Systematic Literature Reviews Explained
A Methodological Overview

Dennis Hoffmann

Abstract

A systematic literature review (SLR) provides a rigorous, replicable method for synthesizing the existing body of evidence on a clearly defined research question. As publication volumes in finance and data science continue to grow rapidly (Bornmann and Mutz, 2015), narrative reviews increasingly fall short in providing unbiased coverage of the literature. This document argues that SLR methodology, while well codified in health sciences and software engineering, requires domain-specific adaptation when applied to machine learning research in finance, and that computational review pipelines are the mechanism through which such adaptation becomes reproducible. It covers the definitional foundations and review typology of SLRs, details the three-phase process of planning, conducting, and reporting, and traces the adoption of systematic reviewing in finance and data science. The document expands on current challenges, including the role of large language models in evidence synthesis, and discusses how computational review pipelines can operationalize these principles through API-based search, automated snowballing, and structured filtering.

1 Introduction

The volume of academic literature in finance and data science has grown substantially over the past two decades. Bornmann and Mutz (2015) estimate that global scientific output has grown at approximately 8–9% per year since the mid-twentieth century. Applied fields such as machine learning have plausibly exhibited even faster rates, given the proliferation of conference venues and preprint servers since the 2010s. Financial economics has seen a parallel expansion in published articles, working papers, and conference proceedings, driven by open-access outlets, preprint servers such as SSRN, and the rapid development of ML applications in quantitative finance. Researchers now face the challenge of navigating this expanding evidence base to identify what is known, what remains contested, and where critical knowledge gaps exist.

Traditional narrative reviews, while valuable for expert commentary and hypothesis generation, are vulnerable to selection bias, lack reproducibility, and struggle with the volume of new publications (Snyder, 2019). The systematic literature review offers a principled alternative: a methodology designed to provide transparent, reproducible evidence synthesis through explicit search protocols, predefined eligibility criteria, and standardized reporting (Gough et al., 2017; Higgins et al., 2019).

The SLR originated in evidence-based medicine, where organizations such as the Cochrane Collaboration (founded 1993) and the Campbell Collaboration (Davies, 2000) operationalized systematic reviewing as a global methodology for synthesizing clinical and social-science evidence. Tranfield et al. (2003) subsequently adapted these principles for management research, and the methodology has since diffused into information systems (vom Brocke et al., 2009), software engineering (Kitchenham, 2004), and finance (Linnenluecke et al., 2020).

Yet the transfer of SLR methodology across disciplines is not frictionless. Finance and ML research present domain-specific challenges (working-paper cultures, conference-dominant dissemination, proprietary databases) that require methodological adaptation beyond what generic SLR guidance provides. This document argues that *computational review pipelines are the mechanism through which such adaptation becomes reproducible.* It examines how the principles of systematic reviewing interact with the practical realities of evidence synthesis in a fast-moving, computationally intensive field.

Section 2 establishes definitional foundations and a taxonomy of review types. Section 3 details the three-phase SLR process, including quality assessment frameworks. Section 4 examines the adoption of SLRs in finance and data science, with particular attention to domain-specific challenges. Section 5 discusses criticisms, computational approaches, LLMs in systematic reviewing, and presents a worked pipeline example with an illustrative PRISMA flow diagram.

2 Foundations: Definition, Principles, and Review Typology

2.1 Defining a Systematic Literature Review

A systematic literature review is a form of secondary research that uses a clearly defined, replicable methodology to identify, select, critically appraise, and synthesize all relevant primary studies addressing a specific research question (Fink, 2019). The Cochrane Handbook defines a systematic review as one that "attempts to collate all empirical evidence that fits pre-specified eligibility criteria in order to answer a specific research question" (Higgins et al., 2019). The defining characteristics, as articulated across the methodological literature, include five core principles (Tranfield et al., 2003; Cooper, 2017; Gough et al., 2017):

1. **Transparency**: every step of the review process is documented, enabling scrutiny by other researchers.

2. **Reproducibility**: another researcher following the same protocol should be able to replicate the search, screening, and synthesis process.

3. **Comprehensiveness**: the review aims to identify all relevant studies, not merely a convenient subset.

4. **Explicit methodology**: inclusion and exclusion criteria, search strategies, and synthesis methods are specified *a priori*.

5. **Minimization of bias**: systematic procedures reduce the risk of subjective study selection and interpretation.

It is important to distinguish between a systematic review and a meta-analysis. A systematic review is the overarching methodology: the structured process for identifying and synthesizing evidence. A meta-analysis is one possible quantitative synthesis technique within a systematic review, using statistical methods to combine results of individual studies (Borenstein et al., 2009; Cooper et al., 2019). Not all systematic reviews include a meta-analysis; many employ narrative or thematic synthesis instead.

2.2 Systematic vs. Narrative Reviews

The distinction between systematic and narrative reviews is best understood as a spectrum rather than a strict dichotomy. Table 1 summarizes the key differences across seven criteria (Snyder, 2019).

Table 1: Comparison of narrative and systematic literature reviews.

