IMPERIAL

Protocol for the Systematic Literature Review on Electoral Integrity Strategies (ELIS 2025)

Author:
Carlos Rocha
Visiting Researcher
Imperial College Business School

Supervisor:
Tommaso Valletti
Professor of Economics
Department of Economics and Public Policy - Business School

Date:
November 2025

Document Type:
Protocol for the Systematic Literature Review
Adapted from PRISMA-P 2015 Guidelines

IMPERIAL

Protocol for the Systematic Literature Review on Electoral Integrity Strategies (ELIS 2025)

Adapted from PRISMA-P 2015 Guidelines
Version: 1.8 | **Date**: 19 November 2025
Principal Investigator: Carlos Rocha (Imperial College Business School)

1. Administrative Information

1.1 Title

Systematic Review Protocol: Electoral Integrity Strategies (ELIS 2025)

This document describes the initial protocol; substantial amendments will be versioned (v1.x, v2.x) and documented in the Amendments section.

1.2 Registration

Primary registry

This protocol will be registered on Spiral (Imperial College London) and mirrored on OSF prior to full-text screening. The OSF registration ID and URL will be added to this section once issued. Given the subject area, PROSPERO registration is not applicable.

Institutional repository

A copy of the registered protocol and subsequent amendments will also be deposited in Spiral, the institutional repository of Imperial College London, for long-term preservation and institutional record.

1.3 Authors and Contributions

- **Carlos Rocha** – Protocol design, AI-assisted methodology, review strategy, lead author.
  Visiting Researcher
  Department of Economics and Public Policy
  Imperial College Business School
  e-mail: c.rocha@imperial.ac.uk
  https://orcid.org/0009-0009-6741-2193
  Guarantor of the review: Carlos Rocha (responsible for the integrity of the protocol, conduct of the review, and final reporting).
- **Tommaso Valletti** – Conceptual advisor, research supervisor.
  Professor of Economics
  Department of Economics and Public Policy
  Imperial College Business School
  e-mail: t.valletti@imperial.ac.uk

1.4 Amendments

Amendments will follow semantic versioning (v1.7, v1.8, etc.) and be documented in:

- A public CHANGELOG.md file in the ELIS SLR Agent GitHub repository (detailing date, version, and nature of changes); and
- The OSF registration's "Version history" and "Registered report" sections. Major methodological amendments (e.g., changes to eligibility criteria or core outcomes) will be explicitly flagged in the final review report.

Version 1.8 (19 November 2025)

Updated Section 3.2 to reflect finalized information sources based on API availability assessment. Alternative sources (Semantic Scholar, OpenAlex, CrossRef, CORE) replace originally specified sources (ACM, JSTOR, Google Scholar) that lacked suitable APIs for systematic retrieval. This change maintains disciplinary coverage while enhancing reproducibility. Full rationale documented in project repository.

1.5 Support and Sponsor

Institutional: Imperial College Business School (Visiting Researcher programme), providing intellectual environment and access to library resources.

Sponsor: Instituto Voto Legal (IVL), Sao Paulo, Brazil
Role of funder/sponsor: The sponsor has no role in protocol design, data collection, data analysis, decision to publish, or preparation of the manuscript. No financial incentives are linked to the results.

2. Introduction

2.1 Rationale

This review investigates the technological, operational, and institutional dimensions that influence electoral integrity. It seeks to generate robust academic evidence on the strategies, mechanisms, and design features that strengthen auditability, publicity, and public trust, across diverse electoral system models. Particular emphasis is placed on interdisciplinary approaches to evaluating voting systems, including electronic and paper-based modalities, in light of rapid technological evolution and renewed concerns over democratic resilience. The synthesis aims to inform academic research and policy design toward effective and independent elections auditing.

To date, there is no consolidated systematic review that jointly examines technological, operational, and institutional strategies for electoral integrity across both electronic and paper-based systems. Existing reviews tend to focus on single technologies (e.g., DRE security, risk-limiting audits) or broader indices of "electoral integrity", without systematically linking concrete design features to empirically observed outcomes.

2.2 Objectives

Primary Research Question (PRQ)

What operational and technological strategies have been shown to improve the integrity or auditability of electoral systems since 1990?

Methodological Sub-question (MSQ)

What types of empirical designs and evaluation frameworks have been used to assess the effectiveness of electoral integrity strategies since 1990?

The phrase *"have been shown to improve"* in the Primary Research Question is used to reflect a range of empirical evidence types. It includes findings from experimental and quasi-experimental studies, comparative observational research, technical evaluations, and structured qualitative analyses that support a causal interpretation. Where causal inference is limited or contested, the review will clearly distinguish between robust findings and those that are suggestive or correlational in nature.

Analytical Sub-questions

a) **Systems & Mechanisms:** What specific technological or operational mechanisms have been associated with increased auditability or verifiability in voting systems?

b) **Institutional Conditions:** Under what institutional, legal, or regulatory conditions have these mechanisms been implemented?

c) **Trust & Perception:** How have these strategies influenced public trust, voter confidence, or perceptions of electoral integrity?

d) **Regional Variation or Global Trends:** What regional patterns or cross-national differences are observed in the adoption and evaluation of these strategies?

2.3 Conceptual Framework – SPIDER

To accommodate both quantitative and qualitative evidence relevant to electoral integrity and auditability, the ELIS review adopts the **SPIDER framework** (Sample, Phenomenon of Interest, Design, Evaluation, Research type). This is appropriate given the diversity of empirical designs in the field. Table 1 defines how SPIDER applies to the ELIS review.

