Appendix 5 to *Narratives of Epistemic Agency in Citizen Science Classification Projects: Ideals of Science and Roles of Citizens*.
Submitted to *AI & Society*.
Corresponding Author: Marisa Ponti, Department of Applied Information Technology, University of Gothenburg, Gothenburg, Sweden, marisa.ponti@ait.gu.se, ORCID ID: http://orcid.org/0000-0003-4708-4048

# Coding structure

## Codes

| Name | Description |
|---|---|
| 000. Epistemic Human Agency | |
| 002. Type of human task | It describes the types of tasks performed by humans. |
| 003. Citizen task | |
| 003.1 Data Analysis and Interpretation | |
| 004. Do collaborative analysis | It describes instances of collaboration between citizens and scientists, e.g., to analyze data |
| 003.1 Data Classification and Annotation | |
| 004. Amplify expert decisions | It describes citizen scientists using expert classification for a larger or an entire dataset. |
| 004. Annotate observations | It describes citizen scientists adding context information associated with an observation, e.g., location and date/time of a photo. This code also labels adding other media, e.g,., video, audio, to an observation. |
| 004. Citizen classify | It describes citizen scientists classifying data. They can assign golden standard labels to objects or assign their own labels. |
| 004. Follow authoritative taxonomies | It describes the use of authoritative taxonomies selected by experts for the identification of species. |
| 004. Generate taxonomies | It describes citizen scientists generating tags to classify objects. |

| Name | Description |
| --- | --- |
| 004. Make predictions | It describes citizens using historical data to predict future events. |
| 003.1 Data Validation and Quality Control | |
| 004. Review machine annotations | It describes citizen scientists reviewing machine annotation of data. |
| 004. Review machine training data | It describes citizen scientists revising and selecting machine training data. |
| 003.1 Feedback Provision | |
| 004. Peer review | It describes citizen scientists sharing ideas, giving each other feedback, and acknowledging and validating contributions. |
| 004. Provide feedback and bugs reporting | It describes citizen scientist feedback on aspects of the project, e.g., features of the application. |
| 003. Researcher task | |
| 003.1 Data Analysis and Interpretation | |
| 004. Create hybrid predictions | It describes experts combining machine and human predictions. |
| 003.1 Data Classification and Annotation | |
| 004. Identify classes of interest | It describes the “classes” which are the concepts that experts want to detect in the data. |
| 004. Label data | It describes the process of preparing the data for supervised learning (classification). An example is tagging an object with one or more tags that are informative. |
| 003.1 Data Management | |
| 004. Curate data | It describes the activities involved in the curation of a dataset, e.g., image update, annotation editing, provision of links for connecting collections, removal of duplicates. |
| 004. Extract metadata | It describes the extraction of metadata by experts, e.g., creation date, etc., from images and other types of files. This labels also the correction of metadata. |

| Name | Description |
|---|---|
| 003.1 Data Validation and Quality Control | Ensure that a set of inherent characteristics of an object fulfills requirements. It labels a variety of mechanisms, e.g., consensus scoring. Data accuracy and correctness is one of the components of data quality. This code also labels assessment of citizen scientist classifications. |
| 004. Aggregate labels to reach a consensus | It describes the aggregation of labels from multiple raters for a given example to arrive at a single “consensus” label. |
| 004. Assess expert rating | It describes the assessment of a consistency of a measure. For example, over time (test-retest reliability), across items (internal consistency), and across different researchers (inter-rater reliability). |
| 004. Attribute confidence score | It describes the confidence level attributed by expert raters to their image classification. |
| 004. Create a coding frame | It describes the creation of the guidelines to be used to label and interpret data. |
| 004. Develop a citizen-labelled data set | It describes the development of a “consensus” dataset based on citizen single classifications. |
| 004. Develop a gold-standard dataset | It describes a dataset that is accepted as the most accurate and reliable and needs to be used as a reference for checking and comparing machine labeling. It also includes the development of ground truth datasets. |
| 004. Provide a taxonomic nomenclature | It describes the use of recognized taxonomic nomenclatures to ensure data accuracy and correctness. |
| 004. Validate classifications | It describes methods used by experts to validate the quality of classifications made by humans and machines. For example, it can include experts seeking to ensure that citizen scientist-trained machine model generates correct classifications. Validation can encompass various methods, e.g,., consensus answer, expert validation, etc. |
| 003.1 Training | |
| 004. Boost training data | It describes the implementation of an algorithm aimed at putting together a strong learner, given a collection (possibly infinite) of weak learners. The idea of boosting is to train weak learners sequentially, each trying to correct its predecessor. |
| 004. Train citizen scientists | It describes the process of training citizen scientists to help experts label images. |
| 004. Train machine classification | It describes the use of labeled datasets for training machine learning to classify objects. It also includes the use of pre-trained models. |
| 000. Machine Epistemic Agency | |
| 001. Type of machine task | |

