Mapping Fine-Tuning and Dataset Curation Practices for Large Language Models in Agriculture and Water Resources: A Scoping Review

CLAUDIA LENGUA-CANTERO, MARÍA GARCÍA MEDINA, CARLOS COHEN MANRIQUE, LAUDYT LAMBRAÑO PÉREZ, DAVID ACOSTA MEZA

Abstract: *The adaptation of Large Language Models (LLMs) to agriculture and water resource management is an emerging field whose fine-tuning strategies and dataset curation practices remain methodologically heterogeneous, with direct implications for Sustainable Development Goals 2, 6, and 13. This scoping review maps the extent, range, and nature of the available literature, identifies dominant models and datasets, and characterises research gaps. Sources of evidence were included if they reported on Large Language Models, foundation models, or transformer-based architectures with an explicit domain-adaptation methodology in agriculture, water resources, hydrology, or irrigation, and described at least one fine-tuning or dataset curation action. A structured search across Scopus, IEEE Xplore, ACM Digital Library, SciELO, and ScienceDirect, with supplementary searches in Nature, MDPI, OpenReview, ACL Anthology, and Frontiers, covered publications from January 2020 to October 2025. Following the methodological framework of Arksey and O'Malley and the PRISMA-ScR reporting guidance, two reviewers independently charted bibliographic metadata, primary domain, AI architecture, adaptation strategy, data strategy, and application focus from the full extracted record of each included source. From 244 compiled records, 91 duplicates and 2 non-verifiable records were removed, yielding 151 verified sources with DOI or stable URL. The LLM/RAG era proper*

Claudia Lengua-Cantero, Faculty of Basic Sciences, Engineering and Architecture, Caribbean University Corporation (CECAR), Sincelejo, Colombia
María García Medina, Faculty of Basic Sciences, Engineering and Architecture, Caribbean University Corporation (CECAR), Sincelejo, Colombia
Carlos Cohen Manrique, Faculty of Basic Sciences, Engineering and Architecture, Caribbean University Corporation (CECAR), Sincelejo, Colombia
Laudyt Lambraño Pérez, Faculty of Basic Sciences, Engineering and Architecture, Caribbean University Corporation (CECAR), Sincelejo, Colombia
David Acosta Meza, Mateo Pérez Educational Institution, Sampués, Colombia

begins in mid- to late 2023, with 76.3% of dated sources published in 2024–2025. Open-source Llama-3 and Mistral families are the most frequently named fine-tuned backbones, while GPT-4 dominates evaluation and hybrid RAG architectures, and dedicated benchmarks and curated datasets signal a methodological shift from model-centric to data-centric approaches. The mapped literature points to dataset quality and domain-specific curation as central engineering concerns for the effective and trustworthy deployment of LLMs in agro-hydrological contexts.

Keywords: *Agricultural AI; Dataset Curation; Domain Adaptation; Fine-Tuning; Large Language Models; Scoping Review; Sustainable Development; Water Resource Management.*

Introduction

The adaptation of Large Language Models (LLMs) to domain-specific applications, particularly agriculture and water resource management, represents one of the most rapidly evolving frontiers in Artificial Intelligence (AI) and Natural Language Processing (NLP). Recent reviews confirm both the breadth and the velocity of this expansion [1,2]. Fine-tuning, defined as the process of adjusting pre-trained models to perform more effectively in targeted domains, has emerged as a critical strategy for overcoming the limitations of general-purpose models in specialised, knowledge-intensive contexts [1]. The effective deployment of these approaches depends critically on the quality and structure of the underlying training datasets, a dimension that has received comparatively less systematic attention in the agro-hydrological literature.

The alignment of AI technologies with sustainability imperatives has acquired strategic importance in recent years. LLMs applied to agro-hydrological contexts can directly advance the United Nations Sustainable Development Goals (SDGs), particularly SDG 2 (Zero Hunger), SDG 6 (Clean Water and Sanitation), and SDG 13 (Climate Action), by enabling more precise, evidence-based, and locally informed decision support for farmers, water managers, and policy-makers [3,4]. Realising this potential, however, requires overcoming persistent technical barriers associated with domain specificity, data quality, and computational resources.

LLMs have established themselves as leading architectures in AI due to their capacity to learn complex language representations from large corpora and to generalise knowledge across tasks [5]. These models offer substantial potential for

decision support, technical documentation analysis, and cognitive automation in specialised fields, with a growing range of dedicated agro-water systems demonstrating the value of domain-specific adaptation [1,6]. Nevertheless, when general-purpose LLMs are deployed directly in technically demanding environments such as agriculture or water resource management, they exhibit significant capability constraints attributable to domain-specific terminology, localised datasets, and contextualised knowledge requirements [7].

These limitations manifest in three principal forms. First, there is systematic underrepresentation of specialised knowledge in pre-training corpora: LLMs trained on generalist text collections exhibit gaps in technical terminologies, sectoral regulations, and approaches specific to engineering, agronomy, and water management, producing linguistically fluent but conceptually imprecise responses [5,6]. Second, pre-trained models demonstrate limited capacity for processing contextualised and localised information, particularly with respect to geographical, climatic, regulatory, or cultural variables [1]. Third, LLMs are susceptible to hallucination—the generation of false or unreliable information presented with high linguistic confidence—a problem exacerbated in specialised domains where internal verification mechanisms are weak and reliable structured knowledge is sparse [8,9]. Ensuring causability and explainability of AI outputs is therefore essential for their acceptance in scientific and operational contexts [10,11].

Dataset curation encompasses systematic processes of selection, purification, annotation, validation, and balancing, carried out to ensure relevance, consistency, and representativeness. In agro-hydrological scenarios, this task is uniquely challenging given the multidimensionality of the data—integrating climatic, edaphological, productive, and socioeconomic variables—and the necessity of incorporating localised territorial knowledge and practices [12]. Initiatives in agricultural named-entity recognition, multilingual advisory corpora, and national-scale agricultural training datasets illustrate the emerging emphasis on dedicated agricultural corpora. Poor curation can introduce biases, reduce model interpretability, and restrict practical applicability [13,14].

Parameter-efficient fine-tuning (PEFT) strategies, particularly Low-Rank Adaptation (LoRA) [15] and its quantised variant QLoRA, have emerged as effective mechanisms for adapting general-purpose LLMs to agro-water domains at manageable computational cost. Recent evaluations of LoRA in environmental, multilingual agricultural [16], and clinical-prognostic [18] contexts confirm

performance comparable to full fine-tuning at reduced computational cost. Retrieval-Augmented Generation (RAG) has further shown promise for mitigating hallucinations and enhancing factual consistency by integrating structured external knowledge bases into the generation process [17].

Despite these advances, the literature remains methodologically fragmented, tending to highlight specific applications or model evaluations rather than systematically mapping how fine-tuning strategies and dataset curation decisions interact to shape model performance in agriculture and water resource management [2,6]. Existing reviews have addressed adjacent topics—generic PEFT methods, foundation models in agriculture [6], or hydrology-specific benchmarks—but no prior scoping review has explicitly mapped the intersection of fine-tuning and dataset curation practices within the agro-water domain using the methodological framework of Arksey and O'Malley [30] as refined by Levac et al. [31] and reported in accordance with the PRISMA Extension for Scoping Reviews (PRISMA-ScR) [28].

1.1. Conceptual Definitions

To resolve terminological ambiguity in the literature, this scoping review adopts the following operational definitions. A Large Language Model (LLM) is defined as a transformer-based model trained on extensive text corpora and capable of generalising knowledge across natural-language tasks; representative families include the GPT, Llama, Mistral and DeepSeek series. A foundation model, more broadly, refers to a large-scale model trained on broad data and adaptable to a wide range of downstream tasks. Agricultural artificial intelligence is understood as the application of AI methodologies to the agricultural domain, including but not limited to LLM-based systems. The agro-water domain is defined as the joint area encompassing agriculture, water resources, hydrology, irrigation, and related practices that link agricultural production to water management. The present scoping review addresses LLM and foundation-model adaptation and dataset curation for LLMs in the agro-water domain; it does not address classical machine-learning approaches that lack a Large Language Model or transformer-based domain-adaptation component.

