

MM-Food-100K: A 100,000-Sample Multimodal Food Intelligence Dataset with Verifiable Provenance

Yi Dong*

Yusuke Muraoka†

Scott Shi‡

Yi Zhang§

Binance Wallet Community
Codatta Community

Abstract

We present **MM-Food-100K**, a high-quality, 100,000-sample multimodal food intelligence dataset designed for fine-tuning AI models. Our proposed data sourcing workflow combines community contributions with automated quality review by Large Vision-Language Models, resulting in real-world photos with rich, multi-level annotations. We demonstrate the dataset’s utility by showing that models fine-tuned on MM-Food-100K consistently outperform original Large Vision-Language Models on food classification and regression tasks. We also introduce Codatta Protocol, a novel data contribution framework that uses a privacy-preserving blockchain protocol to track data provenance and enable royalty-based rewards for contributors. MM-Food-100K is released for public use, and the Codatta Protocol provides a new paradigm for building and sustaining high-quality, community-sourced datasets.

1 Introduction

Human-generated data remains indispensable for training reliable multimodal AI systems. While synthetic data can amplify scale, it tends to recycle model priors, misses edge-case realism, and struggles with grounded facts. As a result, real-world sourced data and advanced, knowledge-rich labeling continues to be in high demand. Today’s human-intelligence pipelines exhibit a **dual-market failure**. On the demand side, businesses face high costs to obtain domain-rich, verifiable data. On the supply side, knowledge providers are frequently underpaid or misaligned with downstream value, while attribution is rarely preserved, depressing participation and diversity of expertise.

In this paper we introduce **MM-Food-100K**, a 100,000-sample, multimodal (image + textual annotation) dataset released for public research. The dataset samples are diverse, showcasing a variety of dishes and ingredients, as shown in Figure 1. MM-Food-100K is a curated ~10% open subset of an original **1.2-million**, quality-accepted corpus collected in six weeks from 87,000+ contributors. Data is sourced via a partnership campaign between **Binance Wallet** and the **Codatta protocol**, which combines community sourcing with configurable AI-assisted quality checks. Each submission is linked to an on-chain **wallet address** to provide verifiable provenance and attribution; a fully on-chain protocol is on the roadmap. Importantly, contributors were invited with zero upfront payment; instead, the model emphasizes attribution and prospective revenue sharing for the retained commercial portion of the corpus, aligning incentives without front-loading cost or risk.

Contributions:

- **Dataset (primary).** We release **MM-Food-100K**, a public multimodal food-intelligence dataset annotated for dish/cuisine, visible ingredients, portion and scale cues and etc., with evidence and provenance for auditability. We demonstrate its utility by fine-tuning large vision-language models for image-based food-related prediction (such as dish recognition), observing consistent improvements over out-of-box baselines.

