Journal of Agriculture, Food Systems, and Community Development
ISSN: 2152-0801 online
https://foodsystemsjournal.org

Beyond self-report surveys: Leveraging multimodal large language models (MLLMs) for farmers market data harvesting from public digital resources

Huy Pham [a] and Yue Cui [b] *
Michigan State University

Submitted June 3, 2025 / Revised July 7, July 23, and August 28, 2025 / Accepted August 28, 2025 / Published online October 1, 2025

Citation: Pham, H., & Cui, Y. (2025). Beyond self-report surveys: Leveraging multimodal large language models (MLLMs) for farmers market data harvesting from public digital resources. *Journal of Agriculture, Food Systems, and Community Development*, *14*(4), 137–154. https://doi.org/10.5304/jafscd.2025.144.025

Abstract

Traditional farmers market research using self-reported surveys has been constrained by high costs, extended timelines, recall bias, and frequently outdated findings. To address these limitations, this study introduced multimodal large language models (MLLMs) as a scalable, cost-efficient approach to extracting farmers market data through automated processing of diverse public digital sources, including websites, social media, photographs, and government documents. This study adopted a two-step framework to extract relevant information and transform unstructured multimodal data into an analysis-ready format. Benchmarked against the Michigan Farmers Market Census and Directory (MIFMA, 2024), our framework covered 76% of their topics. The MLLMs demonstrated robust performance, achieving near-zero hallucination rates, 98% accuracy of key variables extractions, and the ability to support real-time updates. While this approach cannot capture confidential or subjective data, it paves the way for a future hybrid framework that integrates the comparative advantage of two methods: MLLMs for efficient, factual data collection and human researchers for conducting targeted surveys to capture subjective insights. This

[a] Huy Pham, Research Assistant, School of Planning, Design and Construction, Michigan State University; 552 West Circle Drive, Room 112; East Lansing, MI 48824 USA; phamdinh@msu.edu;
https://orcid.org/0009-0005-5444-4672

[b] * *Corresponding author:* Yue Cui, Assistant Professor, School of Planning, Design and Construction, Michigan State University; 552 West Circle Drive, Room 112; East Lansing, MI 48824 USA; cuiyue@msu.edu;
https://orcid.org/0000-0002-2735-8913

Conflict of Interest Disclosures
The authors declare no competing interests. No funding sources influenced the study design, data collection, or interpretation of results.

Funding Disclosure
Not applicable.

efficient, reliable, and scalable approach empowered policymakers, market managers, and researchers to dynamically monitor trends and obtain accurate, detailed, and timely data, fostering resilient and inclusive food systems. Beyond farmers markets, the applications of this adaptive framework could extend to other domains, such as public health, urban planning, and economic policy, highlighting artificial intelligence (AI)'s transformative potential for streamlining data-centric decision-making.

Keywords
multimodal large language models (MLLMs), agricultural information systems, farmers market research, local food systems, structured data extraction, public digital sources, artificial intelligence, AI

Introduction

Farmers markets serve as vital nodes within local food systems, fostering economic resilience, community engagement, and equitable access to fresh produce. Effective management and evidence-based policymaking for farmers market sectors rely on accurate, timely, and operational data, such as vendor structures, fee schedules, and participation in nutrition assistance programs. Historically, this type of data has been collected through resource-intensive survey efforts. For example, the Michigan Farmers Market Census (MFMC), conducted by the Michigan Farmers Market Association (MIFMA), required managers to answer over 50 questions during hour-long sessions and provided financial incentives to participants exceeding US$50 per response (MIFMA, 2024). While these instruments have yielded valuable insights, their manual nature leads to standard survey limitations: high costs, recall bias, incomplete records, extended timelines, and rapid obsolescence. For instance, the 2021 MFMC report was released almost three years after the data collection (MIFMA, 2024).

Previous studies (Kreuter, 2013; Low et al., 2015) indicate that a significant amount of survey-collected data, especially factual information, might be optimized through the use of existing administrative records, thereby minimizing redundancy and improving accuracy. A case in point is the collection of data on food assistance program infrastructure in Michigan farmers markets. MFMC collected this data by surveying market managers and relying on their recall (MIFMA, 2024). However, a more reliable alternative existed: publicly available, compliance-grade records from the Michigan House of Representatives (2023). As a verified and complete dataset, it guarantees 100% coverage without any response rate concerns. Consequently, when fact-based content (e.g., program participation, funding allocations) exists in primary sources, continued data collection through surveys constitutes both methodological redundancy and resource misallocation (Meyer et al., 2015).

Recently, the richness of records that are available has significantly reinforced the value of document-based data collection. The breadth of publicly accessible data has grown exponentially since the early 2000s (Aziz et al., 2024; Meyer et al., 2015). This shift has lifted the technological constraints of traditional resourse-intensive survey methods. Nowadays, farmers markets utilize multiple digital platforms for customer engagement (e.g., messaging through social media) and operational administration (e.g., vendor application and agreements, seasonal schedules, bookkeeping DUFB transactions). Market rules, vendor agreements, and regulatory filings contained operationally critical details, from fee structures to product source standards, with precision that may exceed the reliability of the traditional survey responses. These fragmented but rich data streams, which were previously too cumbersome to analyze effectively, can now be efficiently parsed and structured using advanced natural language processing methods, including MLLMs.

Accordingly, this study sought to answer two important questions: (1) Can MLLMs systematically extract comprehensive farmers market data from public digital sources with greater efficiency and accuracy than traditional surveys? (2) To what extent can document-derived data replace traditional survey content? These questions were investigated through a Michigan-based case study, which used Google Gemini-2.0-Flash in a two-step framework to retrieve and structure data from 2,961 public documents (websites, social media, and government records) and approximately 3,000 application programming interface (API) calls for Michigan farmers markets. This methodology combined iterative prompt engineering with domain-specific taxonomies to convert unstructured inputs into analyzable data. Its effectiveness and coverage were assessed against the MFMC and the Michigan Farmers Market Directory (MFMD).

This study makes four major contributions. First, it pioneers a practical framework combining MLLMs with domain-specific taxonomies to systematically harvest structured data from fragmented public sources (e.g., social media, regulatory documents). The approach provides a replicable, cross-disciplinary model for domains such as urban planning, public health, and local food systems while reducing reliance on manual surveys. Second, the framework significantly expands the depth and contextual richness of collected farmers market data by capturing variables often omitted in traditional surveys, such as market stall management and vendor categories. Third, while preserving real-time monitoring capabilities, this automated method allows for smooth scaling, boosting the feasibility of large-scale initiatives both geographically (from Michigan to multistate and nationwide implementation) and temporally (from single-year snapshots to continuous multi-year analysis). Finally, the study positions MLLMs as a strategic addition to regular surveys, not a substitute. While the MLLMs cannot capture confidential data or subjective insights, they can enhance routine survey content, ultimately supporting more complete and policy-relevant datasets.

Literature Review

The following literature review examines the current state of farmers market data collection practices and the emerging potential of larges language models for automated data extraction. The analysis also covers established frameworks for evaluating model performance and accuracy metrics for document-based information extraction.

Methodologies in Farmers Market Data Collections

Farmers market research has largely relied on the same methodological foundation over the past three decades, primarily centered on qualitative interviews and focus groups and quantitative surveys. This dual approach was established through foundational studies conducted in the 1990s and early 2000s, including the work of C. Brown (2003), Hinrichs (2000), Holloway et al. (2007), and Guthrie et al. (2006). Despite survey design and research scope having been refined, the field has seen limited progress in adopting innovative data collection methods. Recent large-scale efforts, such as the USDA's 2019 National Farmers Market Managers Survey (USDA, 2020) and the Michigan Farmers Market Census (MIFMA, 2024), continued to rely on traditional approaches to a certain degree, namely, mail-in questionnaires, online surveys, phone surveys, etc. In parallel, research in the broader agricultural and rural development literature suggested that factual operational data, often gathered through surveys, may be more accurately and efficiently obtained through administrative records and structured public data sources (e.g., Low et al., 2015).