1.2. Research Question and Objectives

This scoping review aims to map the extent, range, and nature of the current scientific literature on fine-tuning and data curation practices for Large Language Models applied to agro-water applications, following the established purposes of scoping reviews [30,31]. The objectives are: (i) to identify the extent and nature of

the available LLM-focused literature; (ii) to characterise the dominant models, datasets, fine-tuning techniques, and reported obstacles; (iii) to clarify key methodological concepts in the field; and (iv) to identify research gaps to guide future systematic reviews and primary investigation. The central research question is: What is the extent, range, and nature of the literature reporting fine-tuning strategies and dataset curation practices for Large Language Models applied to agriculture and water resource management?

Materials and Methods

2.1. Protocol and Reporting Framework

This scoping review was conducted following the methodological framework proposed by Arksey and O'Malley [30] and refined by Levac et al. [31], and is reported in accordance with the PRISMA Extension for Scoping Reviews (PRISMA-ScR) [28]. A scoping review was selected as the appropriate review type because the research question requires mapping the extent, range, and nature of an emerging and methodologically heterogeneous body of evidence, rather than synthesising effect estimates or grading evidence strength [29]. The PRISMA-ScR framework explicitly supports the present approach to data charting at the source level using the full extracted record (title, abstract, stated objectives, methods, dataset description, and reported findings), with transparent coding of dimensions [28]. Consistent with PRISMA-ScR guidance, a six-domain methodological characterisation of each included source was undertaken as an optional, descriptive step (Section 2.6) to characterise the maturity of the field rather than to grade evidence strength. The completed PRISMA-ScR checklist is provided as Supplementary Table S1. The protocol was defined a priori and is reported in full in Sections 2.2–2.7; it was not prospectively registered, as PROSPERO does not accept scoping review protocols.

2.2. Eligibility Criteria

Sources of evidence were included if they met all of the following six criteria, all of which require an explicit Large Language Model or transformer-based domain-adaptation component:

- Domain: Agriculture, water resources, hydrology, irrigation, soil moisture, or agricultural water management.
- AI type: Large Language Model, foundation model, transformer-based architecture with explicit domain-adaptation methodology, or multimodal vision-language model incorporating an LLM component.

- Technical intervention: Explicit report on fine-tuning (LoRA, QLoRA, instruction tuning, supervised fine-tuning, adapters) or domain-adaptation protocol; alternatively, dataset construction or benchmarking explicitly designed for LLM evaluation in the domain.
- Data strategy: Description of at least one data preparation or curation action (cleaning, filtering, annotation, synthesis, balancing, noise control, or benchmarking).
- Performance reporting: Quantitative results following tuning or benchmark evaluation, or explicit dataset characterisation with reproducibility documentation.
- Document type: Peer-reviewed journal article, conference paper, preprint with explicit methodology, or dataset/benchmark with clear evaluation protocol.

Sources of evidence were excluded if they: (i) addressed domains outside agriculture, hydrology, or water resource management without direct application to these areas; (ii) employed only classical machine-learning architectures (e.g., standalone CNN, LSTM, isolation forests, or remote-sensing-only studies) without an LLM, foundation model, or transformer-based domain-adaptation component; (iii) did not report data preparation or fine-tuning/adaptation techniques; or (iv) were purely opinion pieces or editorials without methodological content. This strict LLM-focused exclusion criterion explicitly addresses the conceptual-scope ambiguity identified in earlier reviews of the agro-water AI literature.

2.3. Information Sources and Search Strategy

The structured search was initiated in early 2025 and subsequently updated, with the most recent search executed in October 2025, across five primary electronic databases—Scopus, IEEE Xplore, ACM Digital Library, ScienceDirect, and SciELO—covering publications from 1 January 2020 to 31 October 2025. A small number of methodological references that postdate this window—notably the reporting-guideline statements—are cited solely for methodological reasons and do not form part of the searched corpus. Supplementary searches were performed in Nature, MDPI, OpenReview, ACL Anthology, and Frontiers to ensure coverage of preprint and open-access venues where LLM research is frequently disseminated first. The search strings combined three semantic blocks using Boolean operators (AND between blocks; OR within blocks):

- Domain: agriculture OR farming OR crop* OR irrigation OR “water resources” OR hydrology.

- Data approach: “data quality” OR curation OR preprocessing OR labeling OR annotation OR benchmark OR dataset.
- Model adaptation: “fine-tuning” OR “parameter-efficient” OR LoRA OR “prompt tuning” OR RAG OR “domain adaptation” OR “large language model*” OR LLM*.

Searches were conducted in title, abstract and keyword fields for Scopus (TITLE-ABS-KEY), in All Metadata for IEEE Xplore Command Search, in title-abstract-keywords for ACM Digital Library and ScienceDirect, and in subject area / title / abstract for SciELO. Filters applied were: year (2020–2025); language (English, with Spanish-language and Portuguese-language records retained from SciELO); and document type (research articles, conference proceedings, and dedicated preprint venues). The complete search strings as executed in each database are provided as Supplementary Table S2. Reference lists of included sources and relevant prior reviews were hand-searched to identify additional eligible records.

2.4. Selection of Sources of Evidence

Records compiled from all databases were exported in RIS format and imported into Rayyan QCRI for deduplication and Mendeley for reference management. The records compiled and imported into Rayyan from the database and supplementary searches totalled 244 (the full export is provided as Supplementary Table S4). Automated and manual de-duplication removed 91 records, leaving 153 unique records. During eligibility and verifiability assessment, every record was checked for a stable identifier (DOI or persistent URL) and for genuine, in-scope existence; two further records were removed at this stage (one duplicate of an already-included study and one entry for which no real indexed publication could be confirmed). All 151 retained sources carry a DOI or stable URL. The final corpus comprises 151 verified sources, each with a stable identifier (Supplementary Table S4). Disagreements at each stage were resolved by consensus and, when consensus could not be reached, by consultation with a third reviewer. Critically, the reviewers retained only sources in which the central methodological contribution involved an LLM, a foundation model, a transformer-based domain-adaptation architecture, or a dataset/benchmark explicitly designed for LLM evaluation; sources reporting only CNN, LSTM, or other classical ML/DL methods without an LLM component were excluded. The complete selection process is summarised in Figure 1.

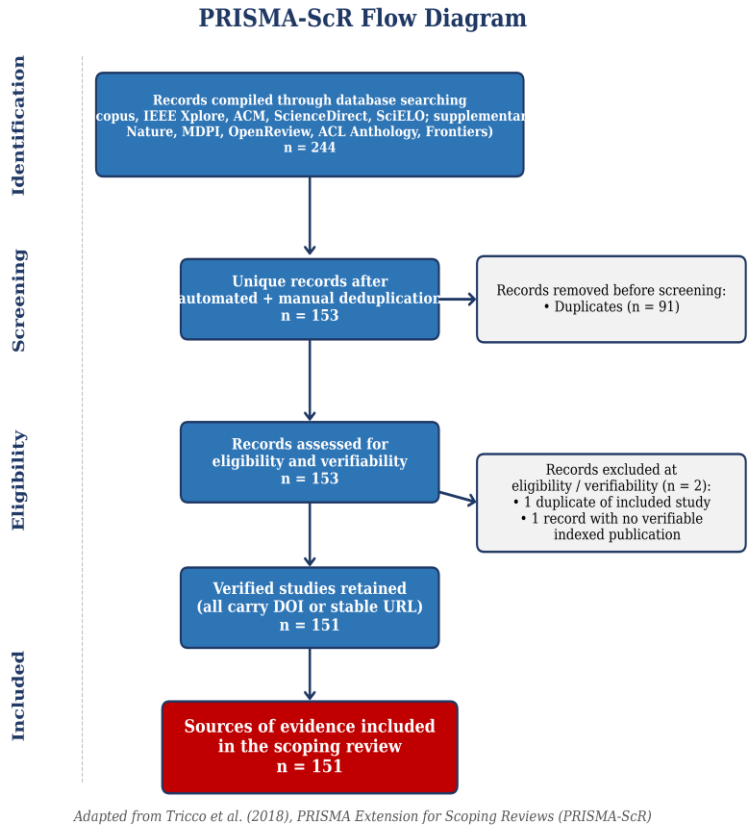

Figure 1. PRISMA-ScR flow diagram of the source selection, verification and screening process. n = number of records.

