolutional neural networks, and recurrent neural networks.
Model Framework	A software library or platform that provides the necessary tools and infrastructure for building, training, and deploying machine learning models. A framework provides the basic building blocks such as data preprocessing tools, optimization algorithms, and evaluation metrics, which can be used to build and train machine learning models. Some popular examples of machine learning training frameworks include TensorFlow, PyTorch, and scikit-learn.
Overfitting	When a machine learning model performs well on the training data but poorly on new or unseen data. Also known as Memorization.
Supervised Learning	A type of machine learning in which a model is trained on labeled input data where each input is associated or matched to a known output. Example: Supervised learning would involve training a model on 1000 labeled images of salmon that include boxes drawn each fish.
Training, Validation, and Testing Sets	Training set: A set of data that a machine learning model uses to identify distinct patterns and set model weights. Validation set: A set of data used to select the best version of the model to save and set some training parameters. Test set: A set of unseen data, not used during training, used to evaluate the performance of a machine learning model. The results from the test set are considered the most reliable measurement of the model’s performance on new data.
Transferability	The capability of a trained algorithm to apply knowledge from previous data to new or unseen data. Example: An algorithm that was trained on data from one fishing vessel would have high transferability if its performance does not drop significantly when it is run on footage from the wider fleet.

Underfitting	Underfitting is when a machine learning model is not complex enough to accurately capture the underlying trend in the training data.
Unsupervised Learning	A type of machine learning in which a model is trained on unlabeled data, where each input is not associated or matched to a known output, resulting in the model categorizing objects without human intervention. Example: Unsupervised learning would involve training a model on 1000 images of salmon that do not include boxes drawn each fish.
Unseen Data	Data not included in training or validation sets.

Metrics

Term	**Definition**
Accuracy	A measurement of the performance of a machine learning algorithm or model that measures the proportion of correct predictions. It is calculated as the sum of the true positives and true negatives as a percentage of all the items in the dataset. Example: For an algorithm that classifies images into two species of fish. If we had 10 images, 5 of trout and 5 of salmon and 4 images were predicted to be of salmon and 6 of trout, our accuracy would be 90% as 9 predictions were correct.
Confidence	A measurement of the relative certainty of a prediction by a machine learning algorithm or model. A high confidence is generally associated with a better prediction. Example: For an algorithm that classifies images into two species of fish, trout, and salmon, if an image is classified as a trout with a 0.95 confidence and as a salmon with 0.30 confidence, it can be determined that the model has predicted the image to contain a trout and not a salmon.

Confusion Matrix	A table that shows the performance of a model or algorithm that classifies items into multiple categories, by comparing the model's outputs to the true values. A confusion matrix can be used to visually display which categories are commonly misclassified by the model.
Precision	A measurement of the performance of a machine learning algorithm or model that measures the proportion of correct predictions among all predictions made. Example: For an algorithm that classifies images into two species of fish, trout, and salmon, if there is a bucket of 8 trout and 2 salmon, and the algorithm classified all 10 fish as trout, then the precision would be 8/(8+2), or 80%.
Recall	A measurement of the performance of a machine learning algorithm or model that measures the proportion of correct predictions among all actual occurrences in the dataset, (i.e., the proportion of actual items the model was able to correctly identify). Example: For an algorithm that classifies images into two species of fish, trout, and salmon, if there is a bucket of 10 salmon and the model classified 6 salmon as salmon and 4 as trout, the recall would be 6/(6+4), or 60%.
False Positive/Type I Error	A false positive occurs when a model or algorithm incorrectly identifies an object or event. False positives can lead to the model raising a false alarm or recommending an unnecessary intervention. Example: An algorithm Identifying a fish as a trout when it is actually a salmon.
False negative/Type II Error	A false negative occurs when the model does not identify an object or event that actually exists. This could lead to the model missing an actual event or failing to take appropriate action. Example: An algorithm not Identifying a trout when it appears in an image.
True Positive	A true positive occurs when a model or algorithm correctly identifies an object or event.
True Negative	A true negative occurs when the model or algorithm correctly identifies that an object or event does not exist.