General Applications of LLMs in Automated Data Extraction

The application of large language models (LLMs) for automated data extraction has rapidly advanced across various scientific domains. In healthcare, LLMs have been employed to extract clinical information such as diagnoses and medications from unstructured patient notes with high accuracy, outperforming traditional rule-based systems (Gu et al., 2025; Siepmann et al., 2025). In the fields of chemistry and materials science, models have been used to mine complex properties from research literature, significantly reducing the need for manual data curation (Schilling-Wilhelmi et al., 2025). Social science researchers have applied LLMs to clean noisy historical records (Schwitter, 2025) and analyze crowdsourced geographic data (Huang et al., 2025). GPT-3 and similar models were able to extract scientific relationships into structured formats like Java Script Object Notation (JSON), as shown in materials chemistry by Dagdelen et al. (2024). Despite these developments, there is currently no documented application of LLM-based data extraction methods in local food system or farmers market research, where traditional surveys still dominate data collection practices.

LLMs and MLLMs in Document Analysis: Techniques and Performance Benchmarking

Extracting structured data from documents, including text, forms, images, and tables, has historically been challenging due to diverse layouts and unstructured text. Recent progress in LLMs and MLLMs research has provided promising solutions by interpreting text alongside its visual layout and semantics, mimicking human-like understanding (P. Liu et al., 2023). MLLMs have improved this capability by directly analyzing scanned documents using both textual and visual cues. However, these models depend heavily on well-crafted prompts to focus on relevant information and deliver accurate outputs, particularly for documents with high variability or novel structures. Research has shown that well-designed prompts can achieve reliable results without extensive retraining (Bommasani et al., 2022; T. Brown et al., 2020). Key strategies for optimizing prompt performance include specifying roles, using chain-of-thought prompting, incorporating exemplars, and employing synonymous terms (T. Brown et al., 2020; Lao et al., 2023; Wu et al., 2024; Y. Zhou et al., 2023). Despite their strengths, LLMs and MLLMs might not perform well when extracting multiple fields or answering numerous questions simultaneously from complex or semi-structured documents since they often struggle with reasoning through a large amount of content, especially when handling several types of information at the same time. A cutting-edge two-step approach might be a solution: first, generating plain-language interpretations, then converting them into structured schemas with a follow-up prompt (Wu et al., 2024; Y. Zhou et al., 2023). This stepwise method has outperformed single-shot extraction, aligned with human analytical processes, and been widely adopted in advanced document analysis systems (R. Liu et al., 2023).

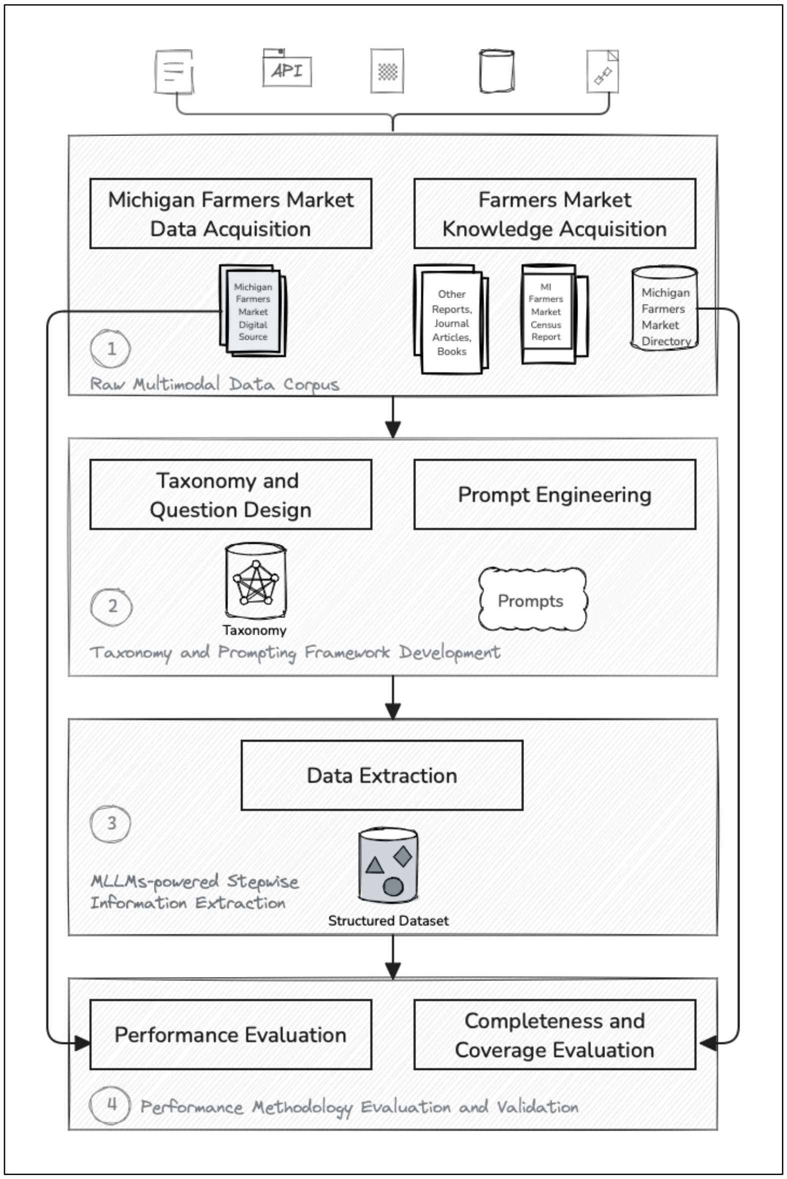
LLMs use prompt-driven reasoning to generate semantically rich outputs, though these responses can occasionally exhibit inconsistencies or gaps. Recent research has prioritized four core evaluation dimensions to measure the models' performance: accuracy, consistency, completeness, and hallucination rates. Accuracy is typically assessed using metrics such as Exact-Match, field-level similarity, token-level precision, question-answer correctness, and manual validation (Kumar, 2024; Ouyang et al., 2025; Sushil et al., 2024). Consistency refers to the model's ability to yield stable responses across identical or similar prompts and is evaluated using prompt sensitivity, output variance, and drift analysis (Greyling, 2024; Joshi, 2025; Lai et al., 2025). Techniques such as the "LLM-as-a-judge" framework also provide scalable solutions for benchmarking consistency in large-scale evaluations (Shah, 2024). Completeness assesses the model's ability to extract all relevant information, with coverage and recall as common metrics. Research has highlighted the role of prompt decomposition and stepwise querying, essentially a divide-and-conquer approach that breaks complex tasks into smaller, more manageable step-by-step tasks, in enhancing completeness, especially for complex documents (P. Liu et al., 2023; Rasool et al., 2024; Zhang et al., 2024). Hallucination, which refers to cases when AI systems generate unverifiable or fabricated information, can be detected from human verification to automated systems such as span/claim detection, retrieval-augmented generation, and knowledge graph comparison (Elchafei & Abu-Elkheir, 2025; Lee & Yu, 2025; S. Liu et al., 2025; Sansford et al., 2024; S. Zhou et al., 2022). These strategies are often integrated into hybrid workflows that combine automated checks with manual oversight to mitigate fabricated or unsupported outputs (Zhang et al., 2024). Together, these studies have established a robust framework for evaluating and improving LLM performance in document-based information extraction.

Methodology

This study used MLLMs to extract information related to farmers markets from digital sources. The workflow is illustrated in Figure 1.