*sky@inductive.network, Codatta

†yusuke.muraoka@gokite.ai, Kite AI

‡Scott.shi@gokite.ai, Kite AI

§PhD, yi@inductive.network, Codatta

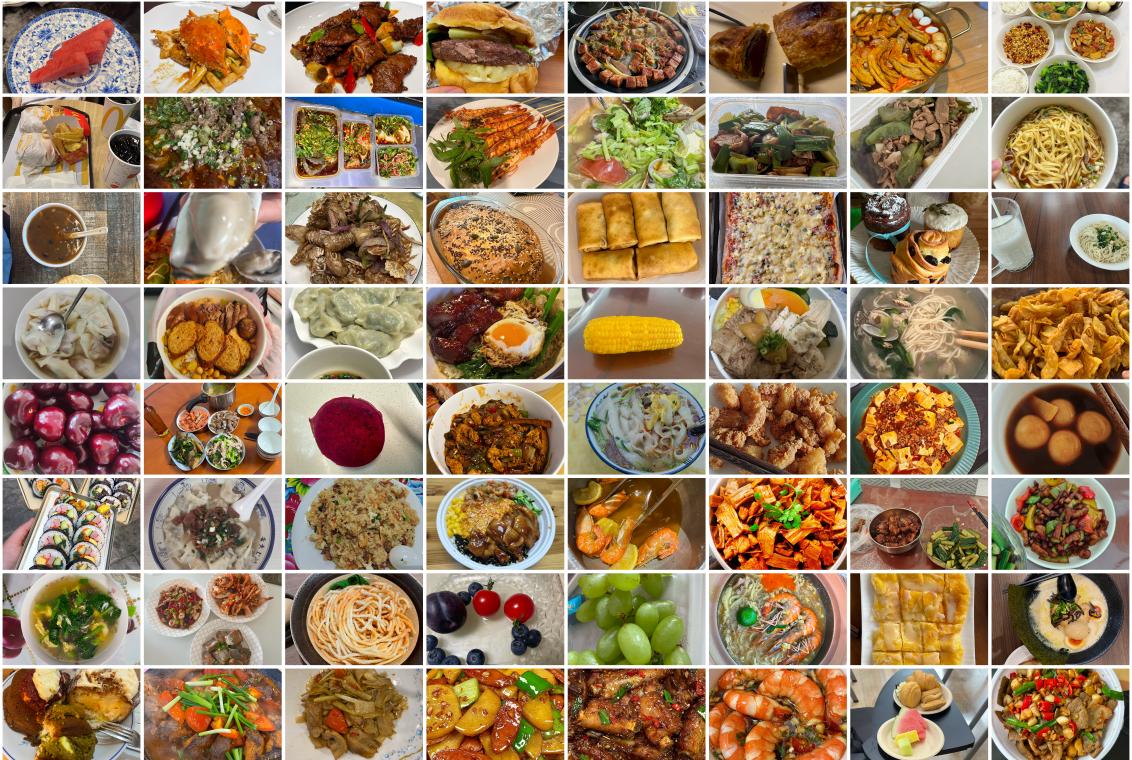


Figure 1: Overview of Food Intelligence Dataset Samples: A collage of diverse dishes and ingredients from multiple cuisines, showcasing the variety in the Food Intelligence dataset—including noodles, dumplings, soups, stir-fries, seafood, meats, breads, fruits, and snacks—captured in varied lighting, perspectives, and serving styles.

- **Framework (secondary).** We describe briefly the **Codatta** blockchain-based data contribution model—data infrastructure for privacy-preserving transparency, structured task/schema design, and configurable human/AI quality assurance, contributor profile and reputation system, and wallet-linked provenance—that enabled verifiable data acquisition at scale and royalty-based payment business model via data classification.

2 The MM-Food-100K Dataset

2.1 Overview and applications

MM-Food-100K is a public, 100,000-sample subset of a larger ~1.2M food corpus curated for **fine-grained food intelligence**. Each record pairs an image with structured annotations that enable recognition (dish/cuisine), ingredient extraction, portion and calorie estimation, cooking-method reasoning, and authenticity checks. The dataset targets practical applications such as image-based nutrition estimation, recipe recommendation, diet logging/health monitoring, and fine-grained vision research.

Why another dataset? A scan of publicly available food datasets highlights three recurring gaps: (i) insufficient diversity and scale, (ii) limited annotation, “monolithic” labels with little schema depth (e.g., no portion/nutrition), and (iii) curated imagery that diverges from real-world photos. MM-Food-100K addresses these limitations by combining breadth with **multi-level, structured annotations** tailored to downstream AI tasks.

2.2 Schema (what a record contains)

Each item includes an image link and a JSON metadata block with the following key fields (non-exhaustive): `dish_name`, `food_type` (homemade/restaurant/packaged/raw), `ingredients`, `portion_size` (ingredient-level estimates), `nutritional_profile` (kcal/protein/fat/carbs when available), `cooking_method`, and authenticity indicators (`camera_or_phone_prob`, `online_download_prob`, `food_prob`). A visual representation of these annotations, which can be a mix of human and AI

labels, is provided in Figure 2, and the core schema fields are detailed in Table 1. This design supports multi-task learning (classification, extraction, and regression) without changing file formats.

