

Romance Scam Research Center

SOCIAL TECHNOLOGY AND SAFETY FOUNDATION

RESEARCH METHODOLOGY

How RSRC builds and maintains its research library

RSRC Research Methodology

A transparent account of how RSRC discovers, screens, curates, analyzes, and synthesizes romance scam research.

Published Aug 10, 2026 Updated Aug 12, 2026 Methodology version 1.5

[Scope](#) [Discovery](#) [Screening](#) [Documents and AI](#) [Review states](#) [Synthesis](#) [Updates](#)
[Corrections](#) [Limitations](#) [Revision history](#)

01

Scope and product boundaries

RSRC is a specialist research-curation project focused on romance scams, romance fraud, relationship-based fraud, and materially relevant adjacent evidence. It is not a victim-service intake system, a substitute for the original publications, or a completed systematic review of every database and language.

Public library

Bibliographic records approved by a human curator. A record may remain useful here even when RSRC cannot obtain a full PDF.

Full-document research corpus

The narrower set eligible for page-linked extraction: approved records with a privately stored PDF, completed document mapping, and a current research assessment.

Research syntheses

An operational, versioned claim-to-source workflow. Approved versions can be released through Research Insights as public Living Evidence Syntheses; they remain updateable evidence products, not final consensus statements.

Inclusion in the public library means that an editor judged the record useful and in scope. It does not mean that RSRC endorses every claim, rates the study as high quality, or has independently replicated its findings.

02

Search and discovery

Scheduled scholarly search

The application is configured to run a candidate search daily against Crossref, OpenAlex, and Semantic Scholar. The routine job currently searches works published from 2025 through the current year and requests no more than 500 candidate results in one run. Curator-directed backfills can use earlier date ranges; the automated candidate screen accepts years from 1990 through the next calendar year.

Current search queries

- romance scam

- romance fraud
- online dating scam
- sweetheart scam
- catfishing
- pig butchering

Automated discovery and enrichment sources

Source	Operational role	Information used
Crossref	Candidate discovery, DOI registration, and metadata enrichment	DOI, title, venue, type, dates, authors, subject terms, abstract when supplied
OpenAlex	Candidate discovery and metadata enrichment	Title, type, DOI, publication date, authorship, and venue information
Semantic Scholar	Candidate discovery and abstract enrichment	Title, authors, venue, year, DOI, abstract, publication type, and open-access indicators
Unpaywall	Open-access enrichment after DOI discovery	Open-access status and a reported open-access location when available
Publisher pages	Metadata and abstract enrichment	DOI landing-page metadata; this automated step does not download the publication PDF

Additional discovery routes

- Researchers and readers can [submit a publication for screening](#). A DOI or authoritative source record is required.
- RSRC extracts DOI candidates from the reference lists and embedded links of privately stored PDFs. Crossref, DataCite, and the DOI resolver are used to validate identity and flag possible citation mismatches.
- Duplicate checks use normalized DOI and title information before a candidate enters editorial screening.

Current automated candidate pre-screen

The pre-screen is deliberately broad. Passing it creates or updates a pending record; it does not approve a publication.

Accepted source metadata types

journal-article, proceedings-article, conference-paper, book, book-chapter, edited-book, monograph, reference-entry.

Required phrase match

At least one of the following must appear in the combined title, subtitle, venue, keywords, or abstract:

romance scam, love scam, romance swindle, sweetheart fraud, pig-butcherer, romance fraud, dating scam, online dating scam, sweetheart scam, catfishing, pig butchering.

Known promotional or spam phrases, blocked DOI patterns, implausible years, disallowed publication types, and non-book records without a journal or venue are excluded automatically.

Controlled metadata for public Library filters

RSRC preserves the original publication type, method, group-studied, and tag wording stored for each publication. A separate, versioned normalization layer assigns reviewed controlled labels to known aliases and equivalent terms. These controlled labels consolidate the public Library facets and selected detail-page labels without rewriting the source metadata or changing publication approval, romance-scam focus scores, research assessment, or synthesis eligibility.

