# PRISMA 2020 Main Checklist

| **Topic** | **No.** | **Item** | **Location where item is reported** |
| --- | --- | --- | --- |
| **TITLE** |  |  |  |
| **Title** | 1 | Identify the report as a systematic review. | Title page — “A Systematic Review of Algorithms...” |
| **ABSTRACT** |  |  |  |
| **Abstract** | 2 | See the PRISMA 2020 for Abstracts checklist |  |
| **INTRODUCTION** |  |  |  |
| **Rationale** | 3 | Describe the rationale for the review in the context of existing knowledge. | Sections 1.1 and 1.2 — “Software Testing: Concept and Advantages” and “Systematic Literature Review on AI Algorithm in Software Testing” |
| **Objectives** | 4 | Provide an explicit statement of the objective(s) or question(s) the review addresses. | Section 1.2.2 — “Planning”: includes the research questions (RQ1–RQ3) and main objectives of the systematic review. |
| **METHODS** |  |  |  |
| **Eligibility criteria** | 5 | Specify the inclusion and exclusion criteria for the review and how studies were grouped for the syntheses. | Section 1.2.3 — “Execution”, coded filtering process (1–4) describing the inclusion/exclusion stages (title, abstract, introduction/conclusion, full text) and Section 1.2.2 — “Planning”:Table 2 |
| **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 1.2.3 “Execution”; and Section 1.2.2 — “Planning”: Table 1 (databases and search strings) |
| **Search strategy** | 7 | Present the full search strategies for all databases, registers and websites, including any filters and limits used. | Section 1.2.2 - “Planning”; Table 1 (search string per database) |
| **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 1.2.3 “Execution”, paragraph “Data Screening and Extraction Process”, and Figure 2 (PRISMA 2020 flow diagram). |
| **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 1.2.3 “Execution”, paragraph “Data Screening and Extraction Process” |
| **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 1.2.4 “Results” — For each included study, data were extracted on the following outcomes: (i) AI algorithms applied to software testing, (ii) testing problem categories (e.g., defect prediction, test automation, effort estimation), (iii) input variables used by the algorithms (structural, historical, semantic, dynamic, interface-related), and (iv) evaluation metrics employed (accuracy, F-measure, precision, recall, AUC, etc.). These outcomes were selected to capture the methodological, technical, and evaluative dimensions of AI applications in software testing between 2014–2024. |
|  | 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 1.2.2 “Planning” and Section 1.2.3 “Execution” — Additional contextual variables were extracted, including:  • Year of publication (2014–2024)  • Source database (Scopus or Web of Science)  • Dataset used (PROMISE, NASA, AEEEM, ReLink, JIT, or proprietary)  • Learning paradigm (supervised, unsupervised, semi-supervised, or hybrid)  • Study type (comparative, proposal, or review). When data were incomplete or ambiguous, cross-references were verified in related studies or author-provided supplementary material. No assumptions were made beyond the information available in the original articles. |
| **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. | N/A, Not applicable. This systematic review focuses on software engineering studies and algorithmic analyses, where traditional bias assessment frameworks (e.g., Cochrane tools) are not relevant. Instead, methodological rigor and reporting completeness were evaluated qualitatively during data extraction. |
| **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. | N/A, Not applicable. This study does not perform statistical comparisons or meta-analytical effect estimation. Instead, performance indicators (e.g., accuracy, precision, recall, F1-score, AUC, and coverage metrics) were descriptively analyzed to assess algorithmic effectiveness in software testing contexts. |
| **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 1.2.3 “Execution” — details the inclusion and exclusion process applied in four sequential stages (title screening, abstract and keywords, introduction/conclusion, and full-text review). A coded filtering matrix (1–4) was used to determine eligibility for synthesis, ensuring that only studies addressing AI algorithms in software testing between 2014 and 2024 were retained. |
|  | 13b | Describe any methods required to prepare the data for presentation or synthesis, such as handling of missing summary statistics, or data conversions. | Section 1.2.4 “Results” — describes how extracted data from eligible studies were organized into analytical dimensions: (i) testing problems, (ii) algorithmic approaches, (iii) input variables, and (iv) evaluation metrics. Each dimension was categorized based on thematic analysis and recorded in a structured Excel matrix to facilitate synthesis and visualization. |
| 13c | Describe any methods used to tabulate or visually display results of individual studies and syntheses. | Section 1.2.4 “Results”, and Figures 2–5 — summarize results using tabular and visual formats, including a PRISMA flow diagram (Figure 2) for study selection and taxonomic tables showing algorithmic categories, input variable types, and evaluation metrics. Graphical summaries (Figures 3–5) illustrate trends and distributions over time. |
| 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. | Section 1.3.3, synthesis was conducted qualitatively through thematic grouping and comparative taxonomy of AI algorithms according to problem categories, input variables, and evaluation metrics. This approach was chosen to identify patterns and evolutionary trajectories in AI applications for software testing, rather than quantitative aggregation. Python was used to generate the graphics |
| 13e | Describe any methods used to explore possible causes of heterogeneity among study results (e.g. subgroup analysis, meta-regression). | Section 2 “Discussion” — analyzes heterogeneity among studies, algorithmic paradigms (machine learning, evolutionary computation, deep learning), and evaluation criteria. Differences were interpreted as reflective of the research evolution within the decade. |
| 13f | Describe any sensitivity analyses conducted to assess robustness of the synthesized results. | N/A, Not applicable. Since this review follows a qualitative synthesis design, no statistical sensitivity analysis was performed. Robustness was ensured through double-review validation during inclusion/exclusion and consistency checks across extracted variables. |