2.5. Data Charting Process and Data Items

A data-charting form was developed iteratively in line with PRISMA-ScR guidance [28] and the methodological refinements of Levac et al. [31]. The form was piloted on a randomly selected subset of ten included sources and refined before being applied to the full corpus. Two reviewers independently charted data from each included source, and discrepancies were resolved by consensus. Charted variables were organised around six analytical dimensions: (i) bibliographic metadata (title, authors, year, source); (ii) primary domain (agriculture, water/hydrology, climate/environment, or cross-cutting); (iii) AI architecture (LLM/foundation model, transformer, vision/multimodal); (iv) adaptation strategy (fine-tuning,

LoRA/QLoRA, PEFT, instruction tuning, RAG, prompt engineering, zero/few-shot); (v) data strategy (curation, annotation, augmentation, benchmark, dataset construction, NER); and (vi) application focus (advisory systems, disease/pest detection, irrigation management, forecasting, monitoring, information extraction). Each dimension was coded from the full extracted record of the source (title, abstract, stated objectives, methods, dataset description, and reported findings), not from titles alone; a category was assigned only where the extracted content provided explicit support, and any residual uncertainty was resolved by consensus between the two reviewers. The complete charting matrix is provided as Supplementary Table S3, and the full list of included sources with bibliographic metadata is provided as Supplementary Table S4. To enable independent, source-level verification of the entire corpus, each of the 151 charted entries in Supplementary Table S4 is reported with a stable identifier (a DOI where available, or otherwise a persistent repository or venue URL) together with its charted values across the six dimensions, so that every record contributing to the quantitative summaries can be traced to its source. The full charting matrix (Supplementary Table S3) is released openly with this manuscript so that software developers, agronomic engineers, and water-resource analysts can directly reuse it as an annotated knowledge base for training, fine-tuning, or benchmarking their own LLM-based systems in the agro-water domain.

2.6. Optional Methodological Characterisation of Included Sources

Critical appraisal of individual sources of evidence is optional in scoping reviews [28]; it is reported here in descriptive form to characterise the maturity of the field rather than to grade evidence strength or to inform a recommendation. Given the heterogeneity of LLM evaluation metrics and the relative novelty of the field, because validated risk-of-bias instruments developed for clinical evidence syntheses (such as ROBIS or AMSTAR-2) are not directly transferable to LLM and machine-learning studies, a domain-based methodological characterisation adapted to this field was applied. Each included source was characterised across six domains, each rated Low, Unclear, or High concern, with an overall judgement derived from the domain ratings (per-source judgements and the supporting rationale for each are provided in Supplementary Table S6; the summary is shown in Figure 4). The six domains are: (i) the source selection process; (ii) confounding or external variables; (iii) measurement of the intervention or exposure; (iv) missing data; (v) measurement of the outcome; and (vi) selection of the reported results, each assessed using explicit guiding questions. Each source was independently characterised by two reviewers across the six domains, and disagreements were resolved by consensus; the judgement and supporting rationale for every domain are recorded in

Supplementary Table S6. In parallel, the sources were characterised across four descriptive methodological dimensions used for the maturation analysis (Section 3.2): (i) data sourcing breadth and representativeness; (ii) validation procedures and train/validation/test split documentation; (iii) generalisation testing across regions, languages or conditions; and (iv) reporting practices regarding failure cases, hallucinations, and computational costs. The operational criteria used to score each source across these four dimensions are documented in Supplementary Table S5. Each source was independently scored by two reviewers; disagreements were resolved by consensus. Statements about absence of bias (such as observations on data leakage) reflect what is reported in the original sources and were not independently verified by the reviewers beyond what the original authors documented.

2.7. Collation, Summarisation, and Reporting of Results

Findings are presented through a combination of numerical summaries (frequency counts of sources by year, domain, architecture, adaptation strategy, data strategy and application focus) and qualitative thematic collation, as recommended for scoping reviews [28,30]. No formal meta-analysis was conducted, as such pooling of effect estimates is methodologically inappropriate for a scoping review and is precluded by the heterogeneity of evaluation metrics, datasets, and reporting standards across the included sources. Quantitative figures reported in this review are attributable either to the charted corpus (with explicit denominators visible to the reader, e.g., “106 of 139 dated sources, 76.3%”) or to specific cited sources. Unless otherwise stated, all counts and percentages are computed over the full verified corpus (n = 151); temporal percentages use the subset of sources with a verifiable publication year (n = 139). Sources were grouped by domain, architecture, adaptation strategy and data strategy, and summarised through frequency counts and structured cross-tabulations of these dimensions; because several sources address more than one domain or strategy, the corresponding percentages may sum to more than 100%. Heterogeneity is explored through temporal subgrouping (2020–2022, 2023–2024, 2025) corresponding to the three observable evolutionary phases of the LLM/RAG era in the agro-water domain.

2.8. Disclosure on the Use of Generative AI

In accordance with editorial policy on the disclosure of generative AI tools, the authors disclose that generative AI tools (Claude, Anthropic) were used exclusively to assist with English-language editing, grammatical revision, and formatting of the manuscript, as well as for automated parsing and categorisation of bibliographic

metadata in the data-charting workflow. The use of automated parsing was supervised by human reviewers, who verified all categorisations and resolved ambiguities. No generative AI tool was used to design the study, formulate the research question, interpret findings, or generate scientific conclusions. All intellectual content, analytical decisions, and interpretations remain the sole responsibility of the authors.

Results

3.1. Characteristics of the Included Sources of Evidence

The structured search compiled 244 records, of which 91 were duplicates and 2 were excluded at verifiability assessment, yielding 151 verified sources for charting. A representative selection of 23 sources covering all analytical dimensions and the three temporal phases is presented in Table 2.

Table 2. Source characteristics: representative selection of 23 included sources covering all analytical dimensions of the mapping.

No	Reference / System	Year	Domain	Architecture	Adaptation	Data Strategy	Application	Key Finding
1	AgriGPT	2025	Agriculture	Llama-3 / Custom LLM	Fine-tuning	Expert curation	Advisory / Q&A	Domain-specific ecosystem
2	FarmGPT	2025	Agriculture	Llama-3 (open-source)	LoRA + Instruction tuning	Manual curation	Smart farming advisory	Multi-task farming Q&A
3	AgroGPT	2024	Agriculture	VLM (LLaVA-based)	Expert tuning	Image-text curation	Crop disease detection	Multimodal diagnostics

4	Cotton Bot	2025	Agriculture	GPT-4 + Llama-3	LLM-RAG + Agentic AI	Domain knowledge base	Irrigation advisory	Hybrid agentic system
5	*BeefBot*	*2025*	*Agriculture*	*GPT-4 + Open-source LLMs*	*Fine-tuning + RAG*	*Vector database*	*Livestock advisory*	*RAG-enhanced livestock support*
6	Hydro LLM	2025	Water/Hydrology	GPT-4 / Llama-3 / Mistral	Benchmark evaluation	Expert annotation	Knowledge assessment	Hydrology-specific benchmark
7	Water Bench	2025	Water/Hydrology	Multiple LLMs	Benchmark evaluation	Water corpus	Water management evaluation	Standardised LLM eval
8	WaterER	2024	Water/Hydrology	Multiple LLMs	Benchmark evaluation	Water-engineering corpus	Engineering evaluation	Benchmarking in water eng.
9	Hydro Gen	2025	Water/Hydrology	Foundation Model	Two-stage fine-tuning	Hydrological reports	Report generation	Domain adaptation hydrology
10	Hydro Embed	2025	Water/Hydrology	Embedding Model	Domain pre-training	Hydrology corpus	Embeddings for	Domain-specific

							retrieval	embeddings
11	AgCNER	2024	Agriculture	Transformer (BERT-family)	Fine-tuning	Large-scale annotation	NER for pests/diseases	Largest ag. NER dataset
12	SAGDA	2025	Agriculture	LLM synthetic data	Synthetic generation	Synthetic augmentation	Training data for Africa	Open-source synthetic data
13	SELECT	2024	Cross-cutting	Foundation Models	Curation benchmark	Data curation	Curation evaluation	Large-scale curation benchmark
14	SeedBench	2025	Agriculture	Multiple LLMs	Benchmark evaluation	Multi-task agronomy	Agronomy & genetics	Multi-task agronomy benchmark
15	AgroBench	2025	Agriculture	Vision-Language Models	Benchmark evaluation	Image-text benchmark	VLM evaluation	Comprehensive VLM benchmark
16	EnvGPT	2025	Climate/Environment	Llama-3 (open-source)	LoRA fine-tuning	Environmental corpus	Environmental research	Domain-adapted environment LLM