Dimension	**Narrative Review**	**Systematic Review**
Protocol	Typically absent	Defined *a priori*
Search strategy	Selective, implicit	Comprehensive, explicit
Study selection	At reviewer's discretion	Predefined criteria
Quality appraisal	Rarely conducted	Standardized assessment
Bias risk	High (selection bias)	Minimized by protocol
Reproducibility	Low	High
Synthesis method	Qualitative commentary	Structured (narrative/quant.)

Narrative reviews remain valuable for providing broad overviews, offering expert interpretation, and generating hypotheses. Semi-systematic and scoping reviews occupy intermediate positions, combining systematic search with more flexible synthesis. The choice depends on the research question, the maturity of the evidence base, and the intended contribution (Petticrew and Roberts, 2006; Paul and Rialp Criado, 2020).

2.3 Review Type Taxonomy

Systematic reviews represent one of several structured review methodologies. Table 2 presents an overview of the main review types and the contexts in which each is most appropriate (Snyder, 2019; Petticrew and Roberts, 2006; Paul and Rialp Criado, 2020).

The choice of review type depends on the research question, field maturity, and the nature of the available evidence. Hybrid approaches are increasingly common: for example, combining a systematic review with bibliometric analysis to both map the intellectual structure of a field and synthesize its substantive findings (Linnenluecke et al., 2020; Donthu et al., 2021).

Table 2: Taxonomy of structured review types.

Review Type	Purpose	Approach	When to Use
Systematic review	Synthesize all evidence on a focused question	Protocol-driven search, screening, and synthesis	Well-defined research question; evidence synthesis needed
Scoping review	Map the breadth of a research area	Systematic search; charting rather than synthesis	Emerging field; identify gaps and key concepts
Semi-systematic review	Thematic analysis of a broad topic	Systematic search; qualitative, thematic synthesis	Broad, multidisciplinary topics
Meta-analysis	Quantitative synthesis of effect sizes	Statistical pooling of results from comparable studies	Sufficient homogeneous quantitative studies exist
Bibliometric review	Map intellectual structure of a field	Citation analysis, co-authorship, keyword co-occurrence	Large literatures; identify clusters and trends

3 The SLR Process

The systematic literature review process comprises three main phases: planning, conducting, and reporting (Tranfield et al., 2003; Kitchenham, 2004). Each phase involves distinct activities, decision points, and documentation requirements.

3.1 Planning Phase

The planning phase establishes the scope, objectives, and methodological blueprint for the review.

Research question formulation. A well-defined research question is the foundation of any SLR. In evidence-based medicine, the PICO framework (Population–Intervention–Comparison–Outcome) provides a standard structure for formulating answerable questions (Richardson et al., 1995). For business and management research, Denyer and Tranfield (2009) developed the CIMO framework (Context–Intervention–Mechanism–Outcome), which accommodates the more complex causal structures typical of organizational research. The choice of framework depends on the discipline: PICO suits intervention-focused questions, while CIMO better serves questions about mechanisms and contextual factors (Fisch and Block, 2018). The research question determines the scope of the search, the inclusion and exclusion criteria, and the synthesis strategy.

Protocol development. A review protocol specifies the objectives, search strategy, eligibility criteria, data extraction procedures, and synthesis methods *before* the review begins. Registering the protocol on platforms such as PROSPERO, an international prospective

register of systematic reviews (Booth et al., 2012), or the Open Science Framework (OSF) enhances transparency and reduces the risk of post hoc modifications.

Inclusion and exclusion criteria. The protocol must define clear criteria for study eligibility. Common dimensions include publication date range, language, publication type (peer-reviewed articles, working papers, conference proceedings), geographic scope, and substantive relevance to the research question.

3.2 Conducting Phase

The conducting phase implements the protocol through systematic search, screening, data extraction, and quality appraisal.

Search strategy. The search strategy must balance comprehensiveness with precision. Key elements include database selection, search string design using Boolean operators and field-specific terms, and a strategy for grey literature. Database selection requires careful consideration of coverage: Scopus and Web of Science provide broad multidisciplinary indexing, while SSRN and NBER cover working papers in economics and finance; OpenAlex offers open metadata for over 250 million works. For ML-in-finance research, conference proceedings (NeurIPS, ICML) and preprint servers (ArXiv) represent primary dissemination channels and must be explicitly included.

Grey literature (working papers, technical reports, and conference proceedings) is not a peripheral concern in ML and finance research but a dominant form of dissemination. Garousi et al. (2019) provide guidelines for systematically including grey literature, arguing that excluding it risks missing substantial portions of the evidence base, particularly in fields where peer-reviewed publication lags behind practice.

Forward and backward snowballing, the practice of examining reference lists of included studies and identifying papers that cite them, is a widely recommended complement to database searches. Wohlin (2014) formalized snowballing procedures for systematic literature studies, demonstrating that snowballing can identify relevant studies missed by database searches alone. vom Brocke et al. (2009) emphasize that the search process must be documented in sufficient detail to allow replication.

Screening. Screening typically proceeds in two stages: title and abstract screening, followed by full-text review. At each stage, studies are assessed against predefined eligibility criteria. When multiple reviewers are involved, inter-rater reliability should be assessed (Templier and Paré, 2015). Tools such as Covidence, Rayyan, and ASReview (van de Schoot et al., 2021) can support the screening process by managing records, tracking decisions, and, in the case of ASReview, prioritizing records using active learning.