| **Reporting bias assessment** | 14 | Describe any methods used to assess risk of bias due to missing results in a synthesis (arising from reporting biases). | N/A, Not applicable. The review does not include a formal reporting bias assessment, as it focuses on descriptive synthesis of studies in software engineering. |
| **Certainty assessment** | 15 | Describe any methods used to assess certainty (or confidence) in the body of evidence for an outcome. | N/A, Not applicable. This systematic review synthesizes heterogeneous studies in software testing without formal grading of evidence certainty. Nevertheless, consistency and validity were ensured by applying clear inclusion criteria, independent data extraction, and cross-verification among reviewers |
| **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 1.2.3 ‘Execution’ Fig. 2 (PRISMA Diagram), and 1.2.4 Result,Table 3. |
|  | 16b | Cite studies that might appear to meet the inclusion criteria, but which were excluded, and explain why they were excluded. | Section 1.2.2 Planning, Table 2 |
| **Study characteristics** | 17 | Cite each included study and present its characteristics. | Section 1.2.4 “Results” and Tables 5-8, (Characteristics of included studies) |
| **Risk of bias in studies** | 18 | Present assessments of risk of bias for each included study. | N/A, Not applicable. The review focuses on algorithmic and methodological characteristics rather than study bias quantification. |
| **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. | ection 1.2.4 ‘Results’, Tables 4-8, 9 |
| **Results of syntheses** | 20a | For each synthesis, briefly summarise the characteristics and risk of bias among contributing studies. | Section 1.2.4 — Results: summarizes the characteristics of the 84 selected studies, including publication years, algorithmic families, testing problem categories, input variable types, and evaluation metrics. Studies were classified according to AI approach and application scope, minimizing selection bias through independent double screening. |
|  | 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. | N/A, Not applicable. This study did not perform a statistical synthesis or meta-analysis. Results are presented qualitatively through categorical analysis of algorithms, problems, and metrics. |
| 20c | Present results of all investigations of possible causes of heterogeneity among study results. | Section 2, discusses sources of heterogeneity across studies, algorithmic paradigms, variables and evaluation metrics. These variations were interpreted as natural heterogeneity reflecting the evolution of research objectives in AI-based software testing. |
| 20d | Present results of all sensitivity analyses conducted to assess the robustness of the synthesized results. | N/A, Not applicable. No sensitivity analysis was conducted since the review followed a qualitative synthesis framework rather than a quantitative meta-analysis. |
| **Reporting biases** | 21 | Present assessments of risk of bias due to missing results (arising from reporting biases) for each synthesis assessed. | N/A, Not applicable. No formal reporting bias assessment was performed. Nonetheless, the potential impact of publication and dataset availability bias was discussed in Section 2.5 (Future Research Directions) |
| **Certainty of evidence** | 22 | Present assessments of certainty (or confidence) in the body of evidence for each outcome assessed. | N/A, Not applicable. This review did not perform a quantitative synthesis or statistical confidence assessment. However, the reliability of findings was strengthened by the inclusion of peer-reviewed studies from Scopus and Web of Science, following transparent selection and coding criteria as outlined in Section 1.2.3 (Execution). |
| **DISCUSSION** |  |  |  |
| **Discussion** | 23a | Provide a general interpretation of the results in the context of other evidence. | Section 2 - Discussion, Section 2.1, 2.2, 2.3, 2.4, 2.5; provides a general interpretation of the results, highlighting how the taxonomy and evolution of AI algorithms in software testing align with or differ from previous literature. The discussion integrates findings with prior studies on defect prediction, test automation, and explainable AI, emphasizing the progression from predictive to autonomous testing approaches. |
|  | 23b | Discuss any limitations of the evidence included in the review. | Section 2 - Discussion, discusses limitations of the reviewed evidence, including potential publication bias toward positive results, variations in dataset quality and availability, and the uneven distribution of algorithmic approaches across problem categories. |
| 23c | Discuss any limitations of the review processes used. | Section 2 - Discussion, acknowledges limitations of the review process itself, such as the restriction to English-language studies and reliance on Scopus and Web of Science databases. It also notes that the inclusion/exclusion decisions were based on manual screening, which may introduce subjective interpretation despite double-review validation. |
| 23d | Discuss implications of the results for practice, policy, and future research. | Section 2 - Discussion, Section 2.5, presents implications for software testing research and industry, suggesting that future studies should integrate explainable AI (XAI) mechanisms, hybrid metrics, and real-world validation pipelines. The section also outlines directions for sustainable and interpretable AI adoption in industrial testing contexts. |
| **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. | N/A, The review was not registered in any public registry, as PRISMA registration databases (e.g., PROSPERO) are primarily oriented to health-related research. |
|  | 24b | Indicate where the review protocol can be accessed, or state that a protocol was not prepared. | N/A, No external protocol was published. The complete methodological protocol, including search strategy, selection criteria, and data extraction procedures, is fully described in Sections 1.2.1–1.2.3 of the manuscript. |
| 24c | Describe and explain any amendments to information provided at registration or in the protocol. | N/A, No amendments were made, as the review followed a predefined methodological framework described in the manuscript. |
| **Support** | 25 | Describe sources of financial or non-financial support for the review, and the role of the funders or sponsors in the review. | Section “Funding” |
| **Competing interests** | 26 | Declare any competing interests of review authors. | Section “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. | Section “Data Availability Statement” |

*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. MetaArXiv. 2020, September 14. DOI: 10.31222/osf.io/v7gm2. For more information, visit: <www.prisma-statement.org>