food_picture						
camera_or_phone_prob	0.8	0.8	0.8	0.8	0.8	0.8
food_prob	1	1	1	1	1	1
dish_name	Croissant	Mapo Tofu	Assorted Baked Goods	Grilled Meat Skewers	Hamburger	Shrimp in Sauce
food_type	Packaged food	Homemade food	Homemade food	Restaurant food	Restaurant food	Homemade food
ingredients	["flour", "butter", "yeast", "salt", "sugar"]	["tofu", "ground meat", "green onions", "sauce"]	["flour", "sugar", "butter", "eggs", "chocolate", "raisins"]	["beef", "pork", "spices", "cilantro", "green onions"]	["beef patty", "bun", "lettuce", "pickles", "mayonnaise"]	["shrimp", "soy sauce", "garlic", "green onions", "chili"]
portion_size	["croissant:100g"]	["tofu:200g", "ground meat:150g", "sauce:50g"]	["sweet roll:200g", "cupcake:150g"]	["meat:400g", "garnish:50g"]	["beef patty:150g", "bun:100g", "lettuce:20g", "pickles:30g", "mayonnaise:10g"]	["shrimp:300g", "sauce:50g"]
cooking_method	Baked	stir-frying	Baking	Grilling	Grilled	boiled
nutritional_profile	{"fat_g": 22.0, "protein_g": 8.0, "calories_kcal": 400, "carbohydrate_g": 44.0}	{"fat_g": 25.0, "protein_g": 30.0, "calories_kcal": 400, "carbohydrate_g": 15.0}	{"fat_g": 15.0, "protein_g": 8.0, "calories_kcal": 450, "carbohydrate_g": 70.0}	{"fat_g": 30.0, "protein_g": 40.0, "calories_kcal": 600, "carbohydrate_g": 20.0}	{"fat_g": 25.0, "protein_g": 30.0, "calories_kcal": 500, "carbohydrate_g": 40.0}	{"fat_g": 10.0, "protein_g": 30.0, "calories_kcal": 250, "carbohydrate_g": 5.0}

Figure 2: Selected Samples from the Food Intelligence Dataset: Images with Mixed Human and AI Annotations. Examples of food images with mixed annotations—green cells are human-labeled, red cells are AI-predicted—showing dish details, ingredients, cooking methods, and nutrition.

Table 1: Core schema fields

Group	Example fields	Purpose
Identity	dish_name, food_type, cuisine tag	recognition / taxonomy
Ingredients	ingredients[]	extraction / reasoning
Portion	portion_size (g/ml per item)	portion / calorie models
Nutrition	nutritional_profile (kcal, P/F/C)	nutrition prediction
Preparation	cooking_method	method reasoning
Authenticity	camera_or_phone_prob, online_download_prob, food_prob	data realism filters

2.3 Codatta’s Two-Stage, AI-Augmented Data Sourcing Workflow

Creating high-quality, large-scale multimodal datasets from diverse community sources is challenging due to the trade-off between quality assurance and operational costs. To address this, we designed and implemented a novel two-stage, AI-augmented data sourcing workflow. As illustrated in Figure 3, this pipeline is a core innovation of the Codatta protocol, turning heterogeneous user-generated content into verifiable, high-fidelity records at scale. Our method consists of the following key steps:

- (1) **Submission and Progressive Enrichment:** Initially, a **Human Data Provider** submits an image and corresponding annotations. These annotations follow a structured schema and taxonomy (L1-L5), allowing for progressive enrichment from basic dish names to complex, evidence-backed nutritional and portion cues.
- (2) **Initial Quality Review:** Each submission proceeds to an **Initial Quality Review** stage, where an **Automated LVM Review** conducts a rapid vetting process. This first-pass check filters out low-quality or irrelevant submissions and provides near-real-time feedback. If a submission is rejected, the data provider can revise and resubmit.
- (3) **Final Quality Review:** Submissions that pass the initial review are forwarded to a more thorough **Final Quality Review**. This stage leverages a more sophisticated, **Programmed LVM** to perform detailed checks, ensuring the accuracy and integrity of the data.

- (4) **Data Curation and Distribution:** Accepted and validated data is then processed for **Data Curation** and deduplication. It is subsequently split into two distinct sets: a **Public Access Subset (10%)** for the research community and a **Commercial Access Subset (90%)** for commercial use. This dual-access distribution model retains wallet-linked provenance, laying the foundation for future royalty-based rewards.

This two-stage, AI-augmented workflow achieves high-quality data acquisition at scale and a sustainable balance between cost and data quality. It establishes a transparent process and a reward mechanism that benefits both community contributors and downstream applications.