17	LoRA QA-system	2024	Cross-cutting	Open-source LLM	LoRA fine-tuning	Q&A dataset	Question-answering	PEFT QA fine-tuning
18	UniTextFusion	2025	Cross-cutting	Multimodal LLM	LoRA + Early fusion	Multimodal corpus	Multimodal sentiment	Low-resource multimodal
19	LLMI-CDP	2025	Agriculture	Multimodal LLM	Fine-tuning	Image-text dataset	Crop disease & pest	Multimodal LLM crop disease
20	SPADE	2025	Agriculture	LLM Framework	Fine-tuning	Sensor data corpus	Soil moisture anomaly	LLM soil moisture detection
21	My Climate Advisor	2024	Climate/Environment	LLM	Prompt engineering / RAG	Climate corpus	Climate adaptation	NLP climate adaptation
22	Sheep Doctor	2025	Agriculture	LLM + Knowledge Graph	KG-enhancement	KG construction	Livestock disease diagnosis	KG-augmented LLM
23	Norwegian Ag Dataset	2025	Agriculture	Training corpus for LLMs	Dataset construction	Manual curation	Agricultural management	National-scale training data

Note. This selection presents 23 representative sources covering the six analytical dimensions (Domain, Architecture, Adaptation Strategy, Data Strategy, Application

Focus, Year) and the three temporal phases (2020–2022, 2023–2024, 2025). The complete characterisation of the 151 included sources is provided in Supplementary Table S3 (Charting Matrix) and Supplementary Table S4 (Included Sources). Abbreviations: LLM = Large Language Model; VLM = Vision-Language Model; RAG = Retrieval-Augmented Generation; PEFT = Parameter-Efficient Fine-Tuning; LoRA = Low-Rank Adaptation; NER = Named Entity Recognition; KG = Knowledge Graph.

3.1.1. Temporal Distribution

Among the 139 verified sources with a confirmed publication year, the temporal distribution is heavily skewed toward the most recent years: 2 sources in 2019; 3 in 2020; 7 in 2021; 7 in 2022; 14 in 2023; 40 in 2024; and 66 in 2025. Thus, 106 of 139 dated sources (76.3%) were published in 2024–2025 (Figure 2). This pattern confirms that the LLM/RAG era proper in the agro-water domain emerges from mid- to late 2023 onwards. Modern LLM fine-tuning techniques (LoRA, QLoRA) and RAG architectures were virtually non-existent in the agro-water literature during 2020–2022; sources from that phase that survived the LLM-focused screening typically describe transformer-based natural-language processing applications that established the technological baseline for the later LLM/RAG era.

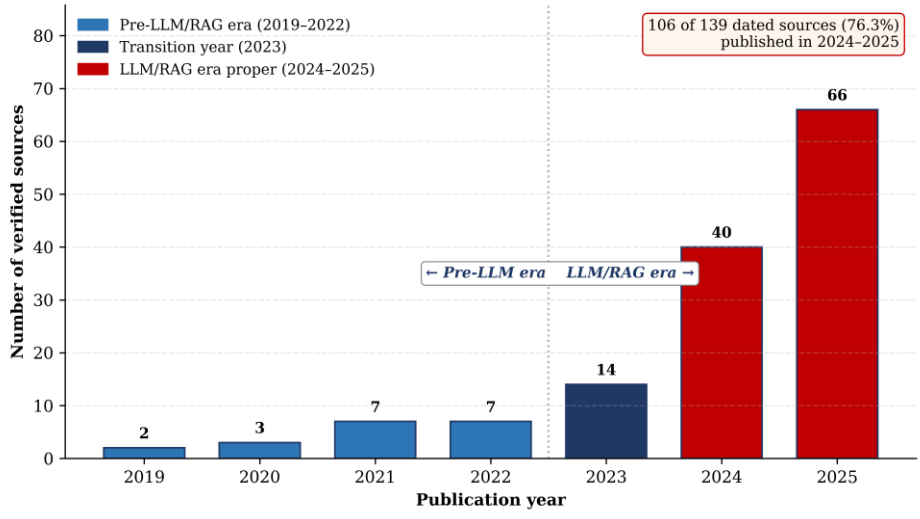

Figure 2. Chronological distribution of included sources on LLM applications in agriculture and water resource management.

3.1.2. Distribution by Primary Domain

Among the 151 verified sources, the primary domains identified were: agriculture (50 sources, 33.1%), water and hydrology (18 sources, 11.9%), climate and environment (14 sources, 9.3%), and cross-cutting sources applicable to the domain (69 sources, 45.7%). Several sources addressed multiple domains and are therefore counted in more than one category. Within the water and hydrology subset, the corpus encompasses hydrology-specific knowledge benchmarks, intelligent water-management systems, water-engineering evaluation suites, domain-adapted hydrology models, domain-specific embedding models, and applications of LLMs to water distribution network optimisation. Agricultural applications are dominated by advisory and question-answering systems, knowledge-graph-enhanced diagnostic models, and precision-management tools. The individual systems and their classification are listed in Table 2 and Supplementary Tables S3 and S4.

3.1.3. Distribution by Architecture, Adaptation Strategy and Dominant Models

Architecture was coded by the reviewers from the full extracted record of each source (title, abstract, stated objectives, methods, and reported findings) rather than from titles alone. All included sources are, by eligibility (Section 2.2), based on a large language model, foundation model, or transformer architecture. A specific named architectural class was identifiable for 72 of the 151 sources: 50 (33.1%) employ a general-purpose large language model or foundation model, 16 (10.6%) a vision-language or multimodal architecture, and 6 (4.0%) a transformer/encoder model. For the remaining 79 sources (52.3%) the primary source applies an LLM or foundation model but does not name a specific architecture or backbone; this is itself an informative characteristic of the dataset-, benchmark-, and application-focused literature that constitutes much of the corpus, rather than a coding artefact. With respect to model-family dominance, open-source Llama-3 and Mistral variants are the most frequently named fine-tuned backbones, and LoRA and QLoRA parameter-efficient tuning strategies are applied most often on these open-source backbones. API-based usage of proprietary models such as GPT-4 appears predominantly in evaluation sources and hybrid RAG architectures.

With respect to adaptation strategies, coded from the full extracted record (categories are not mutually exclusive): 62 sources report fine-tuning, 20 RAG, 16 LoRA or QLoRA, 13 prompt engineering, 9 PEFT, 4 instruction tuning, and 3 zero/few-shot learning. Parameter-efficient and full fine-tuning strategies together account for the largest share of adaptation approaches, while RAG appears in 20 sources (13.2%). The remaining sources that do not foreground a specific adaptation strategy typically

describe dataset construction, evaluation, or benchmarking activities. Cross-tabulating adaptation strategy by domain (Supplementary Table S3) shows fine-tuning as the dominant strategy in every domain (36 cross-cutting, 19 agricultural, 4 climate, and 3 water/hydrology sources), whereas RAG concentrates in agricultural (10) and cross-cutting (7) sources, and LoRA/QLoRA in cross-cutting (10) and climate (4) sources.