Data extraction. Standardized data extraction forms ensure consistent information collection across included studies. vom Brocke et al. (2009) recommend concept matrices to organize extracted data along key thematic dimensions.

Quality appraisal. Study-level quality assessment evaluates the methodological rigour of each included study. While standardized quality appraisal tools are well established in medical research, their adaptation for empirical finance remains limited (Cooper, 2017). Relevant quality dimensions include research design, sample characteristics, data quality, analytical methods, and reporting completeness.

3.3 Reporting and Synthesis

The reporting phase communicates findings in a structured manner.

The PRISMA framework. The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) framework provides the standard for reporting systematic reviews. Originally published by Moher et al. (2009), the PRISMA statement includes a checklist of items that should be reported and a flow diagram documenting the study selection process. The updated PRISMA 2020 statement (Page et al., 2021) expanded the checklist and introduced a revised flow diagram reflecting contemporary search and screening practices. PRISMA-S (Rethlefsen et al., 2021) extends the framework specifically for reporting literature searches, addressing the transparency gap in how search strategies are documented. PRISMA adoption has been increasing across finance and management journals, reflecting broader demand for reproducible review methodologies (Linnenluecke et al., 2020).

Narrative synthesis. Narrative synthesis organizes and interprets findings thematically, grouping studies by research question, methodology, or key findings. This approach is appropriate when included studies are too heterogeneous for quantitative pooling, which is common in finance and data science research.

Quantitative meta-analysis. When studies report comparable quantitative outcomes, meta-analysis provides statistical techniques for combining effect sizes, assessing heterogeneity, and testing for publication bias (Borenstein et al., 2009).

Bibliometric analysis as a synthesis tool. Bibliometric techniques (co-citation analysis, bibliographic coupling, keyword co-occurrence) can serve as a synthesis method *within* an SLR, revealing the intellectual structure and thematic clusters of the included literature (Donthu et al., 2021). This use of bibliometrics as a within-review synthesis tool is distinct from its role as a standalone complementary methodology (discussed in Section 4.3); the two applications are not in tension but serve different analytical purposes.

3.4 Quality Assessment of Systematic Reviews

Beyond assessing individual studies within a review, researchers also need frameworks for evaluating the quality of systematic reviews themselves. AMSTAR 2 (Shea et al., 2017) provides a critical appraisal tool for systematic reviews, identifying 16 domains including protocol registration, search comprehensiveness, and appropriateness of synthesis methods. ROBIS (Whiting et al., 2016) assesses risk of bias across four domains: study eligibility, identification and selection of studies, data collection and study appraisal, and synthesis and findings. These tools were developed for health-sciences reviews; their adaptation to finance and ML, where intervention-outcome logic differs fundamentally from clinical research, remains an open methodological question.

4 SLRs in Finance and Data Science

4.1 Finance-Specific Methodological Challenges

The transfer of SLR methodology from health sciences to finance introduces challenges that generic guidance does not address.

The working-paper problem. SSRN and NBER serve as primary dissemination channels in financial economics. Significant findings often circulate as working papers for years before formal publication, with multiple revised versions appearing under the same title. A systematic search that captures only published articles misses this working-paper discourse; one that includes SSRN must contend with versioning: the same paper may appear as three or four distinct records with different dates and content.

Conference-dominant fields. ML research is disseminated primarily through conference proceedings (NeurIPS, ICML, AAAI) and ArXiv preprints rather than traditional journal articles. Scopus and Web of Science index these sources incompletely, making API-based search through OpenAlex or Semantic Scholar essential for adequate coverage.

Proprietary data and reproducibility. Much empirical finance research relies on proprietary databases (WRDS, Bloomberg, Refinitiv) that are not indexed by any bibliographic database and whose licensing terms restrict data sharing. This creates a reproducibility challenge for both the primary studies under review and the review process itself: replicating a search for "studies using CRSP data" requires knowing which studies used CRSP, information that is rarely captured in metadata.

Rapid methodological evolution. The pace of methodological development in ML means that a review initiated today may be partially outdated within months. This pressure favours living review models (Section 5.5) and computational pipelines that can be re-executed as new evidence accumulates.

4.2 Applied SLRs in ML and Finance

Several reviews have surveyed ML applications in finance using varying degrees of methodological structure. Henrique et al. (2019) reviewed ML techniques for financial market prediction, mapping methods, data sources, and performance metrics. Sezer et al. (2020) reviewed deep learning methods for financial time series forecasting, identifying trends and research gaps spanning 2005–2019. Ozbayoglu et al. (2020) surveyed deep learning across algorithmic trading, portfolio management, and risk assessment. Nosratabadi et al. (2020) reviewed advanced ML and deep learning methods across economics and data science. Goodell et al. (2021) used bibliometric analysis to map the intellectual structure of AI and ML in finance. These reviews illustrate the value of systematic methods in fields characterized by rapid publication cycles and methodological diversity.

4.3 Bibliometric Analysis as a Complementary Method

Bibliometric analysis uses quantitative methods to study patterns of scholarly communication: co-citation networks, bibliographic coupling, keyword co-occurrence, and co-authorship analysis (Zupic and Čater, 2015; Donthu et al., 2021). Software tools such as VOSviewer (van

Eck and Waltman, 2010) and Bibliometrix have made these analyses accessible to researchers without training in scientometrics.

