National Yang Ming Chiao Tung University AI Model and Dataset Management Guidelines

Dataset Management

Metadata Requirements:

To maximize dataset sharing and usability, metadata should include the following fields and be provided in machine-readable formats such as JSON/XML/CSV:

(1) Group
(2) Name
(3) Description
(4) Creation time
(5) Sharing scope and access method (e.g., no application, requires approval, downloadable, VPN-only)
(6) Tags
(7) Creator
(8) Others: e.g., privacy information, file size and type, data labeling details

FAIR Principles:

Datasets should follow FAIR principles (Findable, Accessible, Interoperable, Reusable):

- **Findable**: Include tags and indexing information for searchability.
- **Accessible**: Data can be accessed via API. Sensitive data requires a token.
- **Interoperable**: Use standard vocabularies and formats; ontology recommended (e.g., SNOMED CT).
- **Reusable**: Clearly describe data license and source.

Model Management

Model Metadata Requirements:

Models should be uploaded with metadata in machine-readable formats and include:

(1) Group: AI model
(2) Name
(3) Description

(4) Creation time

(5) License type

(6) Tags

(7) Creator

(8) Framework and version

(9) Training dataset name, source, size, environment, results

(10) Testing dataset details; include platform links if available

Model Testing:

Before upload, models must be tested and results submitted in CSV format, including:

- Testing data source
- Computation resources used (CPU/GPU, memory, storage)
- Usage methods (inference code, evaluation scripts)
- Expected results (e.g., accuracy, F1 score)

Model Performance Metrics:

Optional additional evaluation using separate training/testing sets, ensuring consistent data distribution. Report should include:

- Source of training/testing data
- Data format and size
- Computational resource usage
- Expected results: type, output format, execution time, metrics (e.g., accuracy, precision)