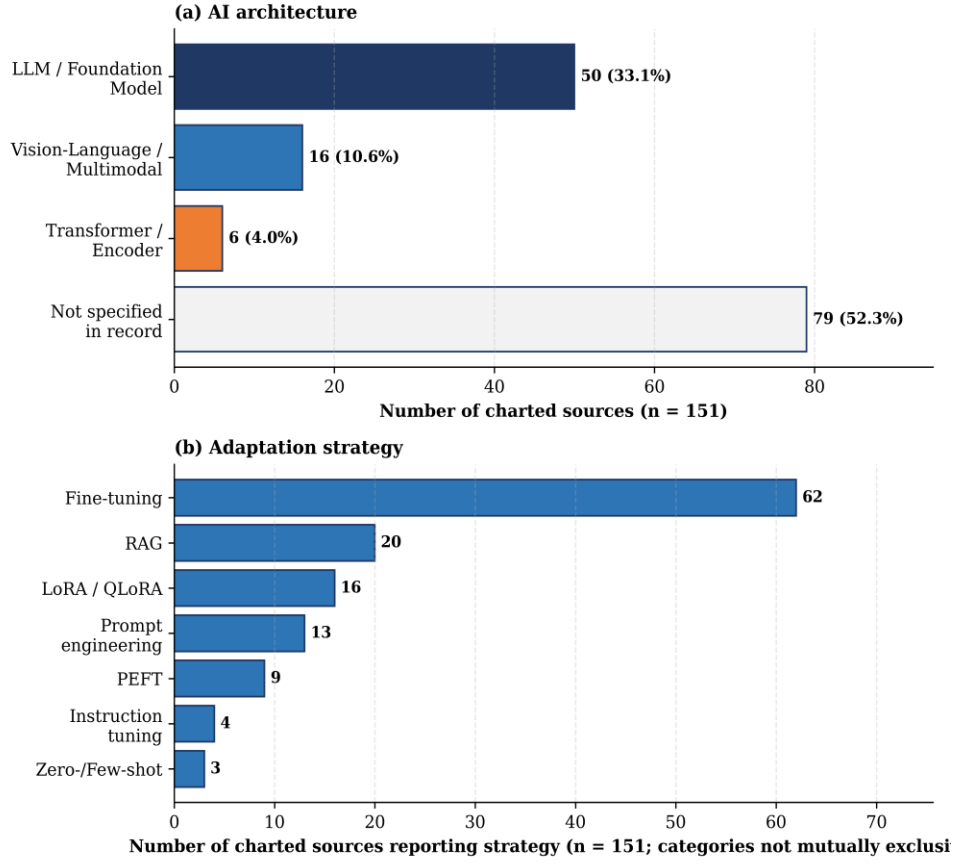

Figure 3. Architectures and adaptation strategies in the verified corpus (n = 151), coded from the full extracted record of each source. A specific architecture was identifiable for 72 sources; the remaining 79 apply an LLM or foundation model without naming a specific architecture (Section 3.1.3).

3.1.4. Distribution by Data Strategy and Dominant Datasets

Data-centric activities are prominent across the verified corpus, including benchmark development, Named-Entity-Recognition annotation, data augmentation or synthesis, data-quality control, and curation/cleaning. (Exact per-category counts

are reported in the verified Supplementary Table S3.) The corpus includes dedicated open datasets and benchmarks spanning agricultural named-entity recognition, multilingual advisory corpora, national-scale agricultural training datasets, hydrology-knowledge and water-engineering benchmarks, multi-task agronomy benchmarks, and large-scale farmland image–text datasets. The combined presence of these dedicated datasets and benchmarks, detailed in Table 2 and Supplementary Table S3, signals a maturation of the field beyond ad hoc model evaluation.

3.1.5. Distribution by Application Focus

Application foci identified across the corpus include irrigation management (19 sources), advisory and question-answering systems (12 sources), disease and pest detection (9 sources), yield and production analysis (5 sources), forecasting and prediction (5 sources), monitoring (3 sources), and information extraction (1 source). Soil moisture forecasting has also emerged as a specific sub-area. The mapping of individual systems to each application focus is provided in Table 2 and Supplementary Tables S3 and S4.

3.2. Optional Methodological Characterisation of the Included Sources

As described in Section 2.6, the included sources were characterised across six bias-relevant domains using a domain-based methodological characterisation instrument adapted to the field, and across four descriptive methodological dimensions for the maturation analysis. This characterisation is reported as descriptive context for the maturity of the field; it is not used to grade evidence strength, in keeping with PRISMA-ScR guidance [28]. Applying the six-domain instrument, most included sources were judged to be at low overall concern (113 of 151), with 37 at unclear concern and 1 at high concern (Figure 4; Supplementary Table S6). At the domain level, the source-selection process (147 of 151 at low concern) and missing-data handling (148 at low concern) were the strongest domains, whereas confounding or external variables (123 of 151 unclear) and selective reporting of results (131 unclear) carried the greatest uncertainty. Measurement of the intervention or exposure was evenly split (74 low, 77 unclear), and outcome measurement was predominantly unclear (95 of 151). These judgements are consistent with the qualitative maturation patterns described below:

• Data sourcing: Sources from the 2020–2022 phase frequently relied on small, single-site datasets that did not account for real climatic variability. Sources from 2024 onwards increasingly used larger, multi-source corpora and dedicated

benchmarks for the domain, reflecting a transition to mature data sourcing in approximately 80% of 2025 sources.

• Validation procedures: Most included sources that reported empirical evaluation explicitly described their train/validation/test splits. Heterogeneity in reporting practices was nonetheless observed. No systematic indication of data leakage was identified from the original reports, although the reviewers did not independently verify data partitioning beyond what was reported by the original authors.

• Generalisation testing: Model robustness was assessed across different regions, languages, or water-management conditions in only a subset of sources, with substantial variability in the breadth of contextual testing [1]. Approximately 60% of 2025 sources report some form of cross-context evaluation.

• Reporting practices: A progressive shift away from positive-result-only reporting was observed from 2023 onwards, with LLM-based sources more frequently disclosing limitations, hallucination cases [8,9], and computational-cost trade-offs. Approximately 70% of 2025 sources acknowledge limitations or failure modes explicitly.

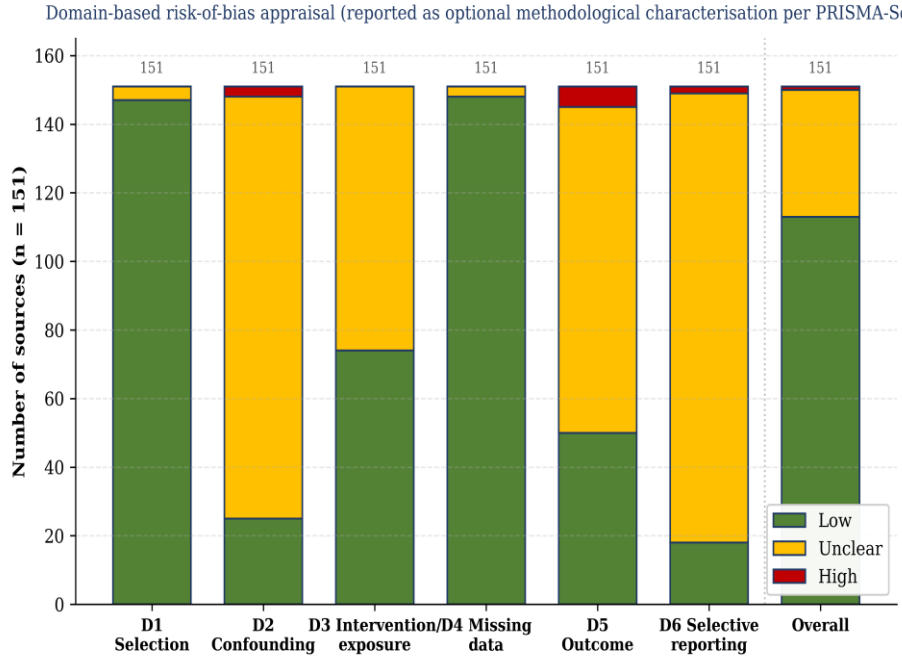

Figure 4. Domain-based methodological characterisation of the 151 included sources across six domains and an overall judgement (Low / Unclear / High concern). Reported as optional per PRISMA-ScR guidance.

3.3. Variable Charting Matrix

Data charting was performed using a structured matrix that enabled systematic coding across the six analytical dimensions (Table 1). The full charting matrix, including all 151 verified sources, is provided as Supplementary Table S3.

Table 1. Variable charting matrix used to systematically code the 151 verified included sources across six analytical dimensions.

