Evidence Appendix — Data Sheet, Licensing, and Governance (Mealyze)

https://drive.google.com/drive/folders/1BsEpLKr7FObXB2jqxgw2l5prwWADBKB7?usp=sharing

A. Overview (Data Governance Summary)

Mealyze uses a hybrid data pipeline that separates and manages data by its role (training, evaluation, and service knowledge).

1. Training data: We train an image classification model using a subset of the AI-Hub “Korean Food Images” dataset (dataSetSn=79).
2. Service knowledge base: We separately build a structured Food DB (300 foods, CSV) and use it as a retrieval-based source for recommendations, explanations, and safety filtering.
3. Evaluation-only data (RWTS): We collect food images from the internet that are labeled as “commercial-use allowed (or reuse permitted)” and use them only for smoke-check observations of failure modes under real-world conditions (low light, camera angle variation, background clutter, mixed dishes). RWTS is not used for training.

Core governance principles

- Purpose separation: We separate training data, evaluation data, and the service DB to reduce data leakage (evaluation data unintentionally entering training) and copyright risk.
- No raw redistribution: We do not include raw AI-Hub images or raw RWTS images in submissions or public repositories (only metrics, visualizations, and metadata).
- Safety-first: Allergen tags are labeled conservatively (set to 1 even if an ingredient is only *possibly* present). The service is disclosed as non-medical (for guidance only), and a manual override option is provided.
- Submission-time evidence: For items that lacked sufficient records during development (RWTS licensing/PII), we supplement evidence retroactively within the “submission-time evidence scope” by providing verification logs and screenshots.

B. Dataset Inventory

B1) AI-Hub Korean Food Images (dataSetSn=79 subset)

- Purpose: Training (food image classifier)
- Size used: 31 classes × 999 images = 30,969
- Access/Policy: Following AI-Hub policy and FAQ guidance, we (i) do not transfer/assign/sell the raw dataset to third parties, (ii) disclose AI-Hub usage when results are publicly shared, and (iii) use the data within an education/research scope.
- Redistribution: We do not include raw images/labels in submissions or public repositories; we include only results such as metrics, graphs, and model artifacts.
- Known risk: Because many training images are captured under clean, food-centered conditions, domain shift may occur in real-world environments.
- Mitigation: RWTS smoke-check + user input guidance (lighting/angle) + manual override.

B2) RWTS (Internet-sourced, evaluation-only)

- Purpose: Robustness smoke-check (NOT for training)
- Size: Small & qualitative (e.g., ~6 images per class)
- Truthful limitation: During development, we did not maintain a standardized per-image log for licensing/provenance/PII actions.
- Submission-time control: For the subset of RWTS samples used to support observations in this submission, we retroactively compile an Appendix log that includes:
  (1) source URL, (2) reuse/commercial-use permission evidence (license/terms/FAQ page),
  (3) checked_date, and (4) a screenshot of the license statement.
  Any RWTS sample without clear permission evidence is excluded from the submission-time evidence scope.
- Redistribution: We do not include raw RWTS images in submissions or public repositories (metadata + screenshots only).

B3) Food DB (meal_advisor_food_db_v1.csv)

- Purpose: Retrieval-based lookup knowledge base (nutrition + tags + safety)
- Size: 300 foods with per-serving nutrition values and 0/1 tags
- Nutrition policy (1-serving standardization):
  (i) kcal/protein_g/fat_g/carbs_g are standardized as a per-serving representative (“guide”) value,
  (ii) serving_desc explicitly states the unit (e.g., 1 bowl/1 plate/1 roll/100g),
  (iii) values are entered based on one of: (a) public/national nutrition DB, (b) manufacturer nutrition labels, or (c) averages from multiple recipes, and item-level provenance is managed via an Appendix (sample) log.
- Primary nutrition references used in DB construction:
  • Food Safety Korea (MFDS) Food Nutrition Database
  • RDA Korean Standard Food Composition Database
  • USDA FoodData Central
- Allergen policy (conservative):
  contains_* is labeled as 1 if an ingredient is even *possibly* present, prioritizing safety warnings.

C. RWTS Governance (evaluation-only) — Retroactive Evidence Protocol (submission-time)

C1) What RWTS is (truthful)

- RWTS is not team-photographed data. It is assembled from internet images labeled as “reuse/commercial-use allowed.”
- During development, we did not maintain a standardized per-image log for licensing/provenance/PII actions.

C2) Evidence scope rule

- RWTS samples referenced in this document are limited to the submission-time evidence scope (RWTS-Evidence Scope).
- Inclusion requirements for the evidence scope:
  (1) source_url is available
  (2) permission evidence is found on license_url (or Terms/FAQ) explicitly stating reuse/commercial-use is allowed
  (3) checked_date is recorded and a license screenshot is saved

(4) PII screening is performed (face/contact info/receipts/license plates, etc.) and an action is recorded (kept/cropped/removed)
- If a sample does not meet these conditions, it is excluded from the evidence scope and is not used to support any claim.

C3) Limited Retroactive License Verification (Submission Scope)

For a limited number of RWTS samples referenced in this report, we retrospectively verified reuse/commercial-use permissions via publicly available license or terms pages. No RWTS raw images are redistributed.