| Name | Description |
| --- | --- |
| 003.1 Data Detection and Classification | |
| 004. Aggregate citizen classifications | It describes the aggregation of citizen-scientists classifications. |
| 004. Automatic classification | It describes machine classifying data by assigning labels to objects. |
| 004. Detect automatically | It describes an algorithm able to recognize objects, e.g., plants and animals. |
| 004. Find missing items in a database | It describes an algorithm searching and finding items missing in a database. |
| 004. Make predictions | It describes machines using historical data to predict future events. |
| 004. Produce consensus dataset | It describes the construction of a ‘consensus dataset’ of citizen final classifications validated against a ‘gold-standard’ dataset. |
| 004. Rank observations | It describes an algorithm ranking observations for information retrieval systems. Training data consists of lists of items with some partial order specified between items in each list. This order is typically induced by giving a numerical or ordinal score or a binary judgment (e.g. “research grade or “casual” in iNaturalist). |
| 003.1 Data Management | |
| 004. Extract metadata | It describes the machine extraction of metadata, e.g., creation date, etc., from images and other types of files. |
| 004. Store data | It describes the features of the system to store data. |
| 003.1 Help Human Observation | |
| 004. Help identify objects of interest | It describes machines helping citizen scientists identify objects of interest in images, for example. |
| 004. Provide geolocation data | It describes machines providing geolocated data about observations. |
| 004. Provide recommendations | It describes the provision of recommendations or suggestions after a system has executed a set of matchers. |
| 003.1 Teaching Itself | |
| 004. Extract training data | In unsupervised learning, it describes the machine extraction of suggestions to train unlabeled data to find patterns, such as inferences or clustering of data points. |
| 004. Learn the training dataset | It describes machines learning the “gold-standard" labeled dataset to predict labels when classifying data. |

| Name | Description |
|---|---|
| 001. Input | It describes the human agent’s input in the system. |
| 002. Domain expert input | It refers to the system requiring the input of a domain expert |
| 002. End-user input | It describes the the input provided by non-expert end-users to the system |
| 002. ML expert input | It describes the system requiring the input of an ML expert |
| 002. Query strategy | It describes which query strategy the algorithm uses to learn. We need to be careful that we don't refer to, e.g. an optimizer as a machine learning algorithm. |
| 003. Active learning | It describes the level of accuracy - for example, an accurate prediction - a machine learning algorithm can achieve if it can choose the data it learns from. |
| 003. Offline query strategy | It describes the data needed to train a machine learning algorithm, data which is collected beforehand. |
| 003. Online query strategy | It describes the data provided by humans and directly fed into a machine learning algorithm. |
| 001. Interaction | |
| 002. Human-Algorithm Interaction | It describes how end-user and algorithms interact |
| 003. Interaction by Data | It describes if data is produced by end-users independently of each other, and then algorithms work on the aggregated collective data |
| 003. Performing tasks together | It describes end-user and machines performing tasks together at the same time |
| 003. Taking turn | It describes end-user and machines taking turns to solve problems together |
| 002. Human-environment interaction | It describes whether humans get the full environment for making an input represented by the data/system or if there is latent information in the task that might not be directly represented in the data but requires external collection/sensing of information |
| 002. Human-human interaction | It describes the interaction between human participants |
| 003. Collaboration | It describes humans performing tasks collectively or cooperatively |
| 003. Independence | It describes humans carrying out tasks entirely or partially on their own |
| 001. Machine learning paradigm | It describes the type of algorithm implemented. |
| 002. Reinforcement learning | It describes an algorithm taking sequential actions in a dynamically evolving environment. The agent searches for a policy (action rules), which maximizes the total reward. |
| 002. Semi-supervised | |