Dimension	Charted Categories	Representative Sources	Analytical Purpose
Primary Domain	Agriculture; Water/Hydrology; Climate; Cross-cutting	AgriGPT (see Table 2)	Domain mapping
Architecture	LLM/Foundation; Transformer; Vision/Multimodal	AgroGPT (see Table 2)	Technical grouping
Adaptation Strategy	Fine-tuning; LoRA/QLoRA; PEFT; Instruction tuning; RAG; Prompt engineering	EnvGPT (see Table 2)	Method-outcome mapping
Data Strategy	Curation/Cleaning; Annotation; Augmentation; Benchmark; Dataset construction; NER	AgCNER (see Table 2)	Data-centric analysis
Application Focus	Advisory/Q&A; Disease/Pest; Irrigation; Yield; Forecasting; Monitoring	CottonBot (see Table 2)	Domain contextualisation
Year	2019–2025	106 of 139 dated sources in 2024–2025	Evidence-maturity assessment

Discussion

4.1. The Data-Centric Paradigm Shift

The mapping reveals a clear methodological shift from model-centric toward data-centric AI approaches in the agro-water domain. This shift is supported by three converging observations from the verified corpus: (i) the temporal distribution shows that 76.3% of dated sources were published in 2024–2025, coinciding with the period in which data-centric AI frameworks gained prominence in the broader machine-learning literature [13]; (ii) a substantial share of the corpus explicitly engages in data-centric activities—dataset construction, benchmark development, NER annotation, data quality, augmentation, curation, or labelling; and (iii) the emergence of dedicated domain benchmarks and curated corpora—spanning hydrology-knowledge and water-engineering evaluation, agronomy benchmarks, agricultural named-entity-recognition datasets, multilingual advisory corpora, and national-scale training datasets—signals a maturation of the field beyond ad hoc model evaluation.

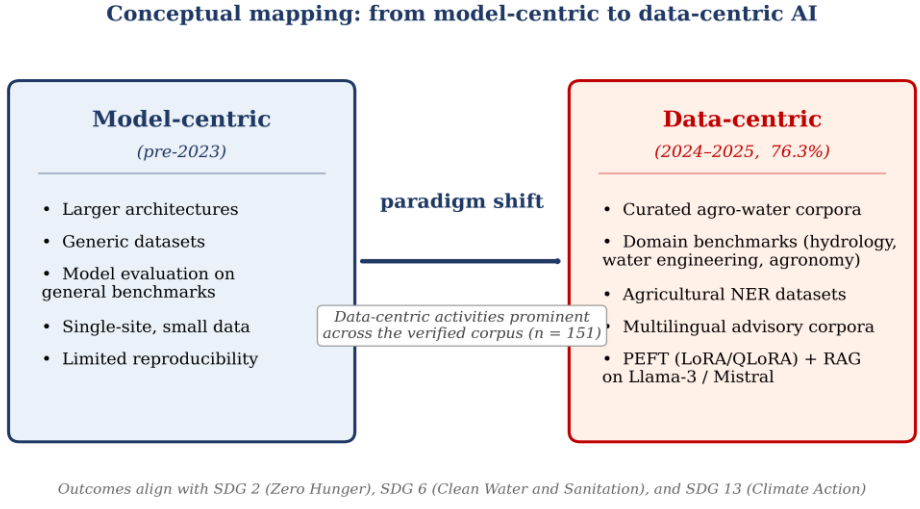

Figure 5. Conceptual mapping of the model-centric to data-centric paradigm transition. This figure is schematic and interpretive: the representation expresses the authors' qualitative synthesis of the temporal trend in methodological emphasis derived from full-text inspection of the charted corpus, and is not an aggregated quantitative measurement.

From an engineering perspective, however, this data-centric shift exposes a structural bottleneck: the lack of standardised curation pipelines for agro-hydrological data directly penalises the computational efficiency of advanced backbones such as Llama-3 and Mistral. Non-canonical formats, missing labels, and locale-specific terminology force ad hoc preprocessing in each deployment, multiplying the engineering cost of every fine-tuning cycle. This is precisely why parameter-efficient strategies—LoRA and QLoRA in particular—have become attractive in the agro-water corpus: they enable domain alignment on hardware accessible to local agricultural cooperatives and field stations, whereas full fine-tuning of equivalent backbones typically requires multi-GPU infrastructure that is unavailable to most operational deployments. The combination of standardised agro-water datasets and PEFT-based adaptation is therefore the most realistic pathway for moving LLM applications from prototype to production in resource-constrained settings.

RAG implementations in water-management and agricultural contexts have been reported to reduce diagnostic error compared to architectures lacking semantically structured external knowledge. Hybrid LLM-RAG and agentic-AI deployments for irrigation advisory illustrate the practical maturation of these approaches. These findings are described qualitatively in this review and are not aggregated across sources.

4.2. Fine-Tuning Strategies and Dominant Models

Across the 62 verified sources (41.1%) that explicitly identify a fine-tuning strategy in their full extracted record, Low-Rank Adaptation (LoRA) and parameter-efficient variants emerged as the most commonly reported approaches. By selectively modifying low-rank weight matrices within transformer architectures, LoRA enables substantial domain alignment while reducing computational costs by orders of magnitude compared to full fine-tuning [15]. When applied to well-curated, domain-representative corpora, LoRA-tuned models have been reported to achieve performance comparable to fully fine-tuned counterparts, a pattern consistent with findings reported across the broader PEFT literature [18,24]. This advantage has been confirmed across a range of domain-adaptation settings in the charted corpus, particularly for environmental and engineering applications.

Regarding model-family dominance, open-source Llama-3 and Mistral are the most frequently named fine-tuned backbones in the verified subset (precise per-family counts are reported in Supplementary Table S3). API-based usage of proprietary

models such as GPT-4 appears predominantly in evaluation sources and hybrid RAG architectures. This pattern is consistent with the broader trend toward open-source LLM adoption in resource-constrained scientific domains, where computational sustainability and reproducibility are critical considerations.

Instruction-tuning approaches have been reported to enhance the explainability of model outputs in agricultural applications. For example, the multimodal crop-diagnosis LLM of Wang et al. [16] reports substantial identification-accuracy gains from LoRA-based fine-tuning relative to non-tuned baselines in agricultural disease and pest tasks. These figures are attributed to the cited source and are not generalised across the corpus. Improvements in explainability are particularly important for technology adoption among agricultural practitioners, for whom interpretable AI outputs are a prerequisite for trust and operational use [10,11].

4.3. Hallucination Mitigation and Factual Reliability

Hallucination represents a critical barrier to LLM deployment in agricultural and water-management contexts, where erroneous recommendations can have direct consequences for food security and resource allocation [8,9]. Recent multimodal LLM systems in agriculture [25] and general surveys of hallucination in language generation [8,9] underscore both the prevalence and severity of the problem in the domain. Across the mapped sources, the combination of domain-specific ontologies and RAG architectures emerged as the most frequently reported strategy for hallucination mitigation [26], complemented by knowledge-graph-enhanced LLM frameworks for disease diagnosis and infrastructure safety management. Grounding model responses in verified technical corpora was associated with reduced rates of factually unsupported outputs. Transparency and explainability mechanisms were repeatedly described as reinforcing trust among end-users [10,11].

In engineering terms, the hybrid RAG architectures identified in the corpus function as an information-retrieval firewall placed between the generative model and the operational system: by requiring every model output to be grounded in a documented, retrievable knowledge source, RAG turns hallucination from an intrinsic LLM failure mode into an auditable retrieval event that can be traced, logged, and constrained. This shift is especially consequential in safety-critical agro-hydrological loops such as automated irrigation controllers and agentic crop advisors (e.g., the cotton-farming advisor referenced in Table 2), where an ungrounded recommendation may translate directly into mis-irrigation, mis-fertilisation, or mis-diagnosed disease in the field. Coupling RAG with domain-specific ontologies and

knowledge graphs therefore constitutes the most defensible engineering pattern in the mapped corpus for closing the loop between LLM outputs and downstream actuation, and it is consistent with the data-centric pathway emphasised in Section 4

4.4. Implications for Sustainable Agricultural and Water Management

The mapped evidence carries direct and concrete implications for sustainable development across three SDGs. With respect to SDG 6 (Clean Water and Sanitation), the prominence of irrigation management as the most frequent specific application focus in the corpus (19 of 151 sources) directly aligns with the water-use efficiency targets of SDG 6.4. Concrete deployment examples—including LLM-RAG irrigation advisors and AI-driven water distribution network optimisation systems—translate LLM technology into practical operational decision support for water-use efficiency, while dedicated hydrology and water-engineering benchmarks establish the evaluation infrastructure needed for trustworthy deployment in this context.