Given the analytical demands posed by diverse multimodal content such as websites, social media posts, images, lengthy texts, and tables like farmers market bylaws, this study chose Gemini-2.0-Flash as the core MLLM based on three features: minimal hallucinations (Hughes et al., 2024), advanced reasoning (Jain, 2025), and scalable multimodal document handling (Google DeepMind, 2024).

Figure 1. Proposed Data Extraction Pipeline



Raw Multimodal Data Corpus

This study aimed to collect publicly available digital sources for Michigan farmers markets. All data acquisition activities adhered to terms of service permissions (e.g., robots.txt files) to ensure ethical data reuse (Hacker & Mason, 2003; Powell et al., 2022). No personally identifiable information was collected or stored, even if it was incidentally encountered during processing. Examples in this article used anonymized references (e.g., market name, detailed address) to prevent re-identification while preserving analytical value (Zimmer, 2018).

This study also did not interact with market personnel directly in order to comply with human-subject research ethics. Table 1 lists Michigan farmers market data sources. In addition, published farmers market surveys, reports, journal articles, and books were reviewed and synthesized to serve as knowledge bases for taxonomy design and prompt framework development.

Taxonomy and Prompting Framework Development

This work began by using Gemini-2.0-Flash to conduct unsupervised knowledge extraction from digital sources detailed in the Raw Multimodal Data Corpus section. The major goal was to identify key thematic areas (e.g., vendor management), entities (e.g., product categories), and contextual relationships (e.g., seasonal product availability). The taxonomy development followed an iterative, structured process. Initial themes and categories emerged from raw textual sources and farmers market knowledge bases. These were then systematically reviewed and refined to generate a coherent set of structured extraction prompts analogous to survey questions. To validate coverage and accuracy, the emerging taxonomy was compared against existing survey instruments to identify conceptual gaps or inconsistencies. New themes were developed to fill the gaps detected during this process. Iterative cycles refined the taxonomy through repeated revisions and testing against the original source documents to maintain interpretive accuracy and empirical relevance. Revisions continued until a stable, conceptually coherent taxonomy was achieved. This finalized taxonomy formed the foundation for developing structured prompts for the MLLM to extract, interpret, and organize relevant content.

One example of this work was the product eli-

Table 1. Public Data Sources for Michigan Farmers Markets

Source	Content	Data Format
Farmers market websites	Operational details, governance policies, vendor applications, annual reports, etc.	Markdown, PDF, Image, Word Document
Social media	Event announcements, community engagement metrics, promotional materials	Markdown, Image
Local news	Market establishment announcements, operational updates	Markdown
Lifestyle magazines, blogs, and travel websites	Curated listings, visitor perspectives, and promotional content positioning markets as cultural attractions and community highlights	Markdown
Government datasets	Nutrition program listings and regulatory documents	Markdown, PDF, Word Document, CSV
Local websites	Events and content related to farmers markets	Markdown, PDF, Image, Word Document
Events websites	Market schedules, locations, organizer details	Markdown
Additional public records	Geographic context (e.g., downtown), demographic data, urbanization levels (census tracts), population density, land use, zoning, public transportation, restroom availability	JSON (via API)

gibility issues for vendor management. The data extraction process started with unsupervised knowledge extraction from the sources and identified four categories (Table 2), which was a significantly different outcome from the MFMC (MIFMA, 2024).

The MFMC approach employed a purpose-driven, traditional survey framework comprising 18 static categories, which integrated product type, vendor type, and origin requirements. On the other hand, the MLLM method utilized discovery-driven data mining to create adaptive categorizations from unstructured document content. The process began by identifying granular product names (e.g., Strawberry) and was then applied to multiple categories, including product type (Fresh produce), vendor role (Direct producer), resale policy (Producer only), and origin requirements (within 50 miles). This method can reach analytical precision that the rigid frameworks (e.g., MFMC, in this example) might not be able to attain. These categories allow researchers to examine farmers markets from a variety of angles, such as (1) tracking seasonal dynamics through product-type ratios (Michigan markets move from early-season microgreens and nonperishables to peak-season vegetables and finally to autumn value-added products like apple cider), (2) quantifying authenticity threats (should farmers markets have fresh produce vendors?), and (3) classifying market typologies (flea markets have loose regulations and little fresh produce; community markets incorporate a variety of vendors with moderate produce; true produce markets prioritize direct-producer sales). It also enables the analysis of how policy enforcement, vendor composition, and targeted customer engagement initiatives collectively contribute to sustaining farmers markets' missions: promoting healthy local food, strengthening community connections, and supporting regional economies (a finding discovered through this MLLM analysis).

Given the complexity of the taxonomy-derived questions, a modular approach was adopted to prompt design by breaking schemas into smaller, related subschemas (e.g., separating "resale policy" from "product origin requirement") to improve output consistency and mitigate hallucinations. This methodology grouped questions into thematic clusters (e.g., vendor types, space or stall management), paired them with tailored instructions (e.g., are there rules or policies for competitive pricing or product dumping), and guided the model to extract specific answers.

Similar to the above taxonomy development,

Table 2. Product Eligibility Frameworks: MFMC vs. MLLM

Items	MFMC	MLLM
Data Collection	Purpose-Driven Survey	Discovery-Driven Data Mining
Categories of Products and Vendors	(1) Michigan farm products, sold by the producer or their representative (2) Honey (3) Cottage foods (4) Cut flowers (5) Food items not produced on farms (6) Plants (7) Handmade body care items (8) Handmade crafts (9) Prepared foods meant for immediate consumption (10) Food trucks (11) Michigan farm products, can be resold (12) Wild-harvested and/or foraged foods (13) Wine, hard cider, mead (14) Non-Michigan farm products (15) Non-handmade items, including multilevel marketing items (16) Topical CBD items (17) Pet supplies and treats (18) Services (e.g., knife sharpening)	**Product Category:** (1) Fresh produce (2) Bread and baking goods (3) Milk and dairy (4) Meat and seafood (5) Beverages (6) Condiments and sauces (7) Food trucks (8) Other food products (9) Plants and seedlings (10) Pet essentials (11) Crafts and artisan items (12) Services (13) Information table (14) Entertainment and activities (15) Others **Vendor Category:** (1) Direct producers (Farmers/Growers - Raw agricultural products) (2) Value-added producers (3) Prepared food vendors (4) Artisans and crafters (5) Nonprofit and community groups (6) Entertainment vendors (7) Other vendors **Resale Policy by Product Types:** (1) Producer only: Vendors must grow, raise, produce, gather, or create all products offered for sale at their booth (2) Allow limited reselling and maximum allowed [xx] percentage of resold products (3) No limitation for reselling **Geographic and Origin Requirements by Product Types** (1) Required radius ([xx] miles) for "locally sourced" products (2) Limited regions (e.g., Michigan Only, Upper Peninsula Only, Adjacent counties or states) (3) No geographic sourcing restrictions apply

the design of the prompts involved a systematically iterative refinement process with human-in-the-loop. The prompts were tested against a sample of documents covering diverse scenarios (e.g., reports, websites, vendor application forms, social media content, digitized survey responses, and market photos) and varied formats (pdf, html, docx, txt, md, jpeg, png, and others), as well as corner cases, such as lengthy texts, and documents with noises (e.g., news articles cluttered with ads). During this process, prompt engineering practices like chain-of-thought and few-shot prompting were adopted to improve the efficiency. The adjustments started with the feedback from an initial model output based on the taxonomy. The revised prompts were then embedded with these modifications, and the process was iterated until the prompts consistently performed well across all test cases in terms of accuracy and avoiding hallucination.