Table 1 – SPIDER Framework for ELIS

Component	Definition in ELIS Review
S – Sample	Public electoral processes and voting systems used in national, subnational, or referendum elections, including official pilots and large-scale trials where outcomes are used to allocate public office or test electoral technologies under realistic conditions.

PI – Phenomenon of Interest	Operational and technological strategies and design features intended to improve the **integrity** and **auditability** of elections. Examples include: voter-verified paper audit trails (VVPAT), risk-limiting audits, parallel vote tabulation, end-to-end verifiability, and public reporting mechanisms.
D – Design	Empirical study designs, including: randomised field experiments, quasi-experimental studies (e.g., RDD, DiD), observational comparative analyses, technical validation studies, structured case studies, and election observation reports using explicit evaluation criteria.
E – Evaluation	Documented empirical outcomes indicating changes in integrity, auditability, or public trust—such as discrepancy detection, audit results, fraud/error identification, voter confidence surveys, or observed implementation effects.
R – Research Type	Quantitative, qualitative, and mixed-methods empirical studies that present systematically collected data or structured evaluations relevant to electoral integrity and auditability.

This framework supports the inclusive but rigorous approach outlined in Section 3.1 (Eligibility Criteria) and reflects the interdisciplinary scope of the review.

3. Methods

3.1 Eligibility Criteria

Inclusion Criteria

- Peer-reviewed academic articles (1990–2025).
- English, French, Spanish or Portuguese.
- Empirical studies (quantitative, qualitative or mixed-methods).
- Focus on voting system integrity, auditability, transparency or public trust.

Exclusion Criteria:

- Opinion pieces or editorials without empirical basis.
- Theoretical or normative texts lacking methodological rigour.
- Articles focused solely on party politics or turnout without relation to voting system design.
- Studies lacking verifiable authorship, institutional affiliation, or publication venue.

3.2 Information Sources

The review will include sources from the following databases and APIs:

Primary Sources:

1. **Scopus** – Multidisciplinary database with comprehensive coverage across political science, governance, law, and engineering.
2. **Web of Science** – High-impact journal indexing platform enabling detailed citation analysis across disciplines.
3. **IEEE Xplore** – Technical literature repository covering electronic voting systems, cryptographic security, and system auditability.
4. **Semantic Scholar** – AI-enhanced bibliographic database covering 200M+ papers across computer science and interdisciplinary research, with citation graphs and semantic indexing.
5. **OpenAlex** – Open bibliographic database (250M+ works) providing comprehensive metadata including institutions, citations, and concept tagging.
6. **CrossRef** – DOI registration agency providing publisher-verified metadata for 130M+ records, enabling robust deduplication and citation tracking.
7. **CORE** – Open access aggregator covering 300M+ papers, theses, and preprints from institutional repositories worldwide.

Source Selection Rationale:

These sources were selected to provide:

- **Disciplinary breadth:** Coverage across political science, computer science, law, and governance
- **Methodological diversity:** Inclusion of both empirical studies and technical evaluations
- **API accessibility:** All sources provide documented APIs enabling reproducible automated searches
- **Complementary coverage:** Combination of subscription databases (Scopus, WoS, IEEE) and open sources (Semantic Scholar, OpenAlex, CORE) maximizes retrieval while supporting open science principles

Technical Implementation:

All search results will be imported into Zotero for de-duplication and metadata management. Initial screening and tagging will be automated using the ELIS SLR Agent, an open-source Python workflow hosted on GitHub. Human reviewers will verify all inclusion decisions, with all decisions and version history recorded via GitHub for full transparency and reproducibility.

3.3 Search Strategy

Boolean queries will target combinations of key terms related to voting technology, auditability, and public trust. Draft queries may be suggested by AI tools but are manually reviewed before application.

Example query:

```
("electoral integrity" OR "e-voting security" OR "ballot auditability" OR "electronic voting trust")
AND
```

("VVPAT" OR "end-to-end verifiability" OR "blockchain voting" OR "cryptographic audit" OR "risk-limiting audit")

AND

("evaluation" OR "empirical study" OR "comparative analysis")

Search logs and filtering decisions will be documented for reproducibility. Search queries are initially generated by LLMs and refined through expert validation.

3.4 Study Records

Screening and Selection Process

This review is conducted by a single researcher. To strengthen reproducibility and reduce bias, the **ELIS SLR Agent**, an open-source Python-based workflow, will perform first-pass screening using natural language filters and structured eligibility rules. All records will then be reviewed manually by the researcher. Disagreements between automated tags and final decisions will be logged to ensure transparency.

Titles and abstracts will be screened based on inclusion criteria. Studies marked “uncertain” or borderline will proceed to full-text screening. No study will be excluded solely by automation. All final decisions and notes will be recorded and version-controlled using GitHub.

Data Management

All references will be imported into **Zotero**, which will be used for metadata management, tagging, and de-duplication. Screening metadata and inclusion/exclusion labels will be exported and synced between Zotero and the ELIS SLR Agent pipeline. GitHub will serve as the versioned record for the screening process and data extraction files.

Data Extraction and Documentation

Structured data (e.g., study design, intervention type, outcomes) will be extracted directly into machine-readable files managed through the ELIS SLR Agent. All summaries, codes, and outputs will be manually verified. AI tools may be used to assist in synthesis, but no final data will be accepted without human validation.