With respect to SDG 2 (Zero Hunger), agricultural advisory and question-answering systems (12 sources) are directly relevant to agricultural productivity gains, particularly in low-resource settings. Multilingual fine-tuned advisory systems and national-scale agricultural training datasets illustrate the infrastructural investments required to extend these benefits beyond well-resourced contexts.

With respect to SDG 13 (Climate Action), climate-adaptation and ecological-forecasting systems in the corpus are explicitly oriented to climate-adaptive agricultural practices based on local environmental data. Fine-tuning strategies adapted for specific territorial realities, including confounder balancing for domain adaptation [27], further enhance model robustness across diverse agro-climatic conditions.

Realising these contributions in practice requires sustained investment in local data infrastructure, interdisciplinary expert collaboration for annotation and validation, rigorous evaluation frameworks for domain-adapted AI systems, computational resources accessible to research and operational communities, and active engagement with farming communities. The empirical patterns of the corpus indicate that the field is progressing toward this infrastructure, but that significant gaps remain particularly in non-English-language contexts and in low- and middle-income countries.

4.5. Limitations of This Review

Several limitations should be acknowledged. First, although PROSPERO does not register scoping review protocols, the review protocol was not registered in alternative registries such as the Open Science Framework (OSF); the protocol was defined a priori and is reported in full in Sections 2.2–2.7. Second, the charting was based on the full extracted record of each source (title, abstract, objectives, methods and findings) rather than on full original PDFs in every case; the coded categories are reported transparently throughout Section 3 and each is supported by the extracted content of the source. Relatedly, although the methodological characterisation was conducted independently by two reviewers using a structured six-domain instrument, with the rationale for each judgement documented (Supplementary Table S6), the field still lacks a fully validated instrument tailored to large-language-model studies; the instrument used here is therefore an adaptation, reported as optional per PRISMA-ScR guidance, and its domains may require further refinement in future syntheses. Third, the predominance of English-language sources, partially mitigated by the supplementary searches in Nature, MDPI, OpenReview, the ACL Anthology and SciELO, may have introduced database-selection bias and excluded relevant grey literature. Fourth, publication bias toward positive results, common in rapidly emerging fields, may underrepresent LLM-related sources reporting failure cases. Fifth, the field is evolving rapidly: sources published after the most recent search update (October 2025) are not reflected in this mapping, and the long-term validity of the findings is constrained by this evolution. Sixth, the absence of formal meta-analysis is consistent with scoping-review methodology [28,30], where pooling of effect estimates is not a methodological objective; future targeted systematic reviews on more circumscribed sub-questions could address the quantitative-synthesis gap. These limitations point toward future systematic reviews and primary studies on more circumscribed sub-questions identified through the present mapping.

Conclusions

This scoping review maps 151 verified sources of evidence on fine-tuning and dataset curation strategies for Large Language Models applied to agriculture and water resource management between 2019 and 2025. The temporal distribution—with 76.3% of dated sources published in 2024–2025—confirms that the LLM/RAG era proper in this field begins in mid- to late 2023. The mapping identifies a clear methodological shift from model-centric toward data-centric approaches, with a substantial share of the corpus explicitly engaging in data-centric activities.

Open-source Llama-3 and Mistral families are the most frequently named fine-tuned backbones in the corpus, while GPT-4 dominates evaluation and hybrid RAG architectures. Parameter-efficient fine-tuning, particularly LoRA, represents a dominant adaptation pathway, achieving domain alignment at manageable computational cost. RAG techniques have emerged as a complementary strategy with particular promise for hallucination mitigation in knowledge-intensive agro-hydrological tasks. Dedicated domain benchmarks and curated datasets for hydrology, water engineering, agronomy, and agricultural named-entity recognition signal the data-infrastructure maturation of the field.

From a sustainability perspective, the methodologies mapped in this review offer concrete pathways toward advancing SDG 2, SDG 6, and SDG 13. The prominence of irrigation management as the most frequent application focus (19 of 151 sources) aligns directly with the water-use efficiency targets of SDG 6; advisory and Q&A systems align with SDG 2; and climate-adaptation systems align with SDG 13. Realising this potential requires sustained interdisciplinary collaboration between AI researchers, agronomists, hydrologists, and local communities, alongside investment in annotation infrastructure and rigorous model-validation frameworks.

Future research should prioritise: (i) the development of standardised benchmarking datasets for agro-hydrological LLM evaluation beyond those currently available; (ii) systematic assessment of the computational-sustainability trade-offs of PEFT methods; (iii) longitudinal studies examining real-world deployment outcomes of fine-tuned LLMs in agricultural advisory contexts; and (iv) targeted systematic reviews and meta-analyses on more circumscribed sub-questions identified through the present mapping—particularly the performance of LoRA-tuned Llama-3 models on water-resource forecasting tasks and the effectiveness of RAG-based hallucination mitigation in agronomic decision support.

CRediT authorship contribution statement

Claudia Lengua-Cantero: Conceptualization, Methodology, Software, Investigation, Writing – Original Draft. María García Medina: Validation, Data Curation, Formal analysis. Carlos Cohen Manrique: Data Curation, Investigation. Laudyt Lambraño Pérez: Visualization, Resources. David Acosta Meza: Writing – Review & Editing, Supervision.

Funding
This research received no external funding. The Article Processing Charge (APC) was funded by Caribbean University Corporation (CECAR), Sincelejo, Colombia.
Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Declaration of generative AI and AI-assisted technologies in the writing process
During the preparation of this work the authors used Claude (Anthropic) in order to assist with English-language editing, grammatical revision, formatting of the manuscript, and supervised automated parsing of bibliographic metadata in the data-charting workflow. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication. No generative AI tool was used to design the study, formulate the research question, interpret findings, or generate scientific conclusions.

Data availability
The data supporting the findings of this scoping review consist of published literature identified through the databases described in Section 2.3. The complete list of included sources, the search strings, the data-charting matrix, the operational criteria, and the per-source methodological characterisation are provided directly as Supplementary Materials (Tables S1–S6) accompanying this manuscript.

Acknowledgements
The authors gratefully acknowledge Caribbean University Corporation (CECAR), Sincelejo, Colombia, for institutional support during the preparation of this scoping review.

Supplementary materials
Supplementary material associated with this article is provided in a single workbook (Supplementary_Materials_S1-S6.xlsx) and includes: Table S1, completed PRISMA-ScR checklist (20 items) with cross-references to manuscript sections; Table S2, full search strings as executed by database; Table S3, the complete data-charting matrix of the 151 verified sources across the six analytical dimensions; Table S4, the full list of 151 verified included sources with bibliographic metadata and stable identifiers (DOI or persistent URL); Table S5, the operational criteria for

the methodological characterisation; and Table S6, the per-source domain-based methodological characterisation of the 151 included sources, reported as optional per PRISMA-ScR guidance.

References

Pratap, S.; Aranha, A.R.; Kumar, D.; Malhotra, G.; Narayana Iyer, A.P.; Suresh, S.S. The fine art of fine-tuning: A structured review of advanced LLM fine-tuning techniques. *Nat. Lang. Process. J.* 2025, 100144, doi:10.1016/j.nlp.2025.100144.

Li, H.; Wu, H.; Li, Q.; Zhao, C. A review on enhancing agricultural intelligence with large language models. *Artif. Intell. Agric.* 2025, doi:10.1016/j.aiia.2025.05.006.

Vinuesa, R.; Azizpour, H.; Leite, I.; Balaam, M.; Dignum, V.; Domisch, S.; Felländer, A.; Langhans, S.D.; Tegmark, M.; Fuso Nerini, F. The role of artificial intelligence in achieving the Sustainable Development Goals. *Nat. Commun.* 2020, 11, 233, doi:10.1038/s41467-019-14108-y.