Stepwise Information Extraction

This study used a two-step structured data extraction process to improve the accuracy and output consistency of farmers market data extraction. First, we adopted Gemini-2.0-Flash as the MLLM to pinpoint and retrieve relevant plain language information from various sources. Prompts described in the Taxonomy and Prompting Framework Development section guided this process to focus the model on extracting the truly relevant, but unstructured, details. Then, the information extracted in the first step was combined with a predefined schema and specific prompts to transform the unstructured natural language answers into a structured JSON dataset that could handle both the flat tables in traditional surveys and hierarchical structures in data analysis aspects. The following example illustrated this process by transforming an unstructured social media post into analyzable data, using a Facebook post from a Michigan farmers market:

> We are thrilled to announce that vendor applications for the first [market name] are now open! The market will be held every other Saturday 9 am - 1 pm starting June 1st running through August. Please Like and Share this post so that we can reach as many folks as possible with this news. We are trying to increase access to fresh food in the [community name] area of our neighborhood and this is a big step in the right direction! Location: City High School Main Parking Lot, [address]. To apply for a vendor spot please complete the application here: [application link]

This message mixed announcements, operation details, and promotions in natural language. Table 3 illustrates how this two-step method was used to process the Facebook post with the designed framework. In the first step for contextual information retrieval, the system demonstrated its ability to understand the nuances of human language by locating pertinent data in plain language.

For example, it identified the context of phrases like "*first [Market Name]*" (indicating an inaugural year), "*every other Saturday 9 am - 1 pm starting June 1st running through August*" (an operating schedule), and "*Location: [school name] High School Main Parking Lot, [address]*" (a location description with detailed address). This step accurately isolated contextually significant details embedded in free-form text and focused on extracting the truly relevant details, whether structured or not. In Step 2, the MLLM applied the Taxonomy and Prompting Framework Development's structured prompts to organize unstructured data. For example, the model inferred that "*June 1st running through August, 2024*" corresponded to the date range 2024-06-01 to 2024-08-24. It interpreted "*Every other Saturday*" as implying a biweekly frequency on Saturday. "*High School Main Parking Lot*" was inferred to represent a site type as an educational institution. By using contextual reasoning, the model distinguished explicitly stated facts from implied content. Table 3 (Step 2 Output) shows the resulting JSON-formatted farmers market data, structured for seamless integration in statistical analysis and other applications.

Topic Coverage and Completeness Evaluation

Following USDA (2020), this study defined a farmers market as a venue that includes two or more vendors selling agricultural products directly to customers at a common, recurrent physical location. Because location is integral to this definition, markets that relocate seasonally (e.g., moving from an outdoor summer site to an indoor winter site) were considered separate market locations. To check how complete the data are and how well they cover topics related to farmers market research, this study used datasets from Michigan to compare two sources: (1) the MFMD (MIFMA, n.d.), a dataset for a consumer-oriented directory that was up to date as of May 15, 2025, matching our data collection deadline; (2) the MFMC (MIFMA, 2024), a management-focused survey report for Michigan farmers markets. The MFMD dataset contained a completed array of variables that permitted a direct and rigorous assessment of data availability and completeness against the MLLM dataset. However, a direct comparison with MFMC was not feasible for two reasons: (1) the MFMC report was published as a summary report without survey instruments and methodological documentation, and (2) the report was organized

Table 3. Example of the Two-Step Data Extraction Procedure

Data Item	Source Text	Answers from Step 1	Output from Step 2
Metadata	*Facebook, posted in 2024*	2024	{"post_year": 2024 }
Market Establish Year	*Vendor applications for the first [market name] are now open*	Market established in 2024	{"start_year": 2024, "#year": 1 }

Operation Schedule	*Every other Saturday 9 am –1 pm starting June 1st running through August.*	frequency: every other Saturday;
		time: 9 am – 1 pm;
		start: June 1st, 2024;
		end: August, 2024
			{
			"start_date": "2024-06-01"
			"end_date": "2024-08-24",
			"operation_month": [6, 7, 8],
			"weekdays": "Saturday",
			"frequency": "bi-weekly", "operation_time": {
			"start": "9:00",
			"end": "13:00" }
			}

Market Mission	*Increase access to fresh food in the community*	Increase access to fresh food	{
			"access_to_local_food": "yes",
			"connect_community": "",
			"boost_local_economy": "",
			"other": ""
			}

Location	*City High School Main Parking Lot, [address]*	Site: High School Main Parking Lot;
		Address: [address]
			{
			"site": "education institution",
			"surface": "parking lot",
			"street": "[street name]",
			"city": "[city name]",
			"state": “Michigan",
			"zipcode": "[zipcode] "
			}

Vendor Application	*To apply, complete application: [application link]*	Vendor application available at: [application link]	{
			"application_form": "yes"
			}

around thematic areas, which provided an overall response rate without itemized response metrics at the topic level. Therefore, the comparative analysis in this study adopted an “apples-to-apples” approach and concentrated on the topics explicitly reported within the MFMC.

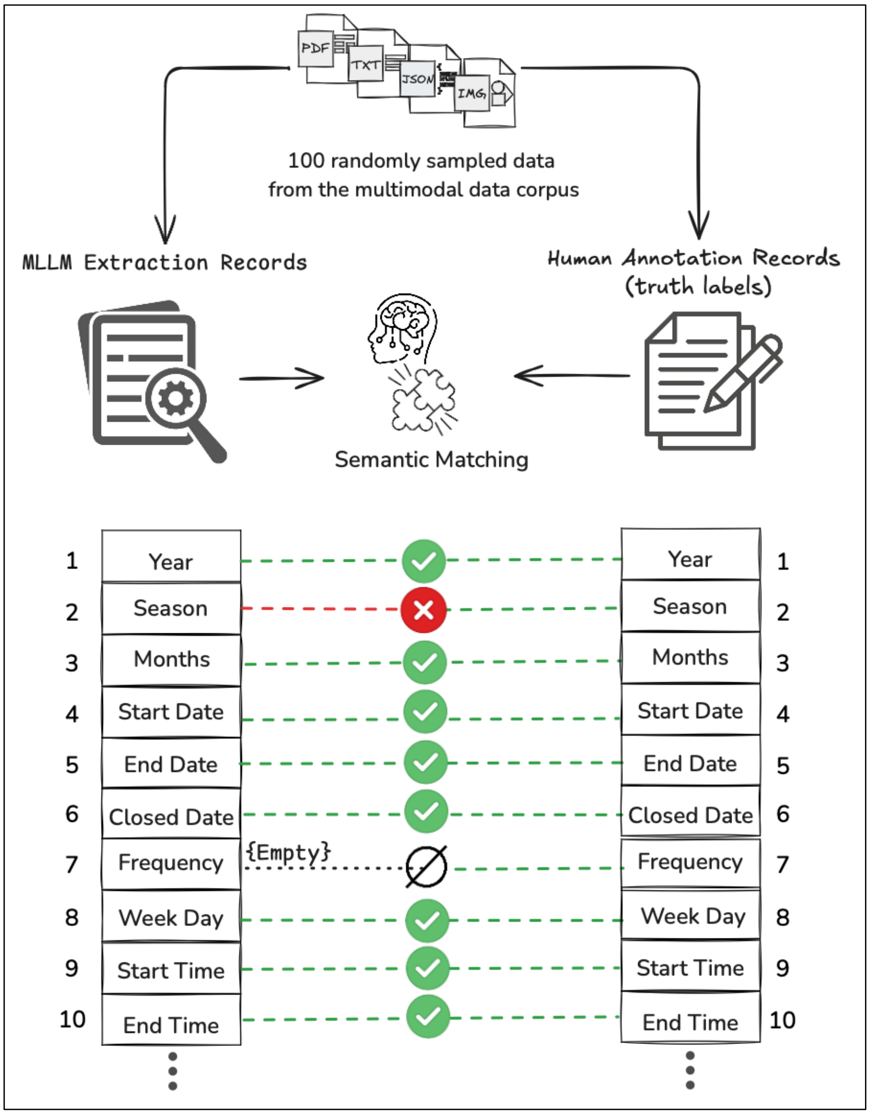
Model Performance Evaluation

The evaluation measured model performance by accuracy, output consistency, and hallucination rates, based on a variety of data extraction tasks typical of farmers market data processing. Each task represented a defined information retrieval objective where the MLLM was prompted to extract a predefined set of data items. This framework included five distinct tasks reflecting common farmers market data types, testing model performance across modalities and complexities:

(1) Address Retrieval Task: Assessing basic entity extraction.
(2) Operational Schedule Retrieval Task: Evaluating the extraction of structured temporal data.
(3) Food Assistance Programs (FAP) Retrieval Task: Testing the identification of specific program participation.
(4) Vendor Fee Structure Retrieval Task: Assessing the ability to extract complex, often nuanced, financial details.