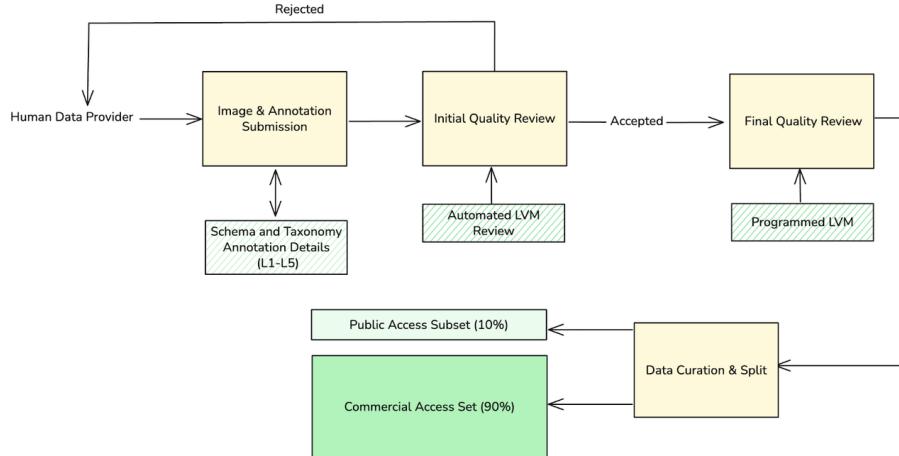


Figure 3: Data Pipeline: Two-Stage Quality Review and Dual-Access Distribution. A workflow for reviewing food image–annotation pairs with quick initial checks, detailed final reviews, and a split into public and commercial sets to balance openness and provider rewards.

2.4 Descriptive statistics (what the open subset looks like)

The dataset card reports **100,000** images with a broad mix of source contexts ('food_type'), illustrating real-world capture conditions and variety. The distribution of 'food_type' is as follows: Homemade (46,555), Restaurant (35,461), Raw vegetables & fruits (9,357), Packaged (8,354), and Others (273). The card also includes authenticity distributions (e.g., `camera_or_phone_prob` bins at 0.8 and 0.7 totaling ~99k images), underscoring the emphasis on user-captured photos. A snapshot of these distributions is presented in Table 2. For a complete overview, see the full dataset card on [Hugging Face](#).

Table 2: Snapshot of reported distribution.

Dimension	Values (examples)	Counts
Food type	Homemade / Restaurant / Raw / Packaged / Other	46,555 / 35,461 / 9,357 / 8,354 / 273
Authenticity (<code>camera_or_phone_prob</code>)	0.9 / 0.85 / 0.8 / 0.7 / 0.6	200 / 161 / 47,879 / 51,629 / 131

3 Codatta Protocol

3.1 Challenges of legacy data markets

Knowledge-rich labeling and real-world data sourcing are expensive to acquire and in high-demand for improving foundational language models. Buyers face high and uncertain upfront costs, opaque provenance, uneven quality assurance and limited auditability or compliance guarantees. Contributors receive relatively low flat fees, forfeit attribution, and capture no upside from downstream reuse, which depresses long-term participation, skill development and discouragement for skillful experts to take part-time contribution tasks. Operationally, static rule checks police formatting

rather than truthfulness, evidence is rarely captured in a reusable form, and usage cannot be reliably traced once data leave the platform.

3.2 Solution: Royalty-based Payment for Data Access.

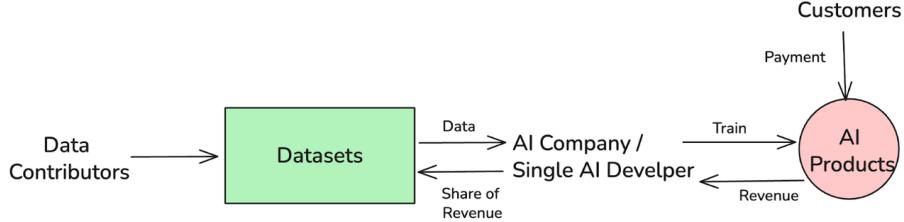


Figure 4: Royalty-Based Payment Flow in the Codatta Protocol. Usage-linked revenue sharing from AI products back to data contributors (wallet-linked provenance; quality/usage-weighted payouts).

In Codatta, royalties link contributor rewards to realized product value. Data contributors supply items that enter a curated dataset; AI developers or companies license and train on this dataset to build models and applications. When customers pay for products derived from the dataset—directly through licenses or indirectly through model use—a pre-agreed share of that revenue flows back to the dataset and is distributed to contributors or data owners according to provenance records and quality-/usage-weighted splits. This arrangement reverses the one-off, up-front payment paradigm: buyers assume less risk because costs scale with actual use, while contributors retain ongoing upside as their data continue to power deployed AI products. Figure 4 illustrates this royalty-based payment flow.

Economically, the royalty model converts data acquisition from a large, upfront capital expense into a usage-based operating expense: buyers pay only as models trained on the dataset generate value. This lowers prior risk, allows small pilots before scale, and ties spend to actual adoption and performance rather than speculative volume. For contributors, recurring royalties reward durable quality and sustained participation, creating a feedback loop in which better data lead to better products and, in turn, higher payouts distributed via wallet-linked provenance.