3.5 Outcomes and Prioritisation

This review focuses on identifying empirical evidence linking operational and technological strategies to improvements in the **integrity** and **auditability** of electoral systems.

Primary outcomes will be prioritised in synthesis and interpretation. Secondary outcomes will be reported when directly linked to specific system design features or operational strategies. Studies reporting only attitudinal or correlational findings without system-level design relevance will be included only if clearly within scope.

Primary Outcomes include:

- Successful implementation of audits or verifiability mechanisms
- Detection and reporting of anomalies, errors, or tampering
- Evidence of procedural transparency or compliance improvements
- Reduction in error rates through system design interventions

Secondary Outcomes include:

- Changes in public trust or voter confidence
- Voter usability and satisfaction with systems
- Institutional capacity or oversight effectiveness

3.6 Risk of Bias and Quality Assessment

Risk of bias will be assessed using a structured checklist adapted from PRISMA-P, Cochrane ROB-2, and Norris et al. (2014). Key domains include: study design suitability, transparency of reporting, outcome measurement, and data availability.

The ELIS SLR Agent will assist in generating preliminary bias flags and summarizing risk signals using rule-based and LLM-aided methods. These tools help pre-classify studies as likely low, moderate, or high risk of bias based on reproducible criteria.

The lead reviewer will validate and finalize bias ratings, focusing attention where automated tools indicate uncertainty or elevated risk. Ratings and rationales will be version-controlled on GitHub and summarized in Annex E for transparency.

3.7 Data Synthesis

Data synthesis will be conducted in parallel for quantitative and qualitative studies using structured, reproducible procedures.

- **Quantitative studies** will be summarized using descriptive statistics, frequency distributions, and tabulated outcome types. Where appropriate, trends in intervention effectiveness or outcome domains will be aggregated across comparable studies.
- **Qualitative studies** will be analyzed using thematic synthesis, clustering findings around core conceptual dimensions (e.g., transparency, auditability, public trust). Themes will be derived inductively and refined iteratively.

AI tools (e.g., ChatGPT, Claude.ai, NotebookLM) may assist in suggesting thematic clusters, summarizing data fields, or identifying co-occurring variables. However, all synthesis steps will be manually verified by the reviewer, and no AI-generated interpretation will be used without explicit validation.

3.8 Confidence in Evidence

For qualitative and mixed-methods findings, this review will apply a **modified GRADE-CERQual*** **approach** to assess confidence in each synthesized theme or outcome. Confidence ratings will follow four levels: **High**, **Moderate**, **Low**, or **Very Low**.

The following four domains will guide each assessment:

- **Methodological Limitations**: Any concerns about study design or risk of bias.
- **Coherence**: The clarity and consistency of patterns across studies.
- **Data Adequacy**: The richness and quantity of data supporting each finding.
- **Relevance**: Applicability of the evidence to the ELIS research question.

All ratings will be applied at the **finding/theme level**, not the study level. Preliminary scoring may be assisted by structured heuristics from the ELIS SLR Agent, but final judgments are made and documented by the lead reviewer. Summary tables and rationales will be provided in **Annex E** and versioned in GitHub.

(*) GRADE-CERQual is a method used in qualitative evidence synthesis to assess the confidence one can have in the findings of a qualitative review.

4. Ethical Considerations

This review does not involve human subjects research or any personal identifiable data. Therefore, it does not require formal ethical approval from an institutional review board. Nonetheless, the review adheres to general principles of research integrity, especially given the use of automation and AI tools in the methodology:

- **Human Oversight:** All final decisions in the review (study inclusion/exclusion, data interpretations, conclusions) are made by the human investigator, not by AI. At no point are AI recommendations accepted without review.
- **Transparency:** Every use of an AI or automation tool in the process is logged and documented. For example, if an LLM is used to summarize an article, that summary and the prompt used are saved to the project log (see Annex F on automation workflow). This ensures an audit trail of how AI contributed.
- **Accountability:** The project follows the UK Research Integrity Office (UKRIO) Principles for the use of AI in research, ensuring responsible integration of AI and reproducibility of results. All outputs of the ELIS SLR Agent and other AI tools are subject to verification, and the human lead researcher accepts responsibility for the integrity of the review's findings.
- **Data Handling:** All reference data and results will be handled in accordance with open science best practices. Since no sensitive personal data is involved, the main ethical consideration is proper citation and avoidance of plagiarism in data synthesis, which will be diligently observed.

In summary, the protocol incorporates AI to improve efficiency but maintains rigorous human oversight and ethical research standards throughout.

5. Collaborative Review Process

This systematic review is initially being conducted by a single reviewer (the lead author). No additional independent reviewers are assigned at the protocol stage. However, the workflow has been designed to simulate certain benefits of a multi-reviewer process and to remain open for future collaboration:

- The **ELIS SLR Agent** automation tool acts as a second-pass reviewer in some respects, offering an independent (albeit algorithmic) perspective on screening and data extraction. While not a substitute for a human second reviewer, it provides a consistency check and helps reduce individual oversight errors.

- All major decisions and intermediate products (like screened-out lists, extracted data, bias assessments) are being version-controlled on GitHub. This means that if another researcher were to audit or join the project later, they could trace every decision made to date. The transparency via the GitHub repository serves as a form of "virtual collaboration," enabling external input or replication.