(5) Product from Images Task: Specifically testing the MLLM's multimodal capability to interpret visual data in conjunction with any accompanying text.

Figure 2. Illustration of Operation Schedule Retrieval Verification



Technically, these tasks covered a spectrum of challenges, from straightforward directory facts (e.g., street and city variables in Address Task) to multimodal scenarios requiring visual-textual integration (e.g., product name in vendor images). This spectrum of difficulty helped align MLLM outputs and human annotations for a consistent and fair comparison. To systematically evaluate the performance of the proposed MLLM-based extraction pipeline, the study built a validation dataset consisting of 100 documents randomly selected for each task. Human annotators underwent training to carefully review the source documents and extract the relevant information according to the defined schema. The annotated values provided a ground truth benchmark used to assess the model's outputs.

Figure 2 illustrates the accuracy evaluation workflow (one of three performance tests), benchmarking MLLM-extracted items against human-annotated ground truth per task, where:

(1) True Positive (TP): An item correctly extracted by the MLLM that semantically matches the ground truth (represented by a green tick in the conceptual figure).
(2) False Positive (FP): An item incorrectly extracted by the MLLM, or an item extracted by the MLLM that is not present in the ground truth (represented by a red cross).
(3) False Negative (FN): An item present in the ground truth that the MLLM failed to retrieve (represented by an empty set symbol).
(4) True Negative (TN): An item correctly not retrieved by the MLLM because it is also absent from the ground truth.

The aggregated counts of TPs, FPs, and FNs across the validation dataset were then used to calculate standard metrics *Precision*, *Recall* and *F-1* score to represent the model's accuracy performance.

To gauge MLLM's reliability, the measurement incorporated hallucination rate, calculated as the number of fabricated or non-existent items retrieved by MLLM divided by the total number of items it retrieved across the validation set. In addition, this study measured output consistency to evaluate the model's stability by executing each extraction task on the validation documents 10 times using identical prompt settings. The consistency score for each task was measured by comparing the similarity of outputs across these 10 runs, and an overall consistency score was then derived by averaging these measurements across all validation cases.

Results

The data collection and extraction process produced comprehensive results across the three primary analysis focus. The following findings cover the scope of available digital source, performance benchmarking against established directories, and model performance metrics.

Raw Multimodal Data Corpus

Data acquisition spanned six weeks, followed by two additional weeks to complete the extraction process. As of May 15, 2025, the sources included 424 verified operational farmers market locations representing 348 distinct market organizers. This total has included 49 market locations active in 2024 and 375 confirmed for the 2025 season. This study compiled 2,961 documents from public sources (Table 4) and about 3,000 API calls, capturing fragmented digital presences across multiple documents per market. Newly established markets frequently lack centralized digital footprints (e.g., official websites or farmers market directories) and often rely on decentralized sources like local news articles or community social media posts. Specifically, we obtained 210 rules and regulations documents from publicly accessible market websites, representing a total of 248 market organizations and 320 market locations. Additionally, nine markets were identified as having such documents that were not available online but could be acquired through direct requests to market organizers. This indicates the potential to expand the dataset further through outreach efforts and appropriate data-sharing agreements.

Table 4. Publicly Available Digital Sources for Michigan Farmers Markets

Source	Data Format	Number of documents
Farmers market website	md, pdf, image, docx	830*
Social media channels	md, image, pdf	1,560
Local news	md	127
Lifestyle sources	md	145
Government datasets	md, pdf, txt, docx, table	8
Local community websites	md, pdf, images, txt, docx	90
Events websites	md	201
Other public records	JSON by API Calls	2,998

* Including 210 rules, regulations, and/or application documents

Table 5. Comparison of MLLM Results with Michigan Farmers Market Directory

Topics	MFMD	MLLM
Address	100%	100%
Operation Schedule	100%	100%
Marketing Channels	100%	95%
Food Assistance Programs (FAP)	57%	55%
Manager Name	100%	N/A
MIMFA Trained Manager	25%	N/A
Total Number of Records	**348**	**310**

Topic Coverage and Completeness Evaluation

In this study, the topic coverage was defined as the percentage indicating the proportion of MFMD and MFMC topics within each category that the MLLM dataset successfully captured. Table 5 compares the data coverage of the MLLMdriven dataset, collected in 2024 and 2025, with the MFMD (MIFMA, n.d.). A key methodological divergence on seasonal markets was processed before a direct comparison. The MFMD aggregates seasonal locations under a single market entity per organizer, whereas our initial extraction recorded them as separate locations. The MLLM data was recoded to adopt the MFMD's convention, defining a "market" as an organizational entity rather than a physical location. Under this aligned framework, the MLLM-based dataset identified 348 markets from public digital sources, compared to the 310 markets listed in the Michigan Directory.

Of the total records, 298 farmers markets were found in both the MLLM dataset and the MFMD. The MLLM dataset excluded certain markets listed in the MFMD based on evidence from public sources that they were inactive (e.g., closed for construction) or permanently closed as of 2025. In contrast, the MFMD does not include several newly established markets (2024 and 2025) that were identified by the MLLM model. The MFMD appeared to include some farm stands that were outside the MLLM's farmers market inclusive criteria that at least two vendors are required in order to be considered a market. Both datasets provided full coverage (100%) for essential operational details like market address and schedule. Marketing channel inclusion was high (100% in the MFMD, 95% in MLLM), and coverage of food assistance programs was similar across sources (57% MFMD, 55% MLLM). In total, MLLM covered 4 out of the 6 variables (67%) in MFMD. Moreover, the MFMD uniquely captured administrative details less relevant to consumers: market manager names (100% coverage) and MIFMA-trained manager (25% coverage). This information was not collected in the MLLM dataset due to privacy considerations and limited public availability.

Table 6 compares MFMC and MLLM topic coverage. MFMC reported 176 market responses without topic-specific rates (MIFMA, 2024), while

Table 6. Topic Coverage Comparison of MFMC and MLLM

Topics	MFMC-Reported Items	#MFMC	#MLLM
Operation schedule	Operation Month, Weekdays	2	2
Location	Location Identification, Property Ownership, Indoor/Outdoor, Site Type (e.g., parking lots), Amenities, Customer Transportation Mode	6	5
Vendors	Vendor Count, Vendor Origin	2	2
Products	Products by categories, Cottages Foods by categories	2	2
Payments	Food Assistance Programs (FAP) counts, FAP transaction details, FAP infrastructures, and other accepted payments	4	3
Market organizer	Market Organization	1	1
Market manager	Responsibility, Pay, and Demographics	3	2
Volunteers	Has Volunteers, Total Volunteer Hours	2	1
Community issues	Community Activities and Engagement	1	1
Sponsor	Sponsors, How Sponsor Support Market	2	1
Labor force at markets	Employee Structure, Compensation, Demographics	3	0
Licensing and insurance	Licensing and Insurance	1	1
Budget and spending	Overall Budget, Budget Breakdowns, Total Revenue, Revenue Breakdowns	4	2
Training	Food Safety Training, MIFMA Certified Manager	2	1
Market evaluation	Visitor Counts, Vendor Matrix	2	0
Business Incubators	Market as Business Incubators	1	0
Longevity	Market Operation Years	1	1
Mission/Vision	Market Missions	1	1
Special topics	COVID, MIFMA leadership	2	0
Total		**42**	**26**

MLLM covered 348 markets with partial details. Among these, 210 markets included rules and regulations addressing detailed management topics, such as market organizers, managers, licensing and insurance, mission, vision, etc.

