PRISMA 2020 Checklist

Section and Topic	Item #	Checklist item	Location where item is reported
TITLE			
Title	1	Identify the report as a systematic review.	Title (p.1): the article is identified as "A Systematic Review".
ABSTRACT			
Abstract	2	See the PRISMA 2020 for Abstracts checklist.	Abstract (p.1): objectives, information sources (three databases, 550 records), number of included studies (74), quality appraisal (modified JBI), risk of bias of the review (ROBIS), main results, limitations, and statement that the review was not registered.
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of existing knowledge.	Section 1, Introduction.
Objectives	4	Provide an explicit statement of the objective(s) or question(s) the review addresses.	Section 1, Introduction (RQ1–RQ3) and stated contributions.
METHODS			
Eligibility criteria	5	Specify the inclusion and exclusion criteria for the review and how studies were grouped for the syntheses.	Section 2.2, Study Inclusion Process (inclusion/exclusion criteria; grouping by AI paradigm in

Section and Topic	Item #	Checklist item	Location where item is reported
			Section 3).
Information sources	6	Specify all databases, registers, websites, organisations, reference lists and other sources searched or consulted to identify studies. Specify the date when each source was last searched or consulted.	Section 2.1, Research Strategy (IEEE Xplore, Scopus, ACM Digital Library; publications from January 2019 to December 2025) and Appendix B, Table A2. The exact date of the last search is not reported (acknowledged in Sections 2.1 and 5).
Search strategy	7	Present the full search strategies for all databases, registers and websites, including any filters and limits used.	Section 2.1 (Boolean search string and concept blocks) and Appendix B, Table A2 (search string, fields searched: title, abstract and keywords, filters and limits, and records per database; the same string, fields and filters were used in all three databases). Search limitations and their impact are discussed in Section 2.1.
Selection process	8	Specify the methods used to decide whether a study met the inclusion criteria of the review, including how many reviewers screened each record and each report retrieved, whether they worked independently, and if applicable, details of automation tools used in the process.	Section 2.2: two reviewers (H.M.,

Section and Topic	Item #	Checklist item	Location where item is reported
			M.E.G.) independently screened titles/abstracts of all 550 records and the full texts of retained records; disagreements resolved by discussion or by a third reviewer (M.A.E.K.); no automation tools. Application of the low-quality journal criterion described in Section 2.2.
Data collection process	9	Specify the methods used to collect data from reports, including how many reviewers collected data from each report, whether they worked independently, any processes for obtaining or confirming data from study investigators, and if applicable, details of automation tools used in the process.	Section 2.2: predefined extraction form; data extracted independently by two reviewers and cross-checked; discrepancies resolved as for screening; no automation tools; study investigators not contacted.
Data items	10a	List and define all outcomes for which data were sought. Specify whether all results that were compatible with each outcome domain in each study were sought (e.g. for all measures, time points, analyses), and if not, the methods used to decide which results to collect.	Section 2.3 (outcomes sought and formal

Section and Topic	Item #	Checklist item	Location where item is reported
			definitions: accuracy, precision, recall, F1-score, FPR, latency); Section 3 ("Reading Tables 3–9": accuracy tabulated, other metrics extracted when reported).
	10b	List and define all other variables for which data were sought (e.g. participant and intervention characteristics, funding sources). Describe any assumptions made about any missing or unclear information.	Section 2.2 / Section 3 (variables: AI technique category, dataset, attack type, publication year).
Study risk of bias assessment	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 and whether they worked independently, and if applicable, details of automation tools used in the process.	Sections 2.3–2.4 and Appendix B: modified JBI checklist (8 criteria; adaptation detailed in Table A3; scoring and thresholds), applied independently by two reviewers, not blinded; study-level risk-of-bias indicators (evaluation data, number of datasets, ceiling effects) in Section 2.4; ROBIS applied to the review itself (Table A4).
Effect measures	12	Specify for each outcome the effect measure(s) (e.g. risk ratio, mean difference) used in the synthesis or presentation of results.	Accuracy, precision, recall, F1-score, false