- If at any point additional human collaborators join the review, their roles and contributions will be documented both in an updated protocol amendment and in the GitHub commit history. The protocol and OSF registration would be updated to reflect multiple reviewers, and any necessary changes (such as dual independent screening) would be implemented.

For now, the combination of one human reviewer plus the AI-assisted workflow is deemed sufficient given resource constraints. We have put measures in place to maximize transparency and mitigate bias that could arise from single-reviewer screening (as detailed in Sections 3.4 and 3.6).

Should peer reviewers or external experts provide feedback during the conduct of the review, such feedback will be treated akin to collaborative input: major suggestions would be incorporated via protocol amendments as needed.

6. Annexes

- **Annex A** – PRISMA-P 2015 Checklist
- **Annex B** – Search Strings
- **Annex C** – Inclusion/Exclusion Log Template
- **Annex D** – Data Extraction Form
- **Annex E** – Risk of Bias Assessment Tool
- **Annex F** – Automation Workflow & AI
- **Annex G** – Evidence Gap Confirmation and Supporting Literature *(in progress)*
- **Annex H** – Tools and Workflow Diagram)

AI Disclaimer

All AI-assisted stages are supervised by human researchers. All automation steps are logged for transparency and auditability using the AI Agents Log Templates template in Annex F.

Annex A – PRISMA-P 2015 Checklist

The table below demonstrates how this protocol complies with each item of the PRISMA-P (2015) checklist for systematic review protocols. Each checklist item is addressed in the protocol, with section references provided:

Section / topic		Item #	Checklist item
ADMINISTRATIVE INFORMATION			
Title		1	
	Identification	1a	Identify the report as a protocol of a systematic review
	Update	1b	If the protocol is for an update of a previous systematic review, identify as such
Registration		2	If registered, provide the name of the registry (e.g., PROSPERO) and registration number
Authors		3	
	Contact	3a	Provide name, institutional affiliation, and e-mail address of all protocol authors; provide physical mailing address of corresponding author
	Contributions	3b	Describe contributions of protocol authors and identify the guarantor of the review
Amendments		4	If the protocol represents an amendment of a previously completed or published protocol, identify as such and list changes; otherwise, state plan for documenting important protocol amendments
Support		5	
	Sources	5a	Indicate sources of financial or other support for the review
	Sponsor	5b	Provide name for the review funder and/or sponsor
	Role of sponsor/funder	5c	Describe roles of funder(s), sponsor(s), and/or institution(s), if any, in developing the protocol
INTRODUCTION			
Rationale		6	Describe the rationale for the review in the context of what is already known
Objectives		7	Provide an explicit statement of the question(s) the review will address with reference to participants, interventions, comparators, and outcomes (PICO)
METHODS			
Eligibility criteria		8	Specify the study characteristics (e.g., PICO, study design, setting, time frame) and report characteristics (e.g., years considered, language, publication status) to be used as criteria for eligibility for the review
Information		9	Describe all intended information sources (e.g., electronic

sources			databases, contact with study authors, trial registers, or other grey literature sources) with planned dates of coverage
Search strategy		10	Present draft of search strategy to be used for at least one electronic database, including planned limits, such that it could be repeated
Study records		11	
	Data management	11a	Describe the mechanism(s) that will be used to manage records and data throughout the review
	Selection process	11b	State the process that will be used for selecting studies (e.g., two independent reviewers) through each phase of the review (e.g., screening, eligibility, and inclusion in meta-analysis)
	Data collection process	11c	Describe planned method of extracting data from reports (e.g., piloting forms, done independently, in duplicate), any processes for obtaining and confirming data from investigators
Data items		12	List and define all variables for which data will be sought (e.g., PICO items, funding sources), any pre-planned data assumptions and simplifications
Outcomes and prioritization		13	List and define all outcomes for which data will be sought, including prioritization of main and additional outcomes, with rationale
Risk of bias in individual studies		14	Describe anticipated methods for assessing risk of bias of individual studies, including whether this will be done at the outcome or study level, or both; state how this information will be used in data synthesis
Data		15	
	Synthesis	15a	Describe criteria under which study data will be quantitatively synthesized
		15b	If data are appropriate for quantitative synthesis, describe planned summary measures, methods of handling data, and methods of combining data from studies, including any planned exploration of consistency (e.g., I^2, Kendall's tau)
		15c	Describe any proposed additional analyses (e.g., sensitivity or subgroup analyses, meta-regression)
		15d	If quantitative synthesis is not appropriate, describe the type of summary planned
Meta-bias(es)		16	Specify any planned assessment of meta-bias(es) (e.g., publication bias across studies, selective reporting within studies)
Confidence in cumulative evidence		17	Describe how the strength of the body of evidence will be assessed (e.g., GRADE)

Overall, this protocol adheres to all PRISMA-P checklist items, ensuring comprehensive reporting of the planned methodology.

Annex B – Search Strings

This annex provides the finalized search strategies for each information source listed in Section 3.2. Each search string has been tailored to the syntax of the respective database, but all share the same conceptual structure (electoral integrity terms AND specific intervention terms AND empirical study terms).