The table shows that the MLLM dataset captured 15 out of 19 topics (79%) and 26 out of 42 items (62%) from the MFMC. It reached 100% coverage on factual and operational topics, including operation schedules, locations, site characteristics, market size, vendors, products, season length, and mission statements. These topics were likely well-represented in publicly available digital sources. Partial coverage items included market governance (80%), market manager responsibilities (50%), volunteers (30%), community support (50%), vendor management (50%), food assistance programs (30%), budget and spending (5%), employment and compensation (10%), and market evaluation practices (5%), where data were often non-public. Demographics and time-sensitive special topics (e.g., COVID-19 issues) had no coverage, as MLLMs excluded personally identifiable information, even when such data might have been available in images on websites or social media, due to privacy constraints. By combining MFMD and MFMC, MLLM achieved 76% coverage (19/25 topics).

Model Performance Evaluation

Table 7 presents the MLLM's task-specific performance across five extraction tasks from farmers market datasets, evaluated using flexible semantic matching criteria.

The model demonstrated near-perfect performance on structured, text-based tasks, specifically on operational schedules (*F1* = 0.998), FAP identification (*F1* = 0.981), and vendor fee structures (*F1* = 0.974). Standardized, explicitly stated formats of these items led to high scores by minimizing ambiguity and enabling the model to achieve high precision with zero hallucination. In contrast, product retrieval from images (*F1* = 0.962) showed slightly diminished precision (0.944) and a small hallucination rate (1.9%), indicating the inherent challenges of interpreting visual content. The model performed reliably for prominent, centered items in an image but struggled with obscure, low-resolution, or peripheral elements, sometimes mislabeling or hallucinating objects. Address retrieval, while reaching a perfect recall, recorded the lowest precision (0.878). The decrease was largely due to inconsistent formatting and implicit contextual dependencies. Informal location descriptions (e.g., "Tractor Supply's parking lot") often omitted essential components such as street numbers or ZIP codes, while local shorthand in documents (e.g., "next to the County Courthouse") introduced ambiguity. As a result, the model's attempts to parse these informal descriptions sometimes led to false-positive (FP) records, which reduced the overall precision despite the impeccable identification of relevant descriptive phrases (perfect recall). In sum, these results validated the MLLM's capability to accurately extract structured information from both textual and visual farmers market sources. These consistently high-performing metrics across all tasks confirmed that this approach is a feasible solution for alleviating manual data entry burdens and achieving more comprehensive data collection. Meanwhile, challenges in clarifying complex addresses and subtle visual details also posed

Table 7. LLM Performance Metrics for Five Extraction Tasks

Tasks	Precision	Extraction Recall	Extraction F1-Score	Output Consistency	Hallucination Rate
Address Retrieval	0.878	1.000	0.935	0.995	0.000
Operational Schedule Retrieval	0.996	1.000	0.998	0.989	0.000
Food Assistance Programs (FAP) Retrieval	1.000	0.963	0.981	1.000	0.000
Vendor's Fee Structure Retrieval	1.000	0.950	0.974	0.963	0.000
Products Retrieval from images	0.944	0.981	0.962	0.929	0.019

an opportunity for the model's continued refinement.

Discussion, Limitations and Future Work

The application of MLLMS to farmers market documents analysis reveals both promising results and important consideration for effective utilizations. The following section addresses model's performance characteristics, the complementary relationship between automated extraction and traditional surveys, and the implication for broader research adoptions of the framework.

Performance of MLLM Method

This study introduced the first systematic application of MLLMs to automate farmers market data extraction from open online sources. The model achieved near-perfect retrieval performance on critical variables (e.g., operating schedules), confirming its reliability for structured text and image data extraction. This study used Google's Gemini-2.0-Flash as the MLLM due to its robust performance in analyzing both long-form text and visual data while maintaining low hallucination rates (Hughes et al., 2024). A key limitation of Gemini-2.0-Flash, like many proprietary models, is its lack of transparency in architecture and training data. To mitigate this, we adopted an iterative prompt engineering strategy to tailor the prompts to fit the farmers market data collection. Looking ahead, we plan to enhance the model's capabilities by (1) incorporating farmers market regulatory documents and expert-reviewed annotations to improve interpretation accuracy, and (2) combining Gemini-2.0-Flash with fine-tuned open-source multimodal frameworks, such as Llama (Meta, 2023), to build hybrid systems that prioritize transparency, adaptability, and enhanced multimodal capabilities.

Leveraging MLLMs to Enhance, Not Replace, Traditional Surveys

This study demonstrated the efficient extraction of data from public digital sources with unstructured traces. Based on comparison with the MFMD and MFMC, the MLLM framework successfully covered 76% of directory and survey contents, suggesting that it can take over a large share of routine data collection in farmers market surveys, such as MFMC. This study deliberately focused only on freely available sources in order to explore the utility boundaries of such data and illustrate what could be achieved using automated tools without relying on market manager surveys. This choice also echoed the existing research that has shown administrative records and digital sources often report operational details more consistently than self-reported survey data (Kreuter, 2013; Low et al., 2015) as well as prior calls to avoid redundant data collection while still maintaining rigor in community-based research (Meyer et al., 2015).

The comparison of the MLLMs framework with MFMD and MFMC confirmed that this method was not capable of collecting everything: MLLMs cannot access data that is private, subjective, or requires direct input from humans, such as manager opinions and employee compensation. Therefore, MLLMs cannot fully replace traditional surveys. However, this study showed that an MLLM approach can significantly augment and support the survey method. In addition to replicating the MFMD and MFMC survey content, this method was able to collect more detailed and expansive information that traditional surveys often do not capture. For example, in the "Location" category, it extracted market sites' walkability scores, urbanicity, demographic profiles, and zoning codes, offering a deeper view into the physical and social context of each market. Furthermore, MLLMs surfaced new themes not covered by existing surveys, including market governance practices and operational rules embedded in the markets' rules and regulations documents. These capabilities demonstrated how MLLMs can complement traditional survey methods by filling gaps in data collection and expanding the overall understanding of farmers market environments.

Future research should establish direct partnerships with farmers markets to streamline data collection. Rather than burdening managers with lengthy surveys for already publicly available facts, this approach would (1) request existing documents for objective data and (2) limit surveys to solely subjective or private questions. Under proper data-sharing agreements, such collaborations could provide privileged access to granular records, such as financials, transaction logs, customer traffic analytics, and FAP data, and significantly enhance coverage beyond public sources alone.

Toward Scalable, Sustainable, and Structured Farmers Market Data Collection Across Time and Geography

This framework attained significant scalability by automatically harvesting and organizing decentralized factual data (e.g., operating schedules), offering three main advantages: (1) significant cost efficiency, with data extraction costing less than US$1 per market (API processing of less than 20K tokens at US$1.24/million tokens per market) compared to MFMC's minimum disclosed expenditure of US$50 (MIFMA, 2024); (2) immediate nationwide implementation, using standardized data extraction protocol and requiring no redevelopment across states or regions for consistent public data extraction; and (3) sustainable longitudinal tracking, where annual updates only require minimal marginal costs after the initial setup. This post-deployment cost structure, which solely includes document corpus and MLLM processing, makes longitudinal market surveillance economically viable, where traditional surveys remain highly resource-intensive.