3.3 Privacy-preserving transparency for data assetification

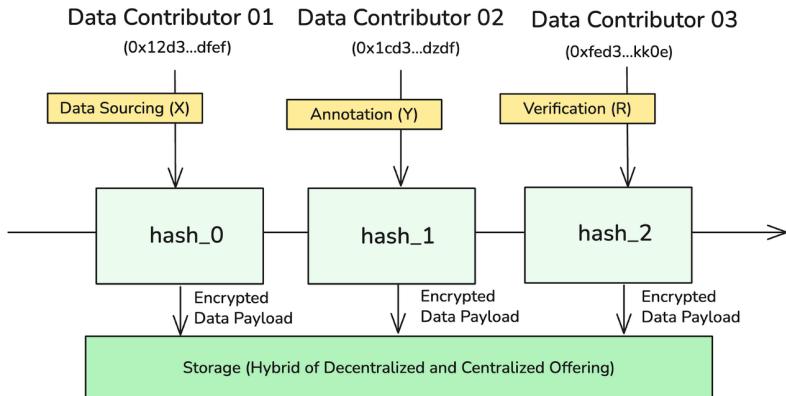


Figure 5: Blockchain-Based Data Contribution Ledger with Privacy-Preserving Transparency. A decentralized attestation layer anchors sourcing, annotation, verification, and adoption events as immutable hashed records, while encrypted payloads are managed in a hybrid decentralized-centralized storage architecture. This preserves verifiable provenance for attribution and royalties, without exposing sensitive data.

A royalty-based market only works if provenance and usage can be trusted without exposing the underlying data. To meet this requirement, we treat each contributed record as a traceable digital asset with two complementary properties: public **attestability** of its existence, authorship, and evolution; and **confidentiality** of its payload for research or commercial use. Attestability enables

fair attribution and accurate royalty allocation; confidentiality preserves contributor privacy and protects the economic value of the dataset.

Our design employs a blockchain-anchored contribution ledger, as depicted in Figure 5. Each sourcing, annotation, verification, and adoption event is recorded in a secure ledger tied to contributor wallets and time-stamped. Periodically, compact cryptographic commitments to batches of these events are anchored to a public chain, making the history tamper-evident and independently checkable without placing raw data on-chain. This provides non-repudiation—contributors can prove inclusion and credit—while keeping operational cost and sensitive content off the ledger.

Payloads (images, evidence screenshots, review notes) are stored in a hybrid architecture. Publicly verifiable metadata—content fingerprints, dataset/version identifiers, role credits—remain visible; the corresponding files are encrypted and gated behind controlled access for research or licensed use. Selective disclosure allows auditors or partners to verify specific claims (for example, that a nutrition label supports a calorie value) without broad data release. The result is **privacy-preserving transparency**: anyone can verify who contributed what and when, while only authorized parties can view the data itself.

This infrastructure enables **assetification** of the corpus. Because each record has a stable identity, provenance, and verifiable usage trail, royalties can be computed from observed adoption (in releases, training runs, or licenses) and distributed to credited wallets. At the same time, contributors retain public attribution for their work, and buyers gain an auditable chain of custody for compliance. Practically, we are rolling this out iteratively: today’s system uses verifiable off-chain records with periodic on-chain attestations; our ongoing “Booster” campaign extends this by publishing an immutable, publicly checkable list of contributions to honor ownership at scale. As the attestation process matures, more of the accounting can be automated, but the core guarantees—traceability, integrity, and controlled access—remain the backbone of the business model.

3.4 Implementation Status and Commercialization Plan

The decentralization of the data infrastructure is in active development rather than fully realized. Our present system anchors provenance and usage with verifiable off-chain records and periodic public attestations; the end-state—contract-mediated revenue pools, on-chain distribution, and community governance—remains on the roadmap. The recent Binance Wallet campaign should be read as a short-term accelerator to stress-test the royalty concept at scale: it provided wallet-linked onboarding, fast growth in contributions, and an initial public basis for attributing ownership without exposing underlying data.

As the dataset transitions from research release to commercial use, royalty sharing will be executed against the contributor registry established during the campaign. In practice, commercialization events will fund a payout pool, and distributions will reference the recorded contributor addresses and role credits. Public attestations will document which contributions were included, while detailed payloads remain access-controlled to preserve privacy and commercial value. This approach preserves the core guarantees—traceable attribution, auditable usage, and fair reward—while allowing the decentralization stack to mature incrementally.