As noted in Section 3.3, bibliometric techniques can serve as a synthesis method within an SLR. They can also function as a standalone methodology, providing a macro-level map of a field's intellectual structure that complements the micro-level synthesis of individual study findings (Linnenluecke et al., 2020; Zupic and Čater, 2015). The distinction matters for protocol design: a researcher must decide whether bibliometric analysis is part of the review's synthesis strategy (requiring integration into the SLR protocol) or a separate, complementary analysis with its own objectives.

5 Challenges, Computational Approaches, and Future Directions

5.1 Criticisms and Limitations

Despite their methodological advantages, SLRs have well-documented limitations. The most commonly cited is resource intensity: a rigorous SLR can take months or years, requiring substantial investment in search, screening, and synthesis. Kraus et al. (2020) note that the quality of published SLRs varies considerably, and that the label "systematic" does not guarantee rigour; some reviews claim systematic status while lacking a registered protocol, comprehensive search, or transparent screening process.

Publication bias, the tendency for studies with statistically significant results to be published at higher rates, affects the evidence base on which any review draws. In empirical finance, this concern is amplified by the "factor zoo" problem: hundreds of published anomalies, many of which may reflect data mining rather than genuine economic phenomena (Harvey and Liu, 2021). A systematic review of factor-based strategies inherits the publication bias embedded in the underlying literature.

In fast-moving fields, a systematic review may be partially outdated by the time it is published, given typical completion and peer-review timelines. This temporal limitation has motivated interest in living systematic reviews (Section 5.5).

5.2 Computational Tools for Evidence Synthesis

Computational tools address the scalability challenges of systematic reviewing, but their maturity varies.

Active learning. ASReview (van de Schoot et al., 2021) uses active learning to prioritize records during screening. The key metric is Work Saved over Sampling at 95% recall (WSS@95): the proportion of records a reviewer can skip while still identifying 95% of relevant studies. Reported WSS@95 values vary by dataset and review context; the metric does not guarantee that 95% of screening can be skipped; it measures the trade-off between recall and screening effort at a specific recall threshold. Active learning performs best when the relevant fraction of the corpus is small, which is typical of systematic reviews but requires careful calibration of the stopping criterion.

API-based search infrastructure. OpenAlex and CrossRef provide programmatic access to bibliographic metadata and citation networks. These APIs enable automated search and snowballing, replacing manual database queries with reproducible, version-controlled retrieval. For finance-specific searches, OpenAlex's concept and topic hierarchies allow filtering by field of study, though coverage of working papers and conference proceedings remains incomplete.

Screening and management tools. Beyond ASReview, tools such as Covidence and Rayyan provide collaborative screening interfaces, conflict resolution workflows, and PRISMA-compliant reporting. The choice of tool depends on team size, budget, and whether active learning is desired.

5.3 Large Language Models in Systematic Reviewing

Large language models are being explored for multiple stages of the SLR process, including search query formulation, abstract screening, data extraction, and narrative synthesis (Khraisha et al., 2024). The evidence to date warrants caution rather than enthusiasm.

Lieberum et al. (2025) conducted a scoping review of LLM use in systematic reviews, concluding that current tools are "not yet ready for use" as a replacement for human reviewers, though they may serve as a useful complement. Gartlehner et al. (2025) evaluated AI-assisted data extraction using a large language model within systematic reviews, finding that AI reduced extraction time and achieved comparable or slightly higher accuracy than human-only extraction, though concordance between methods was moderate and human verification remained essential. Scherbakov et al. (2025) reviewed the emergence of LLMs as tools in literature reviews, identifying substantial variability in accuracy depending on task complexity and prompt design.

The central risk is hallucination: LLMs can generate plausible but fabricated references, misattribute findings, and produce confident but incorrect summaries of study results. For data extraction, where accuracy is paramount, current LLMs introduce error rates that are unacceptable without human verification. For screening, where the cost of a false negative is a missed relevant study, LLMs may serve as a complement to human reviewers but cannot yet replace them. The field would benefit from standardized benchmarks and reporting standards for AI-assisted reviewing, analogous to PRISMA for traditional reviews.

5.4 Reproducible Review Pipelines

The methodological principles discussed above (transparent search, systematic screening, structured reporting) can be operationalized through computational review pipelines. This section describes a typical pipeline architecture built on the OpenAlex API, illustrating how each SLR stage maps to a concrete computational step.

Several design choices in such pipelines respond directly to the finance-specific challenges identified in Section 4.1. API-based snowballing addresses the coverage gaps that arise in conference-dominant fields, where traditional database indexing is incomplete. Provenance tracking mitigates the working-paper versioning problem by recording which search path or snowball step led to each record's inclusion. Configurable quality filters with year-dependent

citation thresholds accommodate the rapid methodological evolution of ML research, where recent work has had less time to accumulate citations.

Pipeline overview. A computational review pipeline is typically implemented in Python, version-controlled in Git, and configured via a YAML file that specifies search queries, database parameters, filtering criteria, and journal tier definitions. The pipeline executes the following stages:

1. **Database search**: Keyword queries, constructed from Boolean combinations of domain-specific terms, are executed against the OpenAlex API. The resulting records form the initial candidate set.