Section and Topic	Item #	Checklist item	Location where item is reported
			positive rate (FPR), and latency are reported as descriptive summary statistics across the included studies. No meta-analysis was performed.
Synthesis methods	13a	Describe the processes used to decide which studies were eligible for each synthesis (e.g. tabulating the study intervention characteristics and comparing against the planned groups for each synthesis (item #5)).	Section 3 and Section 4 (studies grouped by AI paradigm for narrative synthesis).
	13b	Describe any methods required to prepare the data for presentation or synthesis, such as handling of missing summary statistics, or data conversions.	Not applicable — no quantitative data preparation/meta-analysis; summary statistics reported descriptively in Section 4.
	13c	Describe any methods used to tabulate or visually display results of individual studies and syntheses.	Sections 3–4: Tables 3–9 (individual studies), Tables 10–11 (consolidated by category), Tables 15–18 (attack types, deployment indicators, RQ1 and RQ3 syntheses), Figures 5–7.
	13d	Describe any methods used to synthesize results and provide a rationale for the choice(s). If meta-analysis was performed, describe the model(s), method(s) to identify the presence and extent of statistical heterogeneity, and software package(s) used.	Narrative synthesis with descriptive statistics (Section

Section and Topic	Item #	Checklist item	Location where item is reported
			4.1); no meta-analysis because of heterogeneous datasets, tasks, and protocols. Synthesis strategies (within-study contrasts, consistency of patterns, descriptive distributions) described in Section 4.9.
	13e	Describe any methods used to explore possible causes of heterogeneity among study results (e.g. subgroup analysis, meta-regression).	Section 4.1 and Section 4.9 (within-study contrasts of task definition, dataset, and feature set as causes of heterogeneity; Figure 6 cross-tabulates categories and datasets).
	13f	Describe any sensitivity analyses conducted to assess robustness of the synthesized results.	Not applicable — no sensitivity analysis (no meta-analysis performed).
Reporting bias assessment	14	Describe any methods used to assess risk of bias due to missing results in a synthesis (arising from reporting biases).	Section 2.4 (publication and reporting bias): formal tests not feasible (no common effect measure or

Section and Topic	Item #	Checklist item	Location where item is reported
			variance); indicators and likely direction of bias discussed.
Certainty assessment	15	Describe any methods used to assess certainty (or confidence) in the body of evidence for an outcome.	No formal GRADE assessment (not designed for technical benchmark studies); confidence in findings addressed through JBI grades, study-level risk-of-bias indicators, and ROBIS (Sections 2.3–2.4).
RESULTS			
Study selection	16a	Describe the results of the search and selection process, from the number of records identified in the search to the number of studies included in the review, ideally using a flow diagram.	Section 2.2, Figure 1 (PRISMA flow diagram: 550 records identified, 550 screened, 472 excluded at screening, 78 full-text reports assessed, 4 excluded with reasons, 74 studies included) and Table 1.
	16b	Cite studies that might appear to meet the inclusion criteria, but which were excluded, and explain why they were excluded.	Section 2.2 and Table 1 (excluded records that required an explicit eligibility decision, with stage and reason; re-

Section and Topic	Item #	Checklist item	Location where item is reported
			assessment of four records initially excluded on performance or novelty).
Study characteristics	17	Cite each included study and present its characteristics.	Section 3, Tables 3–9 (characteristics of the 49 studies with extractable method, dataset and results; the remaining 25 included studies are not tabulated individually, Sections 3 and 5) and Appendix A.
Risk of bias in studies	18	Present assessments of risk of bias for each included study.	Section 2.3 (JBI scores: mean 6.2, SD 1.1 for the initial 71 studies; 4.5, 5 and 5.5 for the three studies included during revision; all studies ≥4/8) and Section 2.4 (study-level risk-of-bias indicators for the 49 tabulated studies). Per-study JBI scores of the initial 71 studies are not reported individually (Section 5).