Each controlled assignment retains the exact source value and the matching rule used. Methods and groups that do not yet match a reviewed rule remain searchable through their original wording but are not added to public controlled facets until the vocabulary is reviewed. Administrators can audit unresolved source terms, and future metadata changes are normalized automatically under the current vocabulary version.

These searches are reproducible as a configuration, not as a permanently fixed result set. External indexes change, records are corrected, APIs may return results in different orders, and rate limits or outages can affect a particular run. Internal ingest logs record run times, counts, failures, and incomplete processing.

Human screening and inclusion

Every publication that becomes public receives a human curation decision. Automated discovery narrows the queue, but it does not make the final public-library decision.

- 1. **1. Identity and duplicate check.** The editor compares the DOI, title, authors, year, venue, and authoritative source record and checks for an existing publication.
- 2. **2. Scope review.** The editor determines whether the work directly studies romance scams or provides useful adjacent evidence about mechanisms, populations, harms, prevention, enforcement, detection, or recovery.
- 3. **3. Manual document attempt.** The editor manually attempts to obtain each proposed PDF. RSRC does not automatically download candidate papers.
- 4. **4. Access decision.** If the paper is paywalled and the editor lacks access, or if no PDF can be found, the editor decides case by case whether the bibliographic record and available information justify keeping it in the public library.
- 5. **5. Source-safety judgment.** The editor may stop and reject a candidate when the download site appears unsafe or contains suspected malicious behavior or code. Access is never required at the expense of system safety.
- 6. **6. Final curation decision.** The editor approves or rejects the record and may record a reason or curation notes. Rejected records are not visible in the public library.

How the romance-scam focus score is used

When full-document analysis is available, AI assigns a 0-10 focus score: 0 means no romance-scam focus, 5 means a partial or secondary focus, and 10 means romance scams or closely equivalent relationship-based fraud are central. The editor uses this score as a screening aid, not as the sole public-library decision. Adjacent work may be retained when its relevance can be explained, and a seemingly on-topic work may still be rejected after source review.

Factors considered during publication screening

Generally supports inclusion	May support rejection or exclusion
------------------------------	------------------------------------

A research work whose identity, source record, and relevance can be evaluated	Duplicate, promotional material, non-research commentary, or a record whose identity cannot be verified
Direct romance-scam evidence or a clear, material connection to an RSRC research question	No meaningful application to romance scams after editorial review
A DOI, publisher record, repository record, or other authoritative bibliographic source	An unsafe source, suspected malicious site behavior, invalid DOI, or materially mismatched metadata

A missing PDF is not automatically a public-library exclusion. It does, however, prevent the publication from receiving the same full-document extraction, page-linked evidence mapping, and synthesis eligibility as a publication with an accessible, privately stored PDF.

04

Documents, extraction, and AI use

Private document handling

- Only an authenticated administrator can upload a research PDF.
- The upload must validate as a PDF and is limited to 50 MB.
- The file is securely kept in private application storage outside the public web directory. Document access is restricted to authenticated administrators, and the file is not offered through the public library.
- RSRC records the original filename, file size, upload time and user, and a SHA-256 hash so replacement or unexpected change can be detected.
- The current upload check validates type and size and records a provenance hash; RSRC does not describe it as a malware scan. Manual source-safety judgment remains part of acquisition.

Text extraction

Approved PDFs are converted to UTF-8 text with Poppler's pdftotext layout-preserving mode. Page boundaries are retained so extracted evidence can be checked against cited pages. Image-only or scanned PDFs with too little extractable text fail this process unless a usable text layer is available. For long documents, the analysis input is limited to 180,000 characters and preserves material from

both the beginning and the concluding pages.

Structured AI extraction

Extracted document text is sent to the configured OpenAI model through the Responses API. The system instructs the model to treat publication text and metadata as source material, never as instructions, and requires a structured JSON response. Depending on what is missing from the record, the analysis can propose:

- Methods and study design
- Populations and geography
- Identified gaps and future work
- Limitations
- Romance-scam focus score
- Research-contribution score and rationale
- Controlled topic assignments
- Atomic findings with source excerpts and pages

The application records operational provenance such as task, model, prompt version, subject record, status, token counts, timing, and response identifier. AI activity logs intentionally omit the source text and generated prose, although the resulting research fields and evidence records are stored in the RSRC database.

