

PRISMA 2020 Checklist

Agentic Artificial Intelligence in Agriculture: A Systematic Literature Review of Architectures, Applications, Challenges, and Future Directions

Section	Item	Checklist item	Location in manuscript
TITLE	1	Identify the report as a systematic review.	Title page: "A Systematic Literature Review" in the manuscript title.
ABSTRACT	2	See the PRISMA 2020 for Abstracts checklist.	Abstract, page 1: states objective, database and search date, eligibility criteria, screening numbers, included-study count, coding dimensions, principal results with figures, and the review's added value.
INTRODUCTION	3	Describe the rationale for the review in the context of existing knowledge.	Section 1.1-1.4, pages 1-4, and Section 1.5 "Related Reviews and the Gap This Review Addresses", page 4.
	4	Provide an explicit statement of the objective(s) or question(s) the review addresses.	Section 1.6 "Objectives" and Table 1 "Research Questions", page 5.
METHODS	5	Specify the inclusion and exclusion criteria for the review and how studies were grouped for synthesis.	Section 2.3 "Inclusion Criteria" and Section 2.4 "Exclusion Criteria", Table 3, page 6; grouping into nine application domains and six coding dimensions is set out in Section 5.
	6	Specify all databases, registers, websites, organisations, reference lists and other sources searched, and the date each was last searched.	Section 2.1 "Review Protocol", page 5: Scopus, searched 5 August 2026. Single-database use is stated as a limitation in Section 10.2.
	7	Specify all search strategies, including filters and limits used, such that they can be repeated.	Section 2.2 "Search Strategy" and Table 2, page 5, giving the complete Boolean query, field codes, wildcards, year and language filters, and document-type filters; the same query is provided as a standalone supplementary text file.
	8	Specify the methods used to decide whether a record met the inclusion criteria, including how many reviewers screened each record, whether they worked independently, and if applicable details of automation tools.	Section 2.5 "Screening Process", pages 6-7: a scripted, rule-based procedure applied to Scopus metadata by a single research team, followed by manual reading of records reaching the eligibility stage. Single-team coding, and its implications for inter-rater reliability, are stated directly rather than implying independent dual screening.
	9	Specify the methods used to collect data from reports, including how many reviewers collected data, whether they worked independently, and if applicable details of automation tools.	Section 2.6 "Data Extraction", pages 7-8. Coding was performed by a single research team using a documented codebook; uncertain cases were resolved by consensus. A manual validation subsample and a full-text check on an accessible subset are reported in the same section.
	10a	List and define all outcomes for which data were sought.	Not applicable in the outcome-effect sense: this is a technology-mapping systematic review, not an intervention-outcome review. The six coding dimensions and six agentic-capability tags that substitute for outcomes are defined in Section 2.6 and Section 4.3.
	10b	List and define all other variables for which data were sought.	Section 2.6, page 7: bibliographic data, agricultural domain, organisational pattern, AI backbone, data sources, deployment setting, level of autonomy, evaluation strategy, and agentic capabilities, each with its coding rule.

Section	Item	Checklist item	Location in manuscript
	11	Specify the methods used to assess risk of bias in the included studies, including details of the tool(s) used, how many reviewers assessed each study, whether they worked independently, and if applicable details of automation tools.	Not applicable: no formal risk-of-bias instrument (e.g., ROBINS-I, Cochrane RoB2) was applied, since the included studies are heterogeneous technology and system-description papers rather than comparative intervention studies for which such tools are designed. Study-level evidentiary strength is instead captured through the evaluation-strategy coding in Section 2.6 and discussed critically in Section 7.5.
	12	Specify for each outcome the effect measure(s) used in the presentation or synthesis of results.	Not applicable: no pooled effect measure is computed. Results are synthesised as counts and percentages across coded dimensions, reported in Sections 5-7.
	13a	Describe the processes used to decide which studies were eligible for each synthesis.	Section 5 "Using the Taxonomy", page 24, and the domain/dimension definitions throughout Section 5; all 181 included studies are eligible for every descriptive synthesis reported.
	13b	Describe any methods required to prepare the data for presentation or synthesis.	Section 2.6, page 7, describes the coding codebook and priority rules used to prepare categorical data for tabulation.
	13c	Describe any methods used to tabulate or visually display results of individual studies and syntheses.	Tables 4-14 and Figures 3-14 throughout Sections 3, 5, 6 and 7; Table 15 (Appendix A) tabulates every individual included study.
	13d	Describe any methods used to synthesise results and provide a rationale for the choice(s).	Descriptive frequency synthesis (counts, percentages, cross-tabulations) is used throughout Sections 5-7, appropriate for a technology-mapping review without comparable numeric outcomes across studies; the choice is stated in Section 5.7.
	13e	Describe any methods used to explore possible causes of heterogeneity among study results.	Section 7.1 "Contemporary Systems Against the Classical Corpus" and Table 10 explore heterogeneity between the classical and language-model subsets directly; small-sample instability is discussed in Section 7.4.
	13f	Describe any sensitivity analyses conducted to assess robustness of the synthesised results.	The manual validation subsample and full-text check described in Section 2.6 function as a sensitivity check on the coding itself; Section 7.1 functions as a sensitivity comparison of findings across the two traditions.
	14	Describe any methods used to assess risk of bias due to missing results in a synthesis (arising from reporting biases).	Section 2.5 discusses the possible selection bias introduced by excluding studies with insufficiently informative abstracts; Section 10.2 discusses single-database coverage bias.
	15	Describe any methods used to assess certainty (or confidence) in the body of evidence for an outcome.	Not applicable in the GRADE sense used for intervention reviews. Confidence in the coded data is instead addressed through the validation subsample, the full-text check, and the explicit caveats on abstract-based coding in Section 2.6 and Section 10.2.
RESULTS	16a	Describe the results of the search and selection process, from the number of records identified to the number of studies included, ideally using a flow diagram.	Section 2.5 and Figure 2 (PRISMA 2020 flow diagram), page 7: 4111 identified, 181 included, with exact counts at every stage.
	16b	Cite studies that might appear to meet the inclusion criteria, but which were excluded, and explain why they were excluded.	Section 2.4 explains the two named exclusion rules (socio-economic agent-based modelling; non-agricultural word senses of "farm") with worked examples; Table 3 gives