2. **API-based snowballing**: Using the OpenAlex API, the pipeline performs forward and backward snowballing to a configurable depth, discovering all works that cite or are cited by the expanding corpus.

3. **Field-of-study filtering**: Each candidate record is checked against field-of-study classifications (hierarchical subject tags assigned to indexed works) and relevance keywords. Records classified outside the target disciplines are excluded.

4. **Relevance filtering**: Remaining records are screened against domain-specific keywords to ensure topical relevance.

5. **Quality filtering**: Records must satisfy language, DOI availability, abstract presence, journal tier (a ranking of publication venues by quality, where untiered venues are excluded), and citation threshold (a minimum citation count for inclusion) requirements. Citation thresholds can be made year-dependent to account for the fact that recent papers have had less time to accumulate citations.

6. **Export with provenance**: The final corpus is exported as BibTeX, with full provenance tracking (which search query or snowball path led to each inclusion, and at what depth).

Illustrative PRISMA flow. Figure 1 presents an illustrative PRISMA-style flow diagram for an OpenAlex-based review pipeline. The numbers are representative of a typical run in an applied finance domain; actual volumes will vary depending on topic breadth and query specificity.

Methodological lessons. Three observations from building pipelines of this kind are relevant to the broader discussion of SLR methodology.

First, *the field-of-study filter is the most consequential methodological choice.* Filtering by OpenAlex field-of-study classifications can remove the majority of candidate records before any content-based screening occurs. This is efficient but fragile: an interdisciplinary paper published in a statistics or computer science journal might be classified outside the target field and would be excluded. The choice of field-of-study filter encodes a disciplinary boundary that shapes the final corpus.

Second, *citation thresholds embed temporal bias.* Year-dependent citation thresholds serve as a quality proxy, but they systematically disadvantage recent work and work in niche subfields. Pipelines can mitigate this by always including seminal papers and review/survey

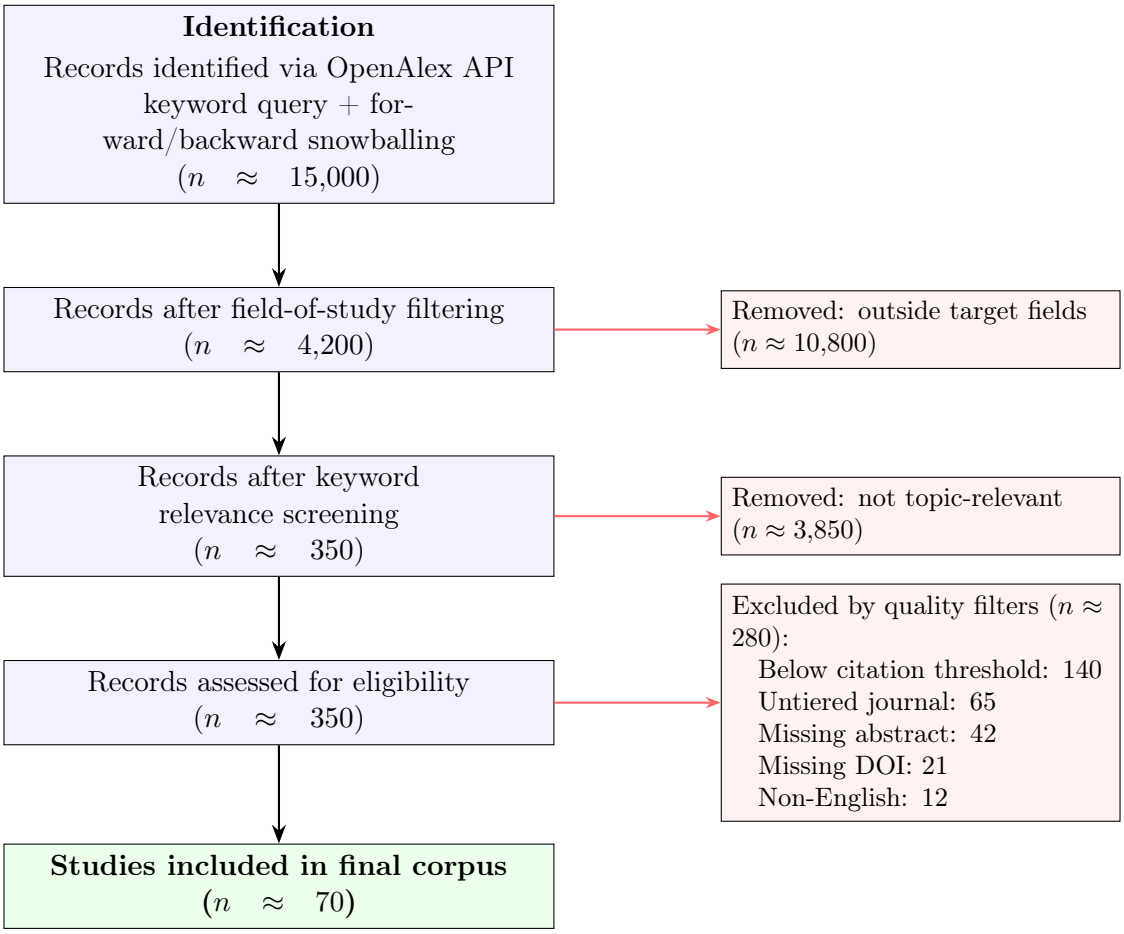

Figure 1: Illustrative PRISMA-style flow diagram for an OpenAlex-based review pipeline. Numbers are representative of a typical run in an applied finance domain; actual volumes will vary depending on topic breadth and query specificity.

articles regardless of citation count, but the trade-off between precision and recall remains a judgement call. Similarly, journal-tier filtering encodes prestige hierarchies that may exclude valid work from newer, interdisciplinary, or open-access outlets that have not yet achieved high rankings.