Scopus (Title/Abstract/Keyword search):

```
TITLE-ABS-KEY(("electoral integrity" OR "election integrity" OR "electronic voting
security" OR "ballot auditability" OR "voter trust")
  AND ("VVPAT" OR "voter-verified paper audit trail" OR "risk-limiting audit" OR
"parallel vote tabulation" OR "blockchain voting" OR "end-to-end verifiability" OR
"cryptographic audit")
  AND ("empirical study" OR evaluation OR "comparative analysis" OR "case study"))
```

Rationale: This query targets documents that mention general concepts of election integrity/trust, specific mechanisms (paper trails, audits, etc.), and indications of empirical research. The use of TITLE-ABS-KEY in Scopus ensures these terms appear in the title, abstract, or keywords.

Web of Science (Topic search):

```
TS=("electoral integrity" OR "election system integrity" OR "voting system security" OR
"auditability")
   AND TS=("risk-limiting audit" OR "VVPAT" OR "parallel vote tabulation" OR "biometric
voting" OR "verifiability")
   AND TS=("empirical study" OR "evaluation" OR "experimental" OR "comparative study")
```

Note: Web of Science TS field searches title, abstract, and keywords. The query includes variations like *experimental* to capture studies even if they do not self-describe as empirical.

IEEE Xplore: The search was adjusted to focus on technical papers:

```
("electronic voting" OR "e-voting" OR "voting system")
  AND ("security" OR "integrity" OR "auditability")
  AND ("verifiability" OR "blockchain" OR "VVPAT" OR "biometric")
  AND ("evaluation" OR "case study" OR "experiment")
```

Note: IEEE Xplore's interface was used to apply year filters (1990-2025) and to search within metadata and full text. Technical keywords like *blockchain* and *biometric* were included given their relevance in recent voting technology discourse.

ACM Digital Library: Similar to IEEE, but with more HCI terms:

```
("electronic voting" OR "voting interface" OR "digital ballot")
  AND ("integrity" OR "security" OR "transparency")
  AND ("usability study" OR "user study" OR "case study" OR "experiment")
```

Note: This captures studies discussing usability or interface aspects that might not appear in purely technical searches, ensuring we gather literature on how system design affects integrity via user interaction (important for ballot design, etc.).

JSTOR: A broad search was executed due to JSTOR's often full-text search:

```
"electoral integrity"
  AND (audit OR "fraud prevention" OR oversight)
  AND (study OR analysis)
```

Google Scholar: Multiple queries were run; a representative query:

```
"electoral integrity" "systematic review" OR
("electoral integrity" "empirical" "audit" 1990)
```

Google Scholar was used in an iterative fashion: initial broad queries identified a set of candidate grey literature and citations within those were followed. The flexible search allowed inclusion of reports, theses, and pre-2025 publications that might not appear in other databases. Results beyond the first 200 were not systematically reviewed, but no new themes emerged in later pages.

Search dates: All database searches were conducted between Nov 6–10, 2025. Each database's search results were exported on the day of search and the counts recorded. The search strategies may be updated if needed during the review (any changes will be noted in an amendment and logged). Full search logs, including exact query strings and any filters used, are available in the project repository for reproducibility.

Annex C – Inclusion/Exclusion Log Template

This annex describes how study selection decisions will be recorded, to ensure transparency at the full-text screening stage. For each study considered at full-text, the following information will be logged:

- **Study ID:** A unique identifier for each reference (e.g., an abbreviation or number, such as *ELIS2025-001*).
- **Citation Details:** Key reference information (author, year, title, source) for clarity.
- **Decision:** Inclusion or Exclusion status after full-text review.
- **Exclusion Reason (if excluded):** A brief reason chosen from a predefined list (e.g., *Population not in scope*, *No empirical data*, *Duplicate study*, *Outcome not relevant*, etc.). For example, *"Excluded – Not empirical (commentary piece)"*.
- **Notes:** Any additional notes or comments by the reviewer (e.g., *"Contains some relevant background, but no primary data"* or *"Pending second opinion on methodology"* if a consultation is needed).

All studies excluded after full-text screening will have an associated reason. The inclusion/exclusion log will be maintained in a spreadsheet or JSON format (managed via the ELIS SLR Agent pipeline) for consistency. An example entry is shown below:

Study ID	Reference (Author, Year)	Decision	Reason for Exclusion (if any)	Notes
ELIS2025-015	Doe *et al.*, 2018 – *Election Audits in X*	Exclude	No relevant intervention	Focuses on voter turnout only
ELIS2025-027	Smith, 2020 – *Electronic Voting Security*	Include	–	Proceed to data extraction

Table: Sample rows from the inclusion/exclusion log.

This log will be updated in real time as screening progresses. By the end of the selection process, it will provide a complete account of how many studies were included and excluded, and for what reasons. The finalized log (with all excluded references and reasons) will be appended to the eventual review report and made available online (e.g., as a supplemental file or in the GitHub repository).

Maintaining this log aligns with PRISMA reporting guidelines and facilitates the construction of the PRISMA flow diagram for study selection.

Annex D – Data Extraction Form

This annex presents the structured form that will be used to extract data from each included study, as well as an example of the JSON data structure that the ELIS SLR Agent will produce for each study. The form captures bibliographic details, context, intervention characteristics, outcomes, and quality indicators.