4 Experiments

4.1 Overview and setup (quick study)

We run a minimal, decision-oriented experiment to obtain an early signal of data value. The primary task is **image-based nutrition prediction** (kilocalories). We compare two foundation models—**ChatGPT-4o** and **Qwen-Max**—in their **out-of-box** form and after supervised fine-tuning (**SFT**) on **MM-Food-100K**. Data are split **80/10/10** (train/validation/test), stratified by cuisine and source type (packaged, chain restaurant, homemade/street); near-duplicates are removed at ingest via perceptual hashing. Schema and check-logic versions associated with each item are fixed to a frozen release index for reproducibility.

The fine-tuning process is intentionally lightweight. For SFT, we use a simple regression head on the shared vision-language backbone. We optimize a Huber/MSE objective with mixed precision, fixed augmentations, and early stopping based on validation MAE. To attribute all performance differences to the data itself, not optimization variables, we keep seeds, batch sizes, and learning rates constant across all runs. The test split is identical for all models.

We design our evaluation to be task-specific:

- **Dish name, ingredients, and cooking method:** We use a text-similarity-based win rate. In each comparison between two models (e.g., base vs. fine-tuned), we report the percentage of test cases where one model’s output is judged to be more accurate or comprehensive.
- **Kilocalories:** We use standard regression metrics: **MAE** (**M**ean **A**solute **E**rror), **RMSE** (**R**oot **M**ean **S**quare **E**rror), and **R**².

4.2 Results

The following tables present our findings, illustrating a clear pattern: SFT-ChatGPT-4o consistently outperforms all others, followed by the statistically similar ChatGPT-4o base and SFT-Qwen-Max, with Qwen-Max base trailing behind. The win rate comparison for categorical predictions is shown in Table 3, and the regression metrics for kilocalorie prediction are detailed in Table 4.

Table 3: Win Rate Comparison for Categorical Predictions

Model	Dish Name (Win Rate)	Ingredients (Win Rate)	Cooking Method (Win Rate)
Qwen-Max (Base)	42.1%	39.8%	41.3%
Qwen-Max (SFT)	57.9%	60.2%	58.7%
GPT-4o (Base)	48.9%	48.6%	48.8%
GPT-4o (SFT)	51.1%	51.4%	51.2%

Table 4: Regression Metrics for Kilocalorie Prediction

Model	MAE (kcal) ↓	RMSE (kcal) ↓	R ² ↑
Qwen-Max (Base)	126.5	185.3	0.521
Qwen-Max (SFT)	104.2	154.5	0.638
GPT-4o (Base)	98.7	148.1	0.685
GPT-4o (SFT)	95.8	144.3	0.706

Observations. Fine-tuning on MM-Food-100K significantly improves performance. For Qwen-Max, fine-tuning reduces calorie **MAE by approximately 17.6%** (126.5 → 104.2 kcal). The **RMSE** sees a larger reduction of **16.6%** (185.3 → 154.5 kcal), and the **R²** value **improves by 22.4%** (0.521 → 0.638). This pattern of improvement across the metrics demonstrates the dataset’s value in enhancing predictive accuracy and reducing large errors. The win rates for dish name, ingredients, and cooking method also show substantial gains of 15.8, 20.4, and 17.4 percentage points, respectively. For ChatGPT-4o, the gains are smaller but still notable: calorie MAE drops by ∼2.9% (98.7 → 95.8 kcal), RMSE by ∼2.6% (148.1 → 144.3 kcal), and R² improves by ∼3.1% (0.685 → 0.706). We observe the largest relative gains on homemade/street items, where the dataset’s structured portion cues (Level 4) and evidence-backed nutrition (Level 2) provide the most valuable additional signal.

4.3 Discussion

These results support the central claim that **human, evidence-linked supervision** provides value beyond generic pretraining. The improvement pattern is consistent across two model families and persists when optimization is held constant, indicating that the gains derive from the data rather than from training tricks. The rough parity between ChatGPT-4o (base) and SFT-Qwen-Max underscores a practical point: access to a strong foundation model is helpful, but targeted supervision can lift a weaker base to comparable accuracy on domain tasks. Conversely, SFT-ChatGPT-4o achieves the best overall performance, suggesting that high-capacity models benefit most from schema-rich supervision.

This is an intentionally minimal study designed to deliver a fast, actionable signal. Next iterations will separate task-specific heads for dish/cuisine, ingredients, and portion; report per-domain strata (packaged, chain, homemade/street); and expand to additional model families and training sizes to characterize scaling. We will also link intake quality to outcomes by weighting

training examples with the dataset’s QA scores. All future phases will reuse the same frozen indices and configuration templates to keep comparisons clean and auditable.