Third, *provenance tracking transforms reproducibility from an aspiration to an implementation.* Every included paper carries a complete provenance chain: which search query or snowball step led to its discovery, through which direction (forward or backward), at what depth, and when. This provenance log, stored as structured JSON, allows any researcher to trace why a specific paper was included or excluded, achieving a level of transparency that manual SLR documentation rarely achieves.

5.5 The Future of Evidence Synthesis

Living systematic reviews represent a shift from static, one-time reviews to continuously updated evidence synthesis (Elliott et al., 2017). In a living review, the search is periodically re-executed, new studies are screened and incorporated, and the synthesis is updated. Computational pipelines make living reviews practical: a pipeline that can be re-run with updated date ranges and citation data produces an auditable delta between review versions.

The registered reports movement aligns with systematic reviewing: pre-registering a review protocol for peer review before conducting the review itself eliminates the risk of post hoc modifications and ensures that the review's contribution is evaluated on methodological grounds rather than the novelty of its findings.

The convergence of open science, computational infrastructure, and methodological rigour points toward a future in which systematic reviews are not isolated publications but maintained research artefacts: version-controlled, continuously updated, and computationally reproducible.

6 Conclusion

This document argued that SLR methodology, while well codified in health sciences and software engineering, requires domain-specific adaptation for ML-in-finance research, and that computational review pipelines provide the mechanism for making such adaptation reproducible.

The evidence presented supports this thesis on several fronts. Finance-specific challenges (working-paper cultures, conference-dominant dissemination, proprietary databases, rapid methodological evolution) are not peripheral complications but fundamental features of the evidence landscape that shape every stage of the review process, from search strategy to quality assessment. Generic SLR guidance, rooted in the Cochrane tradition of clinical trial synthesis, does not address these challenges.

As discussed in Section 5.4, the core SLR principles (transparency, comprehensiveness, bias minimization) can be operationalized computationally: a version-controlled pipeline with structured provenance logging achieves a level of reproducibility and auditability that manual documentation cannot match. At the same time, seemingly technical decisions

(which field-of-study filter to apply, what citation threshold to set) are in fact methodological choices with substantial consequences for the final corpus.

LLMs and active learning tools are reshaping the practical landscape of evidence synthesis, but the current evidence does not support their use as replacements for human judgement in screening or data extraction. They are best understood as complements that reduce effort while preserving the human reviewer's role as the final arbiter of relevance and quality.

For researchers in finance and data science, the adoption of systematic reviewing is not merely a methodological preference but a response to the scale and complexity of the evidence base they confront. The discipline of a systematic approach, codified in a protocol, executed through a reproducible pipeline, and reported against PRISMA standards, strengthens both the credibility and the practical utility of their work.

References

Andrew Booth, Mike Clarke, Gail Dooley, Davina Ghersi, David Moher, Mark Petticrew, and Lesley Stewart. The nuts and bolts of PROSPERO: An international prospective register of systematic reviews. *Systematic Reviews*, 1(1):2, 2012. doi: 10.1186/2046-4053-1-2.

Michael Borenstein, Larry V. Hedges, Julian P. T. Higgins, and Hannah R. Rothstein. *Introduction to Meta-Analysis*. Wiley, 2009. doi: 10.1002/9780470743386.

Lutz Bornmann and Rüdiger Mutz. Growth rates of modern science: A bibliometric analysis based on the number of publications and cited references. *Journal of the Association for Information Science and Technology*, 66(11):2215–2222, 2015. doi: 10.1002/asi.23329.

Harris Cooper. *Research Synthesis and Meta-Analysis: A Step-by-Step Approach*. SAGE, Thousand Oaks, CA, 5th edition, 2017.

Harris Cooper, Larry V. Hedges, and Jeffrey C. Valentine, editors. *The Handbook of Research Synthesis and Meta-Analysis*. Russell Sage Foundation, New York, 3rd edition, 2019.

Philip Davies. The relevance of systematic reviews to educational policy and practice. *Oxford Review of Education*, 26(3–4):365–378, 2000. doi: 10.1080/713688543.

David Denyer and David Tranfield. Producing a systematic review. In David A. Buchanan and Alan Bryman, editors, *The SAGE Handbook of Organizational Research Methods*, chapter 39, pages 671–689. SAGE, London, 2009.

Naveen Donthu, Satish Kumar, Debmalya Mukherjee, Nitesh Pandey, and Weng Marc Lim. How to conduct a bibliometric analysis: An overview and guidelines. *Journal of Business Research*, 133:285–296, 2021. doi: 10.1016/j.jbusres.2021.04.070.