| Name | Description |
| --- | --- |
| 003. Active Learning | It describes a form of semi-supervised machine learning in which data is taken, trained, tuned, tested, and fed back into the algorithm to make it more accurate. The program can actively query an authority source, either the programmer or a labeled dataset, to learn the correct prediction for a given problem. |
| 002. Supervised learning | It describes humans training the machine learning algorithm to generate an answer based on a known and labeled data set. Classification and regression algorithms, such as artificial neural networks, decision trees and support vector machines, are commonly used for supervised learning tasks. |
| 003. Prediction | It describes the use of regression algorithms, including decision trees, and support vector machines, which are used for prediction in supervised learning. |
| 003. Recognition | It describes a recognition of objects, such as images, or natural language text. |
| 003. Recommendation | It describes an algorithm aimed at suggesting relevant items to humans. Recommender systems are an example. |
| 003. Regression | |
| 002. Unsupervised learning | It describes algorithms generating answers on unlabeled data. |
| 003. Clustering | It describes clustering algorithms, such as K-means, often used in unsupervised machine learning to group data points, based on a specified proximity measure. |
| 001. Participation | |
| 002. Activity | It describes the activity performed by an end-user |
| 003. Collecting data | It describes activities that focus on collecting and contributing data |
| 003. Create new puzzles | |
| 003. Create new representations | It describes activities that focus on the use and generation of new external representations (such as the possibility to create a new feature of the system like scripts that can help one solve a problem. |
| 003. Data processing and coding | It describes activities that focus on data processing, e.g., classify image content |
| 003. Defining research questions | It describes activities that focus on identifying research questions or issues |
| 003. Developing methods and material | It describes activities that focus on the development of data collection methods |
| 003. Translation and diffusion | It describes activities that aim at translating/synthesizing findings and content for different audiences, e.g., policy-makers, laymen. |
| 001. Task Characteristics | It describes characteristics of a task. |

| Name | Description |
|---|---|
| 002. Collaborative Output | It describes the output generated by humans and machines |
| 002. Data representation | It describes to the shared data representation, which is the data that is shown to both the human and the machine before executing their tasks |
| 003. Data representation by concept | It describes data represented by a concept. |
| 003. Data representation by feature | It describes data represented by a feature, like human height or weight |
| 003. Data representation by instance | It describes data represented by an instance, such as multiple instances that belong to the same type, e.g., pictures of humans. |
| 002. Goal | It describes the goal(s) of humans and algorithms in solving a problem. |
| 003. Adversarial goal | It describes an adversarial goal (e.g., AI aims to beat humans in games). |
| 003. Common goal | It describes solving a problem through a combination of knowledge and abilities, or both. |
| 003. Goal of the algorithm | It describes the goal of the algorithm in those cases where algorithms are implemented. |
| 003. Human goal | It describes the goal of the human participant |
| 003. Independent goal | It describes independent goals: cooperation between humans and algorithms, such as humans training image classifiers without being involved in the end solution. |
| 002. Type of task instructions | It describes how humans and machines are instructed to perform a task. This code can refer to rules embedded in algorithms, golden-standard datasets, guidelines for labeling, etc. |
| 003. Authoritative taxonomies | It describes authoritative taxonomies for taxonomic identification. |
| 003. Citizen-labelled dataset | It describes a dataset labelled by citizen scientists. |
| 003. Coding frame | It describes the guidelines developed by humans (usually experts) for how to label and interpret (code) observations. |
| 003. Confidence criterion | It describes the confidence score used as a quantitative measure to estimate the confidence of a human classification. |
| 003. Data standard | It describes frameworks for compiling data from varied and variable sources. One example is the Darwin Core Standard. |
| 003. Gold-standard dataset | It describes a particular case of external criterion. A statistical or machine learning algorithm wants to predict a |

| Name | Description |
| --- | --- |
| | criterion that is not dependent on the algorithm (otherwise criterion is "contaminated"). “Gold-standard" is usually a dataset or a set of results which serves as the approved external criterion. |
| 003. Key data fields | It describes the relevant fields that should be either made mandatory or suggested to ensure essential data is gathered. |
| 003. Pre-trained model | It describes a model that is trained on a large benchmark dataset to solve a problem similar to the one that we want to solve. It is used for transfer learning. |
| 003. Rules and patterns | It describes the rules and patterns used by a recommender system for using attributes of items/users, explore content, and making suggestions. |
| 003. Tutorial | It describes tutorials and other forms of support to help citizen scientists walk through the main features of an application, or through the process of classification. |