D. Food DB Schema & Labeling Rules (Data Dictionary + Decision Logic)

D1) Core columns (identity & serving)

- id (int): 1..300, unique key
- name_en (string): English food name (for UI and DB lookup)
- category (string): High-level category (e.g., Korean/Japanese/Chinese/Dessert/Drinks/…)
- serving_desc (string): Unit definition for 1 serving (e.g., 1 bowl, 1 plate, 100g)

D2) Nutrition columns (per 1 serving; numeric, non-negative)

- kcal (number): energy (kcal)
- protein_g (number): protein grams
- fat_g (number): fat grams
- carbs_g (number): carbohydrate grams

D3) Taste & type tags (0/1)

- spicy: 1 if perceived as at least “typical ramen-level” spiciness or higher
- sweet: 1 if dessert-like or sugar/syrup-based sweetness is central
- salty: 1 if classified as high-sodium (soups, salted/fermented sides, pickles, processed meats, etc.)
- oily: 1 if oil usage is high or fatty cuts/creamy sauces are central
- deep_fried: 1 if deep-frying is the core cooking method
- grilled: 1 if grilling/pan-grilling is the core method
- noodle: 1 if noodle-based
- rice_based: 1 if rice-based (bowls, bibimbap, fried rice, etc.)
- soup_or_stew: 1 if broth-based (soups, stews, ramen, pho, etc.)
- one_dish_meal: 1 if it is a self-contained single meal item (bowls, noodles, pizza, burgers, etc.)

D4) Context-aware tags (0/1; for coaching UX)

- soul_food: 1 if commonly perceived as mood-lifting/comfort food or culturally “soul-food”
- stomach_sensitive: 1 if gentle on digestion (not spicy + not oily + soft texture)
- hangover_friendly: 1 if commonly associated with hangover relief (broth-based, stimulating soups)
- late_night: 1 if considered high-risk after 10 PM (high fat/sodium, heavy digestion burden)

- Mutually-exclusive rule: soul_food and stomach_sensitive cannot both be 1 in principle. Exception: A small set of foods can be both satisfying and gentle (e.g., mild hotpots/shabu-shabu/chicken stew), allowing both to be 1.

D5) Goal / nutrition-based tags (0/1; quantitative thresholds)

- cut_friendly = 1 if (kcal ≤ 450) AND (fat_g ≤ 15) AND (category != Dessert)
- bulk_friendly = 1 if (kcal ≥ 600) AND (protein_g ≥ 20)
- high_protein = 1 if (protein_g ≥ 20)
- low_fat = 1 if (fat_g ≤ 10)
- low_carb = 1 if (carbs_g ≤ 40)
- low_sugar: auxiliary tag for sugar control (kept conservative for now; can be refined as provenance expands)

D6) Safety / allergen tags (0/1; conservative)

- contains_egg / contains_milk / contains_wheat / contains_peanut / contains_seafood / contains_cucumber …
- Policy: Label as 1 if an ingredient may be present in common recipes (warning-first).
- UI policy: Allergen tags are used for hard-block (exclusion) or strong warnings.

E. Potential Biases & Failure Modes

E1) Domain shift: lighting/angle/background

- Failure trigger buckets (based on RWTS evidence scope):
  (a) Low-light: indoor restaurants/night scenes/noise increase
  (b) Extreme angle: side shots/close-ups/partial crops
  (c) Background clutter: trays, hands, chopsticks, cups, menus, etc.
  (d) Mixed plate: multiple foods in one image
- Expected behavior: The model may be pulled toward the class with the strongest visible pattern (noodles/broth/fried textures, etc.).
- Evidence hook: These failure patterns were qualitatively observed
- using evaluation-only images within a limited evidence scope.
- Mitigation: Capture guidance (bright lighting + top/45°) + manual override + future expansion by condition buckets.

E2) Coverage limitation: 31-class classifier vs 300-food DB

- Failure: Foods outside the 31 trained labels may be predicted as the “closest-looking” among the 31 (a label-space limit, not a true “bug”).
- Evidence hook: Appendix B includes the 31-class label list file.
- Mitigation: If recognition fails or is uncertain, the user can search/select via text-based DB lookup (override) to continue the flow.

E3) Look-alike confusion pairs (high-impact patterns)

- Noodle/Soup family: ramen vs kalguksu vs jjajangmyeon vs jjamppong, etc.
- Fried family: fried chicken vs yangnyeom chicken, etc.
- Rice bowl family: kimchi fried rice vs bibimbap variants, etc.

- Evidence hook: Confusion-pair patterns were observed in evaluation-only settings and informed the user-correction UX design.
- Mitigation: (If available) show Top-k candidates + user selection/correction UX.

E4) Nutrition uncertainty / recipe variance

- Failure: Nutrition values vary across brands/recipes/portions, so DB guide values may differ from a user's actual meal.
- Evidence hook: Appendix D (provenance sample log) + serving_desc disclosure.
- Mitigation: "Guide value" disclosure + portion adjustment options + future standardization and provenance expansion.

E5) Label noise & ambiguity (especially RWTS)

- Failure: Internet images can be ambiguous or show mixed plates, weakening "ground truth."
- Evidence hook: Only qualitatively clear cases were considered for failure-mode observation, not for quantitative claims.
- Mitigation: Do not use RWTS for quantitative accuracy claims; use it only for failure-mode observation and UX improvements.

F. Storage, Access Control, and Retention

F1) Storage

- Raw datasets (AI-Hub and raw RWTS images) are stored only in a private team repository with restricted access.
- No raw images are included in submissions or public repositories (metrics/visualizations/metadata/logs only).

F2) Retention

- After the course ends, RWTS raw images are kept private or deleted according to a retention policy.
- User inputs/logs are not used for training and follow a minimal collection/minimal retention principle.