Julian H. Elliott, Anneliese Synnot, Tari Turner, Mark Simmonds, Elie A. Akl, Steve McDonald, Georgia Salanti, Joerg Meerpohl, Harriet MacLehose, John Hilton, David Tovey, Ian Shemilt, and James Thomas. Living systematic review: 1. Introduction—the why, what, when, and how. *Journal of Clinical Epidemiology*, 91:23–30, 2017. doi: 10.1016/j.jclinepi.2017.08.010.

Arlene Fink. *Conducting Research Literature Reviews: From the Internet to Paper*. SAGE, Thousand Oaks, CA, 5th edition, 2019.

Christian Fisch and Joern Block. Six tips for your (systematic) literature review in business and management research. *Management Review Quarterly*, 68(2):103–106, 2018. doi: 10.1007/s11301-018-0142-x.

Vahid Garousi, Michael Felderer, and Mika V. Mäntylä. Guidelines for including grey literature and conducting multivocal literature reviews in software engineering. *Information and Software Technology*, 106:101–121, 2019. doi: 10.1016/j.infsof.2018.09.006.

Gerald Gartlehner, Shannon Kugley, Karen Crotty, and Meera Viswanathan. Artificial intelligence–assisted data extraction with a large language model: A study within reviews. *Annals of Internal Medicine*, 178(12):1763–1771, 2025. doi: 10.7326/ANNALS-25-00739.

John W. Goodell, Satish Kumar, Weng Marc Lim, and Debidutta Pattnaik. Artificial intelligence and machine learning in finance: Identifying foundations, themes, and research clusters from bibliometric analysis. *Journal of Behavioral and Experimental Finance*, 32: 100577, 2021. doi: 10.1016/j.jbef.2021.100577.

David Gough, Sandy Oliver, and James Thomas. *An Introduction to Systematic Reviews*. SAGE, London, 2nd edition, 2017.

Campbell R. Harvey and Yan Liu. Lucky factors. *Journal of Financial Economics*, 141(2): 413–435, 2021. doi: 10.1016/j.jfineco.2021.04.014.

Bruno Miranda Henrique, Vinicius Amorim Sobreiro, and Herbert Kimura. Literature review: Machine learning techniques applied to financial market prediction. *Expert Systems with Applications*, 124:226–251, 2019. doi: 10.1016/j.eswa.2019.01.012.

Julian P. T. Higgins, James Thomas, Jacqueline Chandler, Miranda Cumpston, Tianjing Li, Matthew J. Page, and Vivian A. Welch, editors. *Cochrane Handbook for Systematic Reviews of Interventions*. Wiley, 2nd edition, 2019. doi: 10.1002/9781119536604.

Qusai Khraisha, Sophie Put, Johanna Kappenberg, Azza Warraitch, and Kristin Hadfield. Can large language models replace humans in systematic reviews? a comprehensive investigation. *Research Synthesis Methods*, 15(4):616–626, 2024. doi: 10.1002/jrsm.1715.

Barbara A. Kitchenham. Procedures for performing systematic reviews. Technical Report TR/SE-0401, Keele University, 2004.

Sascha Kraus, Matthias Breier, and Sonia Dasí-Rodríguez. The art of crafting a systematic literature review in entrepreneurship research. *International Entrepreneurship and Management Journal*, 16(3):1023–1042, 2020. doi: 10.1007/s11365-020-00635-4.

Judith-Lisa Lieberum, Markus Toews, Maria-Inti Metzendorf, Felix Heilmeyer, Waldemar Siemens, Christian Haverkamp, Daniel Böhringer, Joerg J. Meerpohl, and Angelika Eisele-Metzger. Large language models for conducting systematic reviews: On the rise, but not yet ready for use—a scoping review. *Journal of Clinical Epidemiology*, 181:111746, 2025. doi: 10.1016/j.jclinepi.2025.111746.

Martina K. Linnenluecke, Mauricio Marrone, and Abhay K. Singh. Conducting systematic literature reviews and bibliometric analyses. *Australian Journal of Management*, 45(2): 175–194, 2020. doi: 10.1177/0312896219877678.

David Moher, Alessandro Liberati, Jennifer Tetzlaff, Douglas G. Altman, and The PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Medicine*, 6(7):e1000097, 2009. doi: 10.1371/journal.pmed.1000097.

Saeed Nosratabadi, Amir Mosavi, Puhong Duan, Pedram Ghamisi, Ferdinand Filip, Shahab S. Band, Uwe Reuter, Joao Gama, and Amir Gandomi. Data science in economics: Comprehensive review of advanced machine learning and deep learning methods. *Mathematics*, 8(10):1799, 2020. doi: 10.3390/math8101799.

Ahmet Murat Ozbayoglu, Mehmet Ugur Gudelek, and Omer Berat Sezer. Deep learning for financial applications: A survey. *Applied Soft Computing*, 93:106384, 2020. doi: 10.1016/j.asoc.2020.106384.