Taxonomy-Driven Framework: A Paradigm for Transferable Research

In contrast to traditional surveys that flatten complex realities into tables, this study created a taxonomy-driven, two-step framework that demonstrated the potential of aligning unstructured data with structured schemas, enabling intuitive, human-like representations of information and relationships (such as fee structures) that mirrored the real-world relationships with maximum granularity. The method also drew attention to the practical utility of adopting AI: using this two-step framework, researchers (domain experts) can prioritize specialized knowledge over technical complexities. By minimizing the need for technology-intensive tasks, such as custom AI architecture, this framework allowed experts to iteratively refine outputs through prompt engineering while focusing on high-impact work: designing domain-specific schemas, optimizing data relationship representations, and leveraging extracted data for their research goals by eliminating barriers requiring deep technical skills in low-level engineering. In addition, this method's modular design could support cross-disciplinary transferability for researchers in other domains. By blending practical knowledge with scalable AI tools, this method could help researchers move beyond traditional survey constraints to generate equitable, adaptable findings.

Conclusion

This study pioneered the application of MLLMs to automate farmers market data extraction from public digital sources and provided a scalable and cost-effective data harvesting solution. By adopting Google Gemini-2.0-Flash with a two-step framework, the method systematically extracted Michigan farmers market data from various public sources. Compared to the MFMC and MFMD, this technique achieved an average accuracy of 98% on key variables and 76% on topic coverage. While MFMC's survey methodology cost a minimum US$50 per market in disclosed incentive costs, our AI-driven method extracted all data for less than US$1 per market after the initial setup, supported real-time updates, and had near-zero hallucination rates (0% for schedule extraction and 1% for image-based product identification). Though limited to publicly available data, the methodology could augment traditional surveys by automating the extraction of factual data from public records. The findings of this study provide a practical framework for local food system research, combining AI-processed public records and surveys to broaden access to timely, granular information for resilient and equitable food systems.

Acknowledgment

We acknowledge Qian Yun Zhang, incoming undergraduate researcher at the University of Michigan, for her contributions to preliminary data collection and source identification.

References

Aziz, A., José Herrera-Murillo, D. J., Nogueras-Iso, J., Lacasta, J., & Lopez-Pellicer, F. J. (2024). Identifying the evolution of Open Government Data initiatives and their user engagement. *IEEE Access,* 12, 84556–84566. https://doi.org/10.1109/ACCESS.2024.3414282

Bommasani, R., Hudson, D. A., Adeli, E., Altman, R., Arora, S., von Arx, S., Bernstein, M. S., Bohg, J., Bosselut, A., Brunskill, E., Brynjolfsson, E., Buch, S., Card, D., Castellon, R., Chatterji, N., Chen, A., Creel, K., Davis, J. Q., Demszky, D., … Liang, P. (2022). *On the opportunities and risks of foundation models* (No. arXiv:2108.07258). arXiv. https://doi.org/10.48550/arXiv.2108.07258

Brown, C. (2003). Consumers' preferences for locally produced food: A study in southeast Missouri. *American Journal of Alternative Agriculture*, *18*(4), 213–224. https://doi.org/10.1079/AJAA200353

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A. A., Ziegler, D. M., Wu, J., Winter, C., … Amodei, D. (2020). Language models are few-shot learners. In H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan, & H. Lin (Eds.), *Proceedings of the 34th International Conference on Neural Information Processing Systems* (pp. 1877–1901). Curran Associates. https://dl.acm.org/doi/10.5555/3495724.3495883

Dagdelen, J., Dunn, A., Lee, S., Walker, N., Rosen, A. S., Ceder, G., Persson, K. A., & Jain, A. (2024). Structured information extraction from scientific text with large language models. *Nature Communications, 15,* Article 1418. https://doi.org/10.1038/s41467-024-45563-x

Elchafei, P., & Abu-Elkheir, M. (2025). *Span-level hallucination detection for LLM-generated answers* (No. arXiv: 2504.18639). arXiv. https://doi.org/10.48550/arXiv.2504.18639

Google DeepMind. (2024, December 11). *Introducing Gemini 2.0: Our new AI model for the agentic era* [Technical report]. https://blog.google/technology/google-deepmind/google-gemini-ai-update-december-2024/#gemini-2-0-flash

Greyling, C. (2024, February 23). *LLM drift, prompt drift & cascading*. Medium. https://cobusgreyling.medium.com/llm-drift-prompt-drift-cascading-5a2ea2a5c455

Gu, B., Shao, V., Liao, Z., Carducci, V., Brufau, S. R., Yang, J., & Desai, R. J. (2025). Scalable information extraction from free text electronic health records using large language models. *BMC Medical Research Methodology, 25*, Article 23. https://doi.org/10.1186/s12874-025-02470-z

Guthrie, J., Guthrie, A., Lawson, R., & Cameron, A. (2006). Farmers' markets: The small business counter-revolution in food production and retailing. *British Food Journal, 108*(7), 560–573. https://doi.org/10.1108/00070700610676370

Hacker, K. L., & Mason, S. M. (2003). Ethical gaps in studies of the digital divide. *Ethics and Information Technology, 5*, 99–115. https://doi.org/10.1023/A:1024968602974

Hinrichs, C. C. (2000). Embeddedness and local food systems: Notes on two types of direct agricultural market. *Journal of Rural Studies, 16*(3), 295–303. https://doi.org/10.1016/S0743-0167(99)00063-7

Holloway, L., Kneafsey, M., Venn, L., Cox, R., Dowler, E., & Tuomainen, H. (2007). Possible food economies: A methodological framework for exploring food production–consumption relationships. *Sociologia Ruralis, 47*(1), 1–19. https://doi.org/10.1111/j.1467-9523.2007.00427.x

Huang, Z., Peng, P., Lu, F., & Zhang, H. (2025). An LLM-based method for quality information extraction from web text for crowed [*sic*]-sensing spatiotemporal data. *Transactions in GIS*, *29*(1), Article e13294. https://doi.org/10.1111/tgis.13294

Hughes, S., Bae, M., & Li, M. (2024). *Vectara Hallucination Leaderboard* [Python]. https://github.com/vectara/hallucination-leaderboard (Original work published 2023)

Jain, S. (2025, February 20). *Grok 3—The reasoning agent*. Medium. https://medium.com/@sulbha.jindal/grok-3-the-reasoning-agent-0eb8c97b55a0

Joshi, S. (2025). *Evaluation of large language models: Review of metrics, applications, and methodologies*. Preprints. https://doi.org/10.20944/preprints202504.0369.v2

Kreuter, F. (Ed.). (2013). *Improving surveys with paradata: Analytic uses of process information*. John Wiley. https://doi.org/10.1002/9781118596869

Kumar, S. (2024, February 28). *Detecting LLM hallucinations: Strategies and overview*. Medium. https://medium.com/@techsachin/detecting-llm-hallucinations-strategies-and-overview-57eea69e6a07

Lai, H., Liu, J., Bai, C., Liu, H., Pan, B., Luo, X., Hou, L., Zhao, W., Xia, D., Tian, J., Chen, Y., Zhang, L., Estill, J., Liu, J., Liao, X., Shi, N., Sun, X., Shang, H., Bian, Z., … On behalf of ADVANCED Working Group. (2025). Language models for data extraction and risk of bias assessment in complementary medicine. *npj Digital Medicine*, *8*, Article 74. https://doi.org/10.1038/s41746-025-01457-w