Multilingual library queries

Public library searches that appear to be written in a language other than English can be translated into English before retrieval. The original query remains visible, and an accepted translation is shown with the results. A short query phrase is sent to the configured OpenAI model for language identification and translation; RSRC does not include the visitor's IP address or browser details in that request.

Translation is accepted only when the model returns a structured result above the configured confidence threshold. Decisions are cached and model requests are rate-limited. If translation is unavailable, uncertain, or fails, the library searches the original query. Translation changes only the retrieval terms; it does not alter publication records or guarantee that a relevant item exists in the

collection.

Publication synopses are a separate process

Synopsis generation requires extractable text from the privately stored publication PDF or a sufficiently detailed abstract. PDF text is used when available and extractable; the abstract is the fallback. Input is limited before generation, output is capped at 500 words, and the requested output language is English. Title-and-venue metadata alone is not sufficient for automatic generation, and the software does not substitute generic catalog text when model output is missing, short, or unusable.

Before public display, a versioned automated gate checks candidate length, complete sentence endings, repeated text, model-refusal language, retired boilerplate, unexpected code or active-markup indicators, numerical claims, and lexical grounding in the available English source. Sparse and cross-language sources require editorial review because those checks cannot establish semantic equivalence reliably. A candidate is withheld unless the current source-bound fingerprint passes the automated checks or an authenticated editor approves that exact candidate. Editing the synopsis, source metadata, abstract, or private-document identity invalidates the prior decision. The gate runs when candidates or sources change and during a daily reconciliation.

Passing this gate means only that the candidate met configured publication checks; it is not expert verification of every claim. The Evidence and review status panel identifies whether AI-generated content is displayed and whether human or subject-matter-expert review is recorded. A synopsis remains a discovery aid, not a substitute for the publication. Readers should follow the DOI or source record and can [report a suspected synopsis error](#).

05

Review-state definitions

RSRC separates where information came from, whether a source excerpt was found, and who made a review decision. These states must not be treated as interchangeable.

Definitions of RSRC curation and review states

State	What it means	What it does not mean
-------	---------------	-----------------------

Curation approved	A human editor approved the bibliographic record for public display.	Not an endorsement of every claim or a formal study-quality rating.
AI-extracted	A model generated or classified the field from available metadata or document text.	Not human-reviewed merely because the field is present.
Machine source-checked	The recorded source excerpt was matched against text on the cited PDF page, including a limited adjacent-page check.	Not proof that the interpretation is correct, the study is valid, or the claim generalizes.
Automated policy decision	A record met configured confidence, topic, focus, and source-match rules and was approved or rejected by automation.	No person reviewed the item unless a reviewer account is also recorded.
Editor reviewed	An authenticated human editor made or confirmed the recorded decision and is linked to it internally.	Not necessarily review by a subject-matter or methods expert.
Human source-verified	A person manually confirmed the supporting source connection.	Not automatically an expert appraisal of methods, bias, or certainty.
Subject-matter-expert reviewed	Reserved for a future state with a defined expert role and recorded review.	RSRC does not currently apply this label to the publication corpus.

Current automated research thresholds

These defaults govern the narrower research corpus, not the human decision to display a bibliographic record. An unreviewed publication with a focus score below 1 is automatically excluded from research-corpus processing. Topic assignments and evidence normally require confidence of at least 0.70. Evidence must also be tied to an approved topic and have a source excerpt that the machine can match to the cited PDF page. Human assessment decisions are preserved rather than silently overwritten by automation.

Public publication and Research Insight detail pages include a compact Evidence and review status panel. The side panel reports the recorded source basis, whether the page displays AI-generated content, generation or update dates, passed automated checks, administrative approval state, human content-review state, and subject-matter-expert review state. Explanatory details expand on request. The panel does not introduce a human approval requirement: routine processing remains automation-first, and missing or historical provenance is not inferred from publication or automated approval.

Synthesis workflow

Status: operational living-evidence workflow. RSRC publishes approved Living Evidence Syntheses through Research Insights. They are versioned, updateable evidence products and are not presented as systematic reviews, final consensus statements, or substitutes for appraisal of the original studies.