5 Discussion and Future Work

Data value and limits. MM-Food-100K delivers clear gains on image-based nutrition prediction and improves related signals (dish accuracy, portion error). The benefit comes from evidence-linked fields—nutrition proof and portion/scale cues—that generic pretraining lacks. Limits remain: homemade/street portions and calories carry uncertainty; some cuisines are under-represented; menus drift seasonally; filled-rates vary by field. Our release balances openness and incentives: 100K is public for research; the retained corpus is licensable with revenue-sharing to contributors, converting data spend from a large upfront cost to usage-based payment.

Protocol and royalties. The Codatta protocol targets the dual-market failure of legacy platforms by paying for **adoption**, not just task completion. Wallet-linked provenance and quality/usage-weighted splits connect data value to rewards. Implementation is iterative: accounting runs off-chain with public attestations today; the Binance Wallet campaign established a contributor registry that will be honored during commercialization; fuller decentralization (contract-based pools and automated disbursement) is on the roadmap.

Next studies and the human-AI loop. This quick study is a first signal. We will extend to dish/cuisine, portion, and ingredients with task-specific heads; add more model families and training sizes; and run ablations on schema depth and QA-weighted training, all on fixed splits for clean comparisons. Weekly instrumentation (A, AC, AD, AC/A, AD/AC, review time, cost/AD) guides updates to pre-accept gates and reviewer prompts: AI absorbs routine errors and raises first-pass acceptance; human audit focuses on edge cases. This interaction moves the cost-quality frontier outward and provides leading indicators for downstream model gains.

References

- [1] Bossard, L., Guillaumin, M., & Van Gool, L. (2014, September). Food-101-mining discriminative components with random forests. In *European conference on computer vision* (pp. 446–461). Cham: Springer International Publishing.
- [2] Salvador, A., Hynes, N., Aytar, Y., Marin, J., Ofli, F., Weber, I., & Torralba, A. (2017). Learning cross-modal embeddings for cooking recipes and food images. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 3020–3028).
- [3] Achiam, J., Adler, S., Agarwal, S., Ahmad, L., Akkaya, I., Aleman, F. L., ... & McGrew, B. (2023). Gpt-4 technical report. *arXiv preprint arXiv:2303.08774*.
- [4] Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., ... & Sutskever, I. (2021, July). Learning transferable visual models from natural language supervision. In *International conference on machine learning* (pp. 8748–8763). PMLR.
- [5] Yin, S., Fu, C., Zhao, S., Li, K., Sun, X., Xu, T., & Chen, E. (2024). A survey on multimodal large language models. *National Science Review*, 11(12), nwae403.
- [6] Meyers, A., Johnston, N., Rathod, V., Korattikara, A., Gorban, A., Silberman, N., ... & Murphy, K. P. (2015). Im2calories: towards an automated mobile vision food diary. In *Proceedings of the IEEE international conference on computer vision* (pp. 1233–1241).
- [7] Hu, E. J., Shen, Y., Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S., ... & Chen, W. (2022). Lora: Low-rank adaptation of large language models. *ICLR*, 1(2), 3.
- [8] Dettmers, T., Pagnoni, A., Holtzman, A., & Zettlemoyer, L. (2023). Qlora: Efficient finetuning of quantized llms. *Advances in neural information processing systems*, 36, 10088–10115.
- [9] Xin, Y., Luo, S., Zhou, H., Du, J., Liu, X., Fan, Y., ... & Du, Y. (2024). Parameter-efficient fine-tuning for pre-trained vision models: A survey. *arXiv e-prints*, arXiv–2402.
- [10] Minaee, S., Mikolov, T., Nikzad, N., Chenaghlu, M., Socher, R., Amatriain, X., & Gao, J. (2024). Large language models: A survey. *arXiv preprint arXiv:2402.06196*.