Lao, M., Pu, N., Liu, Y., Zhong, Z., Bakker, E. M., Sebe, N., & Lew, M. S. (2023). Multi-domain lifelong visual question answering via self-critical distillation. In A. El Saddik, T. Mei, R. Cucchiara, M. Bertini, D. P. Tobon Vallejo, P. K. Atrey, & M. S. Hossain (Chairs), *MM '23: Proceedings of the 31st ACM International Conference on Multimedia* (pp. 4747–4758). Association for Computing Machinery. https://doi.org/doi:10.1145/3581783.3612121

Lee, D., & Yu, H. (2025). *REFIND at SemEval-2025 Task 3: Retrieval-augmented factuality hallucination detection in large language models* (No. arXiv:2502.13622). arXiv. https://doi.org/10.48550/arXiv.2502.13622

Liu, P., Yuan, W., Fu, J., Jiang, Z., Hayashi, H., & Neubig, G. (2023). Pre-train, prompt, and predict: A systematic survey of prompting methods in natural language processing. *ACM Computing Surveys*, *55*(9), Article 195. https://doi.org/10.1145/3560815

Liu, R., Jafar, R., Girard, L. A., Thorlund, K., Rizzo, M., & Forsythe, A. (2023). P24 Improving efficiency of living systematic literature reviews (SLR) with artificial intelligence (AI): Assisted extraction of population, intervention/comparator, outcome, and study design (P-I/C-O-S). *Value in Health, 26*(12, Suppl.), S6. https://doi.org/10.1016/j.jval.2023.09.033

Liu, S., Halder, K., Qi, Z., Xiao, W., Pappas, N., Htut, P. M., John, N. A., Benajiba, Y., & Roth, D. (2025). *Towards long context hallucination detection* (No. arXiv:2504.19457). arXiv. https://doi.org/10.48550/arXiv.2504.19457

Low, S. A., Adalja, A., Beaulieu, E., Key, N., Martinez, S., Melton, A., Perez, A., Ralston, K., Stewart, H., Suttles, S. & Jablonski, B. B. R. (2015). *Trends in U.S. local and regional food systems: A report to Congress* (AP-068). U.S. Department of Agriculture, Economic Research Service. https://www.ers.usda.gov/publications/pub-details?pubid=42807

Meta. (2023, February 24). *Introducing LLaMA: A foundational, 65-billion-parameter large language model* [Technical report]. https://ai.meta.com/blog/large-language-model-llama-meta-ai/

Meyer, B. D., Mok, W. K. C., & Sullivan, J. X. (2015). Household surveys in crisis. *Journal of Economic Perspectives, 29*(4), 199–226. https://doi.org/10.1257/jep.29.4.199

Michigan Farmers Market Association [MIFMA]. (n.d.). *Find a farmers market*. Retrieved May 15, 2025, from https://mifma.org/find-a-farmers-market/

MIFMA. (2024). *Census of Michigan farmers markets*. https://mifma.org/for-markets/census/

Michigan House of Representatives. (2023, January 5). *Double Up Food Bucks participating sites*. https://www.house.mi.gov/hfa/PDF/Agriculture/MDARD_DNR_Subcmte_Testimony_DU-MI-Participating-Locations-List_3-7-23.pdf

Ouyang, L., Qu, Y., Zhou, H., Zhu, J., Zhang, R., Lin, Q., Wang, B., Zhao, Z., Jiang, M., Zhao, X., Shi, J., Wu, F., Chu, P., Liu, M., Li, Z., Xu, C., Zhang, B., Shi, B., Tu, Z., & He, C. (2025). *OmniDocBench: Benchmarking diverse PDF document parsing with comprehensive annotations* (No. arXiv:2412.07626). arXiv. https://doi.org/10.48550/arXiv.2412.07626

Powell, A. B., Ustek-Spilda, F., Lehuedé, S., & Shklovski, I. (2022). Addressing ethical gaps in 'Technology for Good': Foregrounding care and capabilities. *Big Data & Society, 9*(2), 1–12. https://doi.org/10.1177/20539517221113774

Rasool, Z., Kurniawan, S., Balugo, S., Barnett, S., Vasa, R., Chesser, C., Hampstead, B. M., Belleville, S., Mouzakis, K., & Bahar-Fuchs, A. (2024). *Evaluating LLMs on document-based QA: Exact answer selection and numerical extraction using Cogtale dataset* (No. arXiv:2311.07878). arXiv https://doi.org/10.48550/arXiv.2311.07878

Sansford, H., Richardson, N., Maretic, H. P., & Saada, J. N. (2024). *Grapheval: A knowledge-graph based LLM hallucination evaluation framework* (No. arXiv:2407.10793). arXiv https://doi.org/10.48550/arXiv.2407.10793

Schilling-Wilhelmi, M., Ríos-García, M., Shabih, S., Gil, M. V., Miret, S., Koch, C. T., Márquez, J. A., & Jablonka, K. M. (2025). From text to insight: Large language models for chemical data extraction. *Chemical Society Reviews, 54*(3), 1125–1150. https://doi.org/10.1039/D4CS00913D

Schwitter, N. (2025). Using large language models for preprocessing and information extraction from unstructured text: A proof-of-concept application in the social sciences. *Methodological Innovations, 18*(1), 61–65. https://doi.org/10.1177/20597991251313876

Shah, D. (2024, November 1). Evaluating prompt effectiveness: Key metrics and tools. *Portkey* [Blog]. https://portkey.ai/blog/evaluating-prompt-effectiveness-key-metrics-and-tools/

Siepmann, R. M., Baldini, G., Schmidt, C. S., Truhn, D., Müller-Franzes, G. A., Dada, A., Kleesiek, J., Nensa, F., & Hosch, R. (2025). An automated information extraction model for unstructured discharge letters using large language models and GPT-4. *Healthcare Analytics, 7*, Article 100378. https://doi.org/10.1016/j.health.2024.100378

Sushil, M., Zack, T., Mandair, D., Zheng, Z., Wali, A., Yu, Y.-N., Quan, Y., Lituiev, D., & Butte, A. J. (2024). A comparative study of large language model-based zero-shot inference and task-specific supervised classification of breast cancer pathology reports. *Journal of the American Medical Informatics Association*, *31*(10), 2315–2327. https://doi.org/10.1093/jamia/ocae146

U.S. Department of Agriculture, Agricultural Marketing Service [USDA AMS]. (2020). *2019 National Farmers Market Managers Survey*. https://www.ams.usda.gov/services/local-regional/research-publications/fmms

Wu, H., Yuan, Y., Mikaelyan, L., Meulemans, A., Liu, X., Hensman, J., & Mitra, B. (2024). *Learning to extract structured entities using language models* (No. arXiv: 2402.04437). arXiv. https://doi.org/10.48550/arXiv.2402.04437

Zhang, Y., Fan, H., & Yang, Y. (2024). *Prompt-aware adapter: Towards learning adaptive visual tokens for multimodal large language models* (No. arXiv: 2405.15684). arXiv. https://doi.org/10.48550/arXiv.2405.15684

Zhou, S., Alon, U., Xu, F. F., Wang, Z., Jiang, Z., & Neubig, G. (2022). *DocPrompting: Generating code by retrieving the docs* (No. arXiv: 2207.05987). arXiv. https://doi.org/10.48550/arXiv.2207.05987

Zhou, Y., Muresanu, A. I., Han, Z., Paster, K., Pitis, S., Chan, H., & Ba, J. (2023). Large language models are human-level prompt engineers [Conference paper]. *The Eleventh International Conference on Learning Representations* (No. arXiv: 2211.01910). arXiv. https://doi.org/10.48550/arXiv.2211.01910

Zimmer, M. (2018). Addressing conceptual gaps in big data research ethics: An application of contextual integrity. *Social Media + Society*, *4*(2), Article 2056305118768300. https://doi.org/10.1177/2056305118768300