Matthew J. Page, Joanne E. McKenzie, Patrick M. Bossuyt, Isabelle Boutron, Tammy C. Hoffmann, Cynthia D. Mulrow, Larissa Shamseer, Jennifer M. Tetzlaff, Elie A. Akl, Sue E. Brennan, Roger Chou, Julie Glanville, Jeremy M. Grimshaw, Asbjørn Hróbjartsson, Manoj M. Lalu, Tianjing Li, Elizabeth W. Loder, Evan Mayo-Wilson, Steve McDonald, Luke A. McGuinness, Lesley A. Stewart, James Thomas, Andrea C. Tricco, Vivian A. Welch, Penny Whiting, and David Moher. The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, 372:n71, 2021. doi: 10.1136/bmj.n71.

Justin Paul and Alex Rialp Criado. The art of writing literature review: What do we know and what do we need to know? *International Business Review*, 29(4):101717, 2020. doi: 10.1016/j.ibusrev.2020.101717.

Mark Petticrew and Helen Roberts. *Systematic Reviews in the Social Sciences: A Practical Guide*. Blackwell Publishing, 2006. doi: 10.1002/9780470754887.

Melissa L. Rethlefsen, Shona Kirtley, Siw Waffenschmidt, Ana Patricia Ayala, David Moher, Matthew J. Page, and Jonathan B. Koffel. PRISMA-S: An extension to the PRISMA statement for reporting literature searches in systematic reviews. *Systematic Reviews*, 10 (1):39, 2021. doi: 10.1186/s13643-020-01542-z.

W. Scott Richardson, Mark C. Wilson, Jim Nishikawa, and Robert S. A. Hayward. The well-built clinical question: A key to evidence-based decisions. *ACP Journal Club*, 123(3): A12–A13, 1995.

Dmitry Scherbakov, Nina Hubig, Vinita Jansari, Alexander Bakumenko, and Leslie A. Lenert. The emergence of large language models as tools in literature reviews: A large language model–assisted systematic review. *Journal of the American Medical Informatics Association*, 32(6):1071–1086, 2025. doi: 10.1093/jamia/ocaf063.

Omer Berat Sezer, Mehmet Ugur Gudelek, and Ahmet Murat Ozbayoglu. Financial time series forecasting with deep learning: A systematic literature review: 2005–2019. *Applied Soft Computing*, 90:106181, 2020. doi: 10.1016/j.asoc.2020.106181.

Beverley J. Shea, Barnaby C. Reeves, George Wells, Micere Thuku, Candyce Hamel, Julian Moran, David Moher, Peter Tugwell, Vivian Welch, Elizabeth Kristjansson, and David A. Henry. AMSTAR 2: A critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. *BMJ*, 358:j4008, 2017. doi: 10.1136/bmj.j4008.

Hannah Snyder. Literature review as a research methodology: An overview and guidelines. *Journal of Business Research*, 104:333–339, 2019. doi: 10.1016/j.jbusres.2019.07.039.

Mathieu Templier and Guy Paré. A framework for guiding and evaluating literature reviews. *Communications of the Association for Information Systems*, 37:112–137, 2015. doi: 10.17705/1CAIS.03706.

David Tranfield, David Denyer, and Palminder Smart. Towards a methodology for developing evidence-informed management knowledge by means of systematic review. *British Journal of Management*, 14(3):207–222, 2003. doi: 10.1111/1467-8551.00375.

Rens van de Schoot, Jonathan de Bruin, Raoul Schram, Parisa Zahedi, Jan de Boer, Felix Weijdema, Bianca Kramer, Martijn Huijts, Maarten Hoogerwerf, Gerbrich Ferdinands, Albert Harkema, Joukje Willemsen, Yongchao Ma, Qixiang Fang, Sybren Hindriks, Lars Tummers, and Daniel L. Oberski. An open source machine learning framework for efficient and transparent systematic reviews. *Nature Machine Intelligence*, 3:125–133, 2021. doi: 10.1038/s42256-020-00287-7.

Nees Jan van Eck and Ludo Waltman. Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics*, 84(2):523–538, 2010. doi: 10.1007/s11192-009-0146-3.

Jan vom Brocke, Alexander Simons, Bjoern Niehaves, Kai Riemer, Ralf Plattfaut, and Anne Cleven. Reconstructing the giant: On the importance of rigour in documenting the literature search process. In *Proceedings of the 17th European Conference on Information Systems (ECIS 2009)*, pages 2206–2217, 2009.

Penny Whiting, Jelena Savović, Julian P. T. Higgins, Deborah M. Caldwell, Barnaby C. Reeves, Beverley Shea, Philippa Davies, Jos Kleijnen, and Rachel Churchill. ROBIS: A new tool to assess risk of bias in systematic reviews was developed. *Journal of Clinical Epidemiology*, 69:225–234, 2016. doi: 10.1016/j.jclinepi.2015.06.005.

Claes Wohlin. Guidelines for snowballing in systematic literature studies and a replication in software engineering. In *Proceedings of the 18th International Conference on Evaluation and Assessment in Software Engineering (EASE '14)*, pages 1–10, 2014. doi: 10.1145/2601248.2601268.

Ivan Zupic and Tomaž Čater. Bibliometric methods in management and organization. *Organizational Research Methods*, 18(3):429–472, 2015. doi: 10.1177/1094428114562629.