Data Extraction Fields and Definitions:

- **Study ID:** Unique identifier for the study (for cross-referencing with screening log). *Example:* ELIS2025_001.
- **Title:** Full title of the article or report.
- **Authors:** Lead author(s) and possibly et al. (for reference listing).
- **Publication Year:** Year the study was published.
- **Publication Type/Source:** Journal name or conference, etc., and whether it's peer-reviewed (Yes/No).
- **Country/Region Studied:** The country or countries where the study's data is from or applicable. e.g., *Brazil*, *Global comparative*, *EU (multi-country)*.
- **Electoral Modality:** What type of voting system is involved – *Electronic*, *Paper-based*, *Hybrid*, or other specific modality.
- **Intervention Type:** The specific integrity mechanism or strategy evaluated. e.g., *Risk-limiting audit*, *VVPAT implementation*, *Blockchain-based ledger*, *Transparency initiative*. If multiple, all are noted.
- **Study Design:** The methodological approach of the study. e.g., *Field experiment*, *Quasi-experiment (DiD)*, *Case study*, *Simulation*, *Survey research*.
- **Evaluation Method:** How outcomes were measured or evaluated. e.g., *Statistical audit*, *Survey of voters*, *Analysis of error logs*, *Interviews*.
- **Outcomes:** For each outcome reported, we will record:
    - Outcome **Type** – whether the outcome is considered Primary or Secondary in our review's context.
    - Outcome **Reported** – a short description of what specific outcome was measured (e.g., *discrepancy rate*, *voter confidence score*, *time to detect fraud*).
    - Outcome **Measure** – how that outcome was operationalized (e.g., *% of votes recounted that differ*, *Likert scale survey response*).
    - **Design Link** – a Yes/No flag indicating if the outcome is clearly linked to a specific system design feature or intervention (for example, an outcome of "error rate" would be "Yes" linked if the errors relate to the new voting machine design).
    - **Reviewer Notes on Outcome** – comments on the outcome's significance or the strength of evidence (e.g., *"Outcome is statistically significant; causal inference strong"* or *"Outcome is perception-based, interpret with caution"*).

- **Risk of Bias Summary:** Key points from the risk of bias assessment for the study. While detailed domain-level assessments go in Annex E, here we may include a short note if a study has major limitations (e.g., *high risk of bias due to no control group*).

All these fields will be captured in a tabular format during extraction for human readability, and simultaneously in a structured JSON format for computational use.

Example Extraction (Tabular Format):

Field	Data Excerpt (Example)
Study ID	ELIS2025_001
Title	Public confidence in electronic voting
Authors	Smith, J.; Nguyen, L.
Year	2021
Peer-reviewed?	Yes
Country/Region	Brazil
Electoral Modality	Electronic
Intervention Type	VVPAT introduction
Study Design	Field experiment
Evaluation Method	Pre-post election survey + audit logs
Outcome 1 Type	Primary
Outcome 1 Reported	Change in discrepancy rate in audit recounts
Outcome 1 Measure	% of ballot counts differing between electronic and paper records
Outcome 1 Design Link	Yes (directly tests VVPAT effect on discrepancies)
Outcome 1 Notes	Discrepancy rate dropped from 2% to 0.5% after VVPAT (causal link plausible)
Outcome 2 Type	Secondary
Outcome 2 Reported	Voter confidence level
Outcome 2 Measure	Mean survey trust score (1–5 scale)
Outcome 2 Design Link	No (many factors influence trust)
Outcome 2 Notes	No significant change in trust score (could be confounded by external events)
Risk of Bias Summary	Moderate risk (no randomization, but good transparency of data)

JSON starter template: Below is a snippet of the JSON structure corresponding to the above example, showing how the ELIS SLR Agent would store the data (keys corresponding to the fields above):

```
{
  "study_id": "ELIS2025_001",
  "title": "Public confidence in electronic voting",
  "authors": ["Smith, J.", "Nguyen, L."],
  "publication_year": 2021,
  "peer_reviewed": true,
  "country_or_region": "Brazil",
  "electoral_modality": "Electronic",
  "intervention_type": "VVPAT",
  "study_design": "Field experiment",
  "evaluation_method": "Survey and audit logs",
  "outcomes": [
    {
      "outcome_type": "Primary",
      "outcome_reported": "Discrepancy rate in audit recounts",
      "outcome_measure": "Percentage of mismatched ballots",
      "design_link": true,
      "reviewer_notes": "Discrepancy reduced from 2% to 0.5% post-intervention."
    },
    {
      "outcome_type": "Secondary",
      "outcome_reported": "Voter confidence score",
      "outcome_measure": "Mean trust score (1-5)",
      "design_link": false,
      "reviewer_notes": "No significant change; external factors likely involved."
    }
  ],
  "risk_of_bias": {
    "study_design_appropriateness": "Moderate",
    "data_transparency": "High",
    "outcome_clarity": "High",
    "notes": "Well-documented methods, but no randomization."
  }
}
```

JSON example for one study; actual dataset will contain an array of such entries.

This JSON format is **schema-compliant**, meaning each field adheres to a predefined schema ensuring consistency across studies. The schema (defined in the project repository) dictates allowed values and formats (for example, outcome_type must be either "Primary" or "Secondary", risk of bias ratings must be "Low/Moderate/High"). Using a schema helps validate the data automatically – the ELIS SLR

Agent will flag any entries that do not conform (e.g., a misspelled rating), thereby reducing data errors.

All extracted data will be made available as a CSV and JSON file. The JSON is especially useful for computational analysis or for sharing data for future evidence synthesis efforts. The final version of the extraction data will be deposited on GitHub (in the ELIS SLR Agent repository) and a permanent archive (such as Zenodo or OSF) upon completion of the review.