Section and Topic	Item #	Checklist item	Location where item is reported
Results of individual studies	19	For all outcomes, present, for each study: (a) summary statistics for each group (where appropriate) and (b) an effect estimate and its precision (e.g. confidence/credible interval), ideally using structured tables or plots.	Not applicable — no meta-analysis / effect estimates; performance metrics reported descriptively in Section 3–4.
Results of syntheses	20a	For each synthesis, briefly summarise the characteristics and risk of bias among contributing studies.	Sections 3.3–3.9 (synthesis paragraph per category), Section 2.4 (risk-of-bias indicators), Tables 10 and 17.
	20b	Present results of all statistical syntheses conducted. If meta-analysis was done, present for each the summary estimate and its precision (e.g. confidence/credible interval) and measures of statistical heterogeneity. If comparing groups, describe the direction of the effect.	Not applicable — no statistical synthesis/meta-analysis performed.
	20c	Present results of all investigations of possible causes of heterogeneity among study results.	Section 4.9 (within-study contrasts: binary vs. multi-class tasks, datasets, feature sets) and Section 4.3 (generalization across environments).
	20d	Present results of all sensitivity analyses conducted to assess the robustness of the synthesized results.	Not applicable — no sensitivity analysis performed.
Reporting biases	21	Present assessments of risk of bias due to missing results (arising from reporting biases) for each synthesis assessed.	Section 2.4 (indicators of publication/reporting

Section and Topic	Item #	Checklist item	Location where item is reported
			bias and likely impact) and Section 5.
Certainty of evidence	22	Present assessments of certainty (or confidence) in the body of evidence for each outcome assessed.	No formal GRADE assessment; Section 4.9 states which conclusions are and are not supported by the evidence.
DISCUSSION			
Discussion	23a	Provide a general interpretation of the results in the context of other evidence.	Section 4, Discussion, including Sections 4.8 (detection vs. prevention), 4.9 (synthesis of non-comparable results), and 4.10 (answers to RQ1–RQ3).
	23b	Discuss any limitations of the evidence included in the review.	Section 5, Limitations (heterogeneity of protocols; reliance on legacy datasets).
	23c	Discuss any limitations of the review processes used.	Section 5, Limitations (three databases, search terms, English only, no registration/protocol, lenient quality threshold, AI-specific appraisal items, outcome-based exclusions, unblinded

Section and Topic	Item #	Checklist item	Location where item is reported
			appraisal) and ROBIS assessment (Section 2.4, Table A4).
	23d	Discuss implications of the results for practice, policy, and future research.	Sections 4.7–4.10 (reporting protocol, Table 14; prevention design recommendations) and Section 6.
OTHER INFORMATION			
Registration and protocol	24a	Provide registration information for the review, including register name and registration number, or state that the review was not registered.	Section 2: the review was not registered (PROSPERO not applicable to non-health reviews; no alternative registry used); implications discussed in Section 2 and Section 5.
	24b	Indicate where the review protocol can be accessed, or state that a protocol was not prepared.	Section 2: no formal protocol was prepared; mitigation measures described in Section 2.
	24c	Describe and explain any amendments to information provided at registration or in the protocol.	Not applicable — no registration/protocol amendments.
Support	25	Describe sources of financial or non-financial support for the review, and the role of the funders or sponsors in the review.	Funding Declaration: this study did not receive any funding.
Competing interests	26	Declare any competing interests of review authors.	Conflict of Interest Statement: the

Section and Topic	Item #	Checklist item	Location where item is reported
			authors declare no conflicts of interest.
Availability of data, code and other materials	27	Report which of the following are publicly available and where they can be found: 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: no new data were generated; all reviewed studies are cited in the References.

From: Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021;372:n71. doi: 10.1136/bmj.n71. This work is licensed under CC BY 4.0. To view a copy of this license, visit https://creativecommons.org/licenses/by/4.0/