Section	Item	Checklist item	Location in manuscript
			the general exclusion categories. Excluded records are not cited individually given the scale of exclusion (2,749-599 records at different stages).
	17	Cite each included study and present its characteristics.	Every included study is cited in-text in Section 6 or listed in Table 15 (Appendix A) and the supplementary coded spreadsheet, both giving domain, architecture, backbone, autonomy and evaluation for each study.
	18	Present assessments of risk of bias for each included study.	Not applicable; see item 11. Evaluation-strategy coding (Table 12) serves as the closest analogue.
	19	Present results for all outcomes.	Sections 5-8 present results for every coded dimension across all 181 studies.
	20a	Briefly summarise the characteristics and risk of bias among studies contributing to the synthesis.	Table 4 (characteristics), Section 3 (bibliometrics), and the limitations on coding reliability in Section 10.2.
	20b	Present results of all statistical syntheses conducted.	Tables 5-14 and Figures 3-14 present every synthesis conducted, with raw counts and percentages throughout.
	20c	Present results of any investigations of possible causes of heterogeneity among study results.	Section 7.1-7.3 and Table 10 investigate heterogeneity between the classical and language-model subsets.
	20d	Present results of any sensitivity analyses.	Section 2.6 (validation subsample and full-text check results) and Section 7.1 (classical vs. language-model comparison).
	21	Present assessments of risk of bias due to missing results.	Section 10.2 discusses selection bias from abstract-informativeness exclusion and single-database coverage.
	22	Present assessments of certainty in the body of evidence.	Not applicable in the GRADE sense; see item 15. Section 10.2 gives the closest equivalent, an explicit statement of how much confidence the coding and evidence base support.
DISCUSSION	23a	Provide a general interpretation of the results in the context of other evidence.	Section 10.1 "Interpretation in Context", page 33-34, read together with Section 1.5's comparison to four related reviews.
	23b	Discuss any limitations of the evidence included in the review.	Section 10.2 "Limitations of the Evidence and the Review Process", page 34, covers abstract-level reporting quality and coverage gaps in the underlying literature.
	23c	Discuss any limitations of the review processes used.	Section 10.2, same location: single database, no protocol pre-registration, single-team coding, taxonomy boundary judgement calls, affiliation-country simplification, partial 2026 coverage.
	23d	Discuss implications of the results for practice, policy, and future research.	Section 9 "Future Research Directions" (research implications) and Section 10.3 "Implications for Practice and Policy" (practice and policy implications).
OTHER INFORMATION	24a	Provide registration information for the review, including register name and registration number, or state that the review was not registered.	Section 2.1, page 5, and Section 10.2: the review was not registered on a public platform such as PROSPERO or the Open Science Framework, stated directly as a limitation.

Section	Item	Checklist item	Location in manuscript
	24b	Indicate where the review protocol can be accessed, or state that a protocol was not prepared.	A written protocol was followed internally (Section 2.1) but was not deposited on a public repository; it is available from the corresponding author on request, as stated in the Data Availability Statement.
	24c	Describe and explain any amendments to information provided at registration or in the protocol.	Not applicable: the review was not registered, so no registered protocol exists to amend. Section 2.1 states the inclusion rules were fixed before screening began and not changed afterward.
	25	Describe sources of financial or non-financial support for the review, and the role of the funders or sponsors in the review.	Funding statement, end of manuscript: this research received no external funding.
	26	Declare any competing interests of review authors.	Conflicts of Interest statement, end of manuscript: the authors declare no conflict of interest.
	27	Report which of the following are publicly available: template data collection forms; data extracted from included studies; data used for all analyses; analytic code; any other materials used in the review.	Data Availability Statement and Supplementary Materials statement, end of manuscript: the coded extraction spreadsheet for all 181 studies, the codebook, the complete Scopus query, and the validation-sample data are provided as supplementary files; Appendix A reproduces the core coding in the manuscript itself.