Annex E – Evidence Quality Assessments

This annex contains tools and templates for appraising the quality of evidence from the included studies, specifically: **E.1) Risk of Bias Tool and Log** for individual studies, and **E.2) CERQual Confidence Assessment Template** for synthesized findings.

E.1 Risk of Bias Tool and Rating Log (per study)

For each included study, a risk of bias assessment will be performed across key domains. The tool is adapted to this review's needs, combining elements from standard instruments. Each domain is rated **Low, Moderate, or High** risk of bias, with a short note explaining the judgment. The domains are:

- **Study design appropriateness:** Is the study design suitable to answer the research question? *Example:* A randomized controlled trial (RCT) in the context of a new voting technology would generally be Low risk on design (if well executed), whereas an uncontrolled before-after comparison might be Moderate or High risk due to potential confounders.
- **Data transparency:** Are the study's data sources and analyses reported transparently? *Example:* High risk if the study does not share important details or data (perhaps only high-level claims without method details); Low risk if data and analysis code are available or at least described in full. We also consider whether outcome data can be verified (e.g., official audit reports published).
- **Outcome measurement clarity:** Are the outcomes measured in a direct and reliable way? *Example:* If a study measures "integrity" via a clearly defined audit discrepancy rate, that's clearer (Low risk) compared to a study using a very indirect proxy for integrity (which could introduce bias). Also, if multiple outcomes were possible but only some reported, we flag potential reporting bias here.

Additional domains that may be noted (if applicable) include selection of reported results (selective reporting within the study) and any conflicts of interest that could bias results (e.g., study authors involved in developing the technology being evaluated).

All these assessments will be logged in a **Risk of Bias Log**, a table where each study has a row with its domain ratings and notes. For example:

Study ID	Design Appropriateness	Data Transparency	Outcome Clarity	Overall Bias Rating	Notes
ELIS2025_001	Moderate	High	High	Moderate	No control group, but methods clearly described and data shared.
ELIS2025_007	Low	Moderate	Moderate	Moderate	RCT conducted, but incomplete reporting of outcome definitions.

ELIS2025_012	High	Low	Moderate	High	Descriptive study with possible selective reporting of positive results.

E.2 CERQual Confidence Rating Template (per synthesized finding)

For each major **finding or theme** that emerges from the synthesis (especially qualitative or mixed-method findings), we will assess confidence using CERQual criteria, as described in Section 3.8. The output will be a structured summary, and we will maintain a JSON record of these assessments as well.

Confidence Assessment Fields: for each finding/theme, we record:

- **Theme/Finding description** – a short identifier (e.g., *"VVPAT improves error detection"*).
- **Methodological Limitations:** Level of concern (None/Minor, Moderate, Serious) with a brief note (e.g., *"Minor concerns: most studies well-designed except one case study"*).
- **Coherence:** Level of consistency across studies (High, Moderate, Low coherence) with note (e.g., *"High coherence: all studies report reduction in discrepancies"*).
- **Data Adequacy:** Are there enough data? (Adequate or Not Adequate) with note (e.g., *"Moderate adequacy: only 3 studies, but all with decent sample sizes"*).
- **Relevance:** Direct relevance to our review context (High, Moderate, Low) with note (e.g., *"High relevance: evidence comes from national elections similar to those of interest"*).
- **Overall Confidence:** High, Moderate, Low, or Very Low – based on the above domains collectively.

We will provide a summary table in the final review for these, but here is an example in JSON format to illustrate how it might be stored (and to ensure schema consistency with our data management):

```
{
  "theme": "VVPAT reduces tally discrepancies",
  "methodological_limitations": "Minor concerns",
  "coherence": "High",
  "data_adequacy": "Moderate",
  "relevance": "High",
  "overall_confidence": "High"
}
```

This example would correspond to a finding that voter-verified paper audit trails (VVPAT) reduce discrepancies between electronic and paper counts. The judgment indicates that across studies, there were only minor methodological issues, findings were consistent (high coherence), data volume was moderate, and all in contexts highly relevant – leading to a high confidence in this conclusion.

All such JSON entries for each finding will be compiled. The **full set of CERQual assessments** will be available in the GitHub repository (for transparency) and referenced in the review's results section.

We will also likely include a concise table in the review manuscript listing each key theme with its overall confidence level and a one-line explanation.

The *overall confidence assessments* will help readers understand which conclusions are strongly supported and which should be interpreted cautiously. This approach to assessing cumulative evidence strength is aligned with GRADE/CERQual recommendations and provides a structured way to discuss the implications of the findings.

Lastly, we will cross-reference the risk-of-bias data (from E.1) with the confidence in evidence (E.2). For instance, if all studies on a theme had high bias, even if coherent, our confidence might be capped at moderate. This interplay will be explicitly discussed in the review.

(The JSON schema for confidence assessments will be included in the repository. It mirrors the fields above, ensuring that each domain is captured for every theme. This ensures our outputs are machine-readable and easily sharable.)

Annex F – Automation Workflow & AI

This annex details the role of automation tools and AI (specifically the **ELIS SLR Agent** and large language models) in supporting the systematic review process, as well as the human oversight measures in place. The ELIS SLR Agent (v2.0) is an open-source, Python-based workflow tool developed for this project that automates certain labor-intensive tasks while preserving full auditability.