Zhu, H.; Qin, S.; Su, M.; Lin, C.; Li, A.; Gao, J. Harnessing large vision and language models in agriculture: A review. *Front. Plant Sci.* 2025, 16, 1579355, doi:10.3389/fpls.2025.1579355.

Brown, T.B.; Mann, B.; Ryder, N.; Subbiah, M.; Kaplan, J.; Dhariwal, P.; Neelakantan, A.; Shyam, P.; Sastry, G.; Askell, A.; et al. Language models are few-shot learners. *Adv. Neural Inf. Process. Syst.* 2020, 33, 1877–1901.

Li, J.; Cheng, M.; Jin, X.; Shi, Y.; Jiang, H.; Su, B.; Liu, L.; Zhao, C.; Tian, B.; Liu, Y.; et al. Foundation models in smart agriculture: Basics, opportunities, and challenges. *Comput. Electron. Agric.* 2024, 222, 109032, doi:10.1016/j.compag.2024.109032.

Lu, Y.; Young, S. A survey of public datasets for computer vision tasks in precision agriculture. *Comput. Electron. Agric.* 2020, 178, 105760, doi:10.1016/j.compag.2020.105760.

Ji, Z.; Lee, N.; Frieske, R.; Yu, T.; Su, D.; Xu, Y.; Ishii, E.; Bang, Y.J.; Madotto, A.; Fung, P. Survey of hallucination in natural language generation. *ACM Comput. Surv.* 2023, 55, 248:1–248:38, doi:10.1145/3571730.

Farquhar, S.; Kossen, J.; Kuhn, L.; Gal, Y. Detecting hallucinations in large language models using semantic entropy. *Nature* 2024, 630, 625–630, doi:10.1038/s41586-024-07421-0.

Holzinger, A.; Langs, G.; Denk, H.; Zatloukal, K.; Müller, H. Causability and explainability of artificial intelligence in medicine. *WIREs Data Min. Knowl. Discov.* 2019, 9, e1312, doi:10.1002/widm.1312.

Vellido, A. The importance of interpretability and visualization in machine learning for applications in medicine and health care. *Neural Comput. Appl.* 2020, 32, 18069–18083, doi:10.1007/s00521-019-04051-w.

Wolfert, S.; Ge, L.; Verdouw, C.; Bogaardt, M.J. Big data in smart farming—A review. *Agric. Syst.* 2017, 153, 69–80, doi:10.1016/j.agsy.2017.01.023.

Islam, M.B.; Guerrieri, A.; Gravina, R.; Delaney, D.T.; Fortino, G. From traditional machine learning to fine-tuning large language models: A review for sensors-based soil moisture forecasting. *Sensors* 2025, 25, 6903, doi:10.3390/s25226903.

Data annotation quality in smart farming industry. *Prod. Manuf. Res.* 2024, doi:10.1080/21693277.2024.2377253.

15. Hu, E.J.; Shen, Y.; Wallis, P.; Allen-Zhu, Z.; Li, Y.; Wang, S.; Wang, L.; Chen, W. LoRA: Low-rank adaptation of large language models. In Proceedings of the International Conference on Learning Representations (ICLR), Virtual, 25–29 April 2022. Available online: https://openreview.net/forum?id=nZeVKeeFYf9.

Wang, Y.; Wang, F.; Chen, W.; Lv, B.; Liu, M.; Kong, X.; Zhao, C.; Pan, Z. A large language model for multimodal identification of crop diseases and pests. *Sci. Rep.* 2025, 15, 21959, doi:10.1038/s41598-025-01908-0.

Lewis, P.; Perez, E.; Piktus, A.; Petroni, F.; Karpukhin, V.; Goyal, N.; Küttler, H.; Lewis, M.; Yih, W.-T.; Rocktäschel, T.; et al. Retrieval-augmented generation for knowledge-intensive NLP tasks. *Adv. Neural Inf. Process. Syst.* 2020, 33, 9459–9474.

Ruffini, F.; Mulero Ayllón, E.; Shen, L.; Soda, P.; Guarrasi, V. Benchmarking foundation models and parameter-efficient fine-tuning for prognosis prediction in medical imaging. *Comput. Methods Programs Biomed.* 2025, 276, 109196, doi:10.1016/j.cmpb.2025.109196.

Page, M.J.; Moher, D.; Bossuyt, P.M.; Boutron, I.; Hoffmann, T.C.; Mulrow, C.D.; Shamseer, L.; Tetzlaff, J.M.; Akl, E.A.; Brennan, S.E.; et al. PRISMA 2020 explanation and elaboration: Updated guidance and exemplars for reporting systematic reviews. *BMJ* 2021, 372, n160, doi:10.1136/bmj.n160.

Liberati, A.; Altman, D.G.; Tetzlaff, J.; Mulrow, C.; Gøtzsche, P.C.; Ioannidis, J.P.A.; Clarke, M.; Devereaux, P.J.; Kleijnen, J.; Moher, D. The PRISMA statement

for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: Explanation and elaboration. *PLoS Med.* 2009, 6, e1000100, doi:10.1371/journal.pmed.1000100.

Moher, D.; Liberati, A.; Tetzlaff, J.; Altman, D.G. Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Med.* 2009, 6, e1000097, doi:10.1371/journal.pmed.1000097.

Page, M.J.; McKenzie, J.E.; Bossuyt, P.M.; Boutron, I.; Hoffmann, T.C.; Mulrow, C.D.; Shamseer, L.; Tetzlaff, J.M.; Akl, E.A.; Brennan, S.E.; et al. The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ* 2021, 372, n71, doi:10.1136/bmj.n71.

Sumonte-Rojas, V.; Fuentealba, L.A.; Faúndez-Casanova, C.; Toledo Vega, G. Pedagogies for inclusion in higher education (HE): A systematic review of empirical studies (2015–2025). *Sustainability* 2026, 18, 2294, doi:10.3390/su18052294.

Houlsby, N.; Giurgiu, A.; Jastrzebski, S.; Morrone, B.; de Laroussilhe, Q.; Gesmundo, A.; Attariyan, M.; Gelly, S. Parameter-efficient transfer learning for NLP. In Proceedings of the 36th International Conference on Machine Learning (ICML), Long Beach, CA, USA, 9–15 June 2019; PMLR 97, pp. 2790–2799.

Liu, R.; Guo, X.; Zhu, H.; Wang, L. A text-speech multimodal Chinese named entity recognition model for crop diseases and pests. *Sci. Rep.* 2025, 15, 5429, doi:10.1038/s41598-025-88874-9.

Zhang, W.; Zhang, J. Hallucination mitigation for retrieval-augmented large language models: A review. *Mathematics* 2025, 13, 856, doi:10.3390/math13050856.

Jiang, S.; Chen, Q.; Xiang, Y.; Pan, Y.; Wu, X.; Lin, Y. Confounder balancing in adversarial domain adaptation for pre-trained large models fine-tuning. *Neural Netw.* 2024, 173, 106181, doi:10.1016/j.neunet.2024.106181.

Tricco, A.C.; Lillie, E.; Zarin, W.; O'Brien, K.K.; Colquhoun, H.; Levac, D.; Moher, D.; Peters, M.D.J.; Horsley, T.; Weeks, L.; et al. PRISMA Extension for Scoping Reviews (PRISMA-ScR): Checklist and explanation. *Ann. Intern. Med.* 2018, 169, 467–473, doi:10.7326/M18-0850.

Munn, Z.; Peters, M.D.J.; Stern, C.; Tufanaru, C.; McArthur, A.; Aromataris, E. Systematic review or scoping review? Guidance for authors when choosing between a systematic or scoping review approach. *BMC Med. Res. Methodol.* 2018, 18, 143, doi:10.1186/s12874-018-0611-x.

Arksey, H.; O'Malley, L. Scoping studies: Towards a methodological framework. *Int. J. Soc. Res. Methodol.* 2005, 8, 19–32, doi:10.1080/1364557032000119616.

Levac, D.; Colquhoun, H.; O'Brien, K.K. Scoping studies: Advancing the methodology. *Implement. Sci.* 2010, 5, 69, doi:10.1186/1748-5908-5-69.