- [11] Liu, J., Peng, S., Long, C., Wei, L., Liu, Y., & Tian, Z. (2020, March). Blockchain for data science. In *Proceedings of the 2020 2nd International Conference on Blockchain Technology* (pp. 24–28).
- [12] Wang, K., Dong, J., Wang, Y., & Yin, H. (2019). Securing data with blockchain and AI. *IEEE Access*, 7, 77981–77989.
- [13] Theodore Armand, T. P., Nfor, K. A., Kim, J. I., & Kim, H. C. (2024). Applications of artificial intelligence, machine learning, and deep learning in nutrition: A systematic review. *Nutrients*, 16(7), 1073.
- [14] Rieke, N., Hancox, J., Li, W., Milletari, F., Roth, H. R., Albarqouni, S., ... & Cardoso, M. J. (2020). The future of digital health with federated learning. *NPJ digital medicine*, 3(1), 119.
- [15] He, K., Mao, R., Lin, Q., Ruan, Y., Lan, X., Feng, M., & Cambria, E. (2025). A survey of large language models for healthcare: from data, technology, and applications to accountability and ethics. *Information Fusion*, 118, 102963.
- [16] Cheng, S. T., Lyu, Y. J., & Teng, C. (2025). Image-Based Nutritional Advisory System: Employing Multimodal Deep Learning for Food Classification and Nutritional Analysis. *Applied Sciences*, 15(9), 4911.
- [17] Muhammad, S. K., Ansari, T. A., Shabbir, B., Almagharbeh, W. T., Rehman, A., Arjmand, R., & Ghulam, A. (2025). THE ROLE OF ARTIFICIAL INTELLIGENCE IN PUBLIC HEALTH SURVEILLANCE: A POST-PANDEMIC PERSPECTIVE. *Insights-Journal of Life and Social Sciences*, 3(3 (Social)), 74–80.
- [18] Jiang, L., Qiu, B., Liu, X., Huang, C., & Lin, K. (2020). DeepFood: food image analysis and dietary assessment via deep model. *IEEE Access*, 8, 47477–47489.
- [19] Park, J. S., O'Brien, J., Cai, C. J., Morris, M. R., Liang, P., & Bernstein, M. S. (2023, October). Generative agents: Interactive simulacra of human behavior. In *Proceedings of the 36th annual ACM symposium on user interface software and technology* (pp. 1–22).

Appendix

Schema and taxonomy-design rationale

The schema and taxonomy are designed to turn heterogeneous food photos into **verifiable, model-ready supervision** while keeping collection costs predictable and quality auditible.

Design goals

- **Verifiability first:** every factual claim (e.g., calories, portion) links to **evidence** (label/menu/URL/receipt) and to **provenance** (wallet-linked contributor record).
- **Progressive enrichment:** annotations are organized into **five levels** so contributors (or AI agents) can start with low-friction L1 submissions and **upgrade** items to richer L2–L5 records.
- **Hybrid QA:** the same fields support **human**, **AI**, or **hybrid** quality checks (e.g., OCR validation, outlier flags, consensus sampling).

Why three domains

We partition sources into **Packaged**, **Chain restaurant**, and **Homemade/Street** to align evidence requirements and QA cost with a **ground-truth gradient** and a **variance gradient**. Packaged items have printed nutrition (high verifiability, low visual variance); chain items have menu nutrition (medium verifiability); homemade/street items rely on estimates and portion cues (lower external ground truth, higher variance). This split lets us tune evidence rules and sampling rates **per domain**.

Why five levels

Levels reflect **increasing effort and model utility**:

- **L1:** image + name (seed the corpus).
- **L2:** nutrition with evidence (make records verifiable).
- **L3:** visible ingredients (enable extraction).
- **L4:** portion & geometry (enable calorie/serving regression).
- **L5:** per-ingredient localization (enable detection/segmentation).

Table 5: Where the data comes from (source types)

Source type	Where it comes from	Photos to include	Nutrition source	Minimum detail level	Proof to attach	Default quality check	Common pitfalls
Packaged food	supermarketfront + packs, barcode pages	back / label	per serving / per pack	2	label photo or clear OCR	auto checks + quick human review	duplicate SKUs; old labels
Chain restaurant	brand web-site/menu, in-store boards	dish photo + menu screen-shot	per serving / menu size	2	screenshot / URL with item marked	OCR match + human confirm	menu changes; seasonal items
Homemade / street	personal photos, chef notes, receipts	dish photo	estimate / per 100g	3	short note or receipt (if any)	human-led review + model assist	portion bias; privacy; mixed plates

Table 6: What each annotation level adds

Source type	Nutrition source	Min detail level	Proof to attach	Default quality check	Common pitfalls
Packaged food	per serving / per pack	2	label photo or clear OCR	auto checks + quick human review	duplicate SKUs; old labels
Chain restaurant	per serving / menu size	2	screenshot/URL with item marked	OCR match + human confirm	menu changes; seasonal items
Homemade / street	estimate / per 100g	3	short note or receipt (if any)	human-led review + model assist	portion bias; privacy; mixed plates