Workflow Overview: The table below summarizes the stages of the review, the tasks handled by automation/AI, and the corresponding human oversight:

Stage	Automation / AI Role	Human Oversight
1. Search Management	Zotero auto-imports metadata and tags sources	Reviewer verifies imports and de-duplicates
2. Title/Abstract Screening	ELIS SLR Agent flags likely inclusions/exclusions via keyword heuristics	Final inclusion decisions made by reviewer
3. Full-Text Screening	LLMs (e.g., ChatGPT, Claude.ai, NotebookLM) assist in summarizing or highlighting	Reviewer reads and validates all inclusions
4. Data Extraction	ELIS SLR Agent outputs structured JSON/CSV fields	Reviewer confirms and edits extractions as needed
5. Bias & Confidence Rating	ELIS SLR Agent suggests bias levels and CERQual inputs	Final ratings assigned by reviewer (Annex E)
6. Thematic Synthesis	LLMs help identify recurring terms/themes across studies	Themes confirmed, refined, and interpreted by reviewer
7. Audit Trail	All AI-assisted outputs logged in GitHub; no auto-deletion	Reviewer commits changes and version history

Tools Used

Tool	Purpose
Zotero	Reference management and de-duplication
ELIS SLR Agent	Automated tagging, bias flagging, JSON extraction
ChatGPT / Claude.ai / NotebookLM	Pattern detection, synthesis suggestions
GitHub	Version control, audit log, reproducibility

Oversight and Integrity Commitments

- All final study selections, bias ratings, and syntheses are performed or validated by the **lead reviewer**.
- LLMs are used to **assist**, not decide — particularly in cases requiring interpretation or inclusion judgement.
- The protocol adheres to the **UK Research Integrity Office (UKRIO) Principles on AI Use in Research**.
- Versioning and traceability are ensured through open-access logging on GitHub.

Annex G – Evidence Gap Confirmation and Supporting Literature

This annex documents the validation of the following statement in the ELIS Rationale:

"To date, there is no consolidated systematic review that jointly examines technological, operational, and institutional strategies for electoral integrity across both electronic and paper-based systems."

G.1 Purpose and Scope

To confirm the accuracy of this statement, a structured review of peer-reviewed literature from the past 10 years (2015–2025) was conducted. The objective was to identify any existing **systematic literature reviews**, **scoping reviews**, or other formal evidence syntheses that:

- Jointly examine **technological**, **operational**, and **institutional** strategies affecting electoral integrity; and
- Include both **electronic** and **paper-based** voting systems within the same review framework.

The search covered political science, governance, and election technology domains, prioritising systematic reviews published in high-impact journals and by major academic publishers.

G.2 Summary of Findings

The review confirmed that **no consolidated systematic review** has yet addressed this combined scope. Existing literature falls into distinct silos:

- **Technology-focused reviews** (e.g. on internet voting, blockchain, or biometric authentication) are common but narrowly focused on specific digital systems or use cases. They rarely incorporate paper-based modalities or address operational design.
- **Operational studies** tend to examine logistics, election-day procedures, or administrative performance, often as isolated variables. These do not address the interaction between process and technology or include broader institutional strategies.
- **Institutional analyses** and **integrity indices** (e.g. the Perceptions of Electoral Integrity dataset) provide high-level assessments but do not systematically review the impact of concrete system design features or specific interventions.

As a result, **no existing synthesis provides an interdisciplinary, comparative overview** of how technological, operational, and institutional features affect electoral integrity and auditability across voting system types.

This confirms the novelty and relevance of the ELIS review within both academic and practitioner communities.

G.3 Notable Existing Reviews and Gaps

Domain	Representative Review	Coverage	Gap
Electronic voting	Turnbull-Dugarte & Devine (2023) – *Systematic review of i-voting pilots*	Strong on internet voting and public trust	Does not address paper-based systems or institutional context
Election logistics	Apiriba & Lim (2025) – *SLR on logistics and election performance*	Covers supply chain and operational design	Does not address technology or governance
Institutional frameworks	Norris et al. (2014) – *PEI Index and integrity metrics*	Global scope on electoral quality	Not a systematic review; lacks mechanism-level detail
Biometric voting systems	Asimakopoulos et al. (2025) – *Impact of ICT on democratic processes*	Covers biometrics in Kenya and Asia	Focused on technology only
Election observation & trust	Kelley (2012), IDEA (2024), IFES (2021)	Emphasise transparency and oversight	Not structured as systematic reviews; limited to post-election practices

These examples underscore that while robust literature exists in each area, there is a **lack of integrated synthesis** across design dimensions and voting system types. The ELIS protocol addresses this gap.

G.4 Justification for PRQ Framing

In light of this evidence, the ELIS primary research question is framed with causal language:

> *"What operational and technological features of electoral systems have been shown to improve the integrity and auditability of elections since 1990?"*

To ensure methodological clarity, the protocol specifies that this phrasing includes:

- Findings from experimental and quasi-experimental studies,
- Observational studies and comparative analyses,
- Qualitative case studies and structured evaluations with empirical grounding.

Where causal inference is weak or contested, the review will clearly distinguish between robust evidence and more tentative findings. This framing is appropriate for informing electoral reform and policy design.

Annex H – Workflow Diagram & Tools

This annex provide a visual and textual description of the ELIS protocol's streamlined toolchain, including:

- Zotero for reference management and metadata tagging
- ELIS SLR Agent for automated screening, tagging, and extraction
- GitHub for version control and full audit trails
- (Optional) AI-assisted summarization tools with manual verification

This diagram illustrates the flow from literature retrieval to final inclusion, showing how automation and human validation interact across the pipeline.