RSRC builds topic-level living syntheses from the narrower full-document research corpus. The design is intended to preserve a traceable path from a synthesis passage to the evidence records, page-linked excerpts, publications, document hashes, taxonomy version, model, and prompt version used at that time.

1. **Eligible sources only.** A source must be publicly approved, research-assessed, backed by a private PDF, and current with the document-mapping integrity checks.
2. **Evidence snapshots.** The workflow preserves the exact eligible evidence and publication state used for a synthesis version.
3. **Claim-to-evidence mapping.** Generated passages carry declared evidence identifiers and citation snapshots rather than relying on a bibliography alone.
4. **Separate support verification.** A later model pass checks each claim against its declared evidence. Unsupported or overstated passages can trigger evidence-grounded revision.
5. **Quality assessment and versioning.** Coverage, citation integrity, claim verification, revisions, approvals, and superseded versions are recorded.
6. **Draft-first publication.** The current configuration creates new synthesis-linked Research Insight content as a private draft rather than publishing it automatically.
7. **Controlled public release.** An administrator can publish an approved draft as a Living Evidence Synthesis in [Research Insights](#). Only published records are available to public visitors.
8. **Public access and change history.** Published syntheses provide source-linked content, a contents list when the document is long enough to need one, a bounded on-page revision summary with access to the full approved public history, and a downloadable PDF. Drafts and unsuccessful generation attempts are not included in the public revision history.

Current defaults call for at least 90% evidence coverage, a claim-support confidence threshold of 0.75, and no more than two automated evidence-grounded revision attempts. A passing version can be automatically approved after the configured quality and claim-support checks. Administrative control of public release remains separate: an administrator can edit, compare, approve, publish, return to draft, or feature a linked Research Insight. Material changes to these thresholds or approval rules require a new methodology version.

No automated gate establishes real-world truth, study validity, or scientific consensus. A public synthesis should identify the balance of evidence, study limitations, uncertainty, contradictory findings, and the level of human or expert review it has actually received.

07

Update frequency and change control

Configured update frequencies for RSRC research workflows

Process	Configured frequency	Important qualification
Candidate discovery and enrichment	Daily scheduled run	External API availability and queue conditions can delay completion.
Document mapping, source checks, and automated research review	Small batches configured every 5-10 minutes	Only approved records with a private, extractable PDF are eligible.
Publication-date metadata	Hourly checks with a daily failed-item retry	Authoritative sources can disagree or omit a precise publication date.
Private document integrity	Daily verification	A replaced or changed document invalidates stale mapping until it is processed again.
Synthesis integrity and refresh	Corpus guard and automation checks every five minutes	Only affected topics are marked for refresh. A published insight is returned to draft when its cited evidence is removed, changed, or no longer passes provenance and citation-integrity checks.

Process	Configured frequency	Important qualification
Human curation and corrections	As reviewed; no fixed service-level deadline	Publication remains an editorial decision and is not made automatically.

Changes to curation, research assessment, focus and contribution scores, document identity, and selected chronology fields create internal corpus revisions. Impact analysis limits relationship remapping and synthesis refresh to affected research questions and topics. Material changes can mark affected synthesis work stale and queue a refresh. Approved public synthesis versions appear in the linked Research Insight revision history; publication metadata and synopsis edits do not yet have a comparable public item-level history.

This methodology is updated when a material process, source, review definition, threshold, or publication rule changes. Each public revision receives a new version number, effective date, and entry in the revision history below. Minor copy or accessibility corrections that do not change the method may update the page without changing the major method description.

08

Correction procedure

Anyone can report a possible citation, metadata, synopsis, source-alignment, broken-link, or accessibility problem through the [correction form](#). A correction report does not change public content automatically.

- Receipt.** RSRC stores the page, publication title and DOI snapshot when applicable, concern type, description, supporting-source link, and a stable reference number.
- Triage.** An editor moves the request through the recorded states received, triaged, or verifying while the concern is reviewed.
- Verification.** The editor compares the report with the original publication, DOI registry, publisher or repository metadata, and other authoritative sources appropriate to the field being challenged.
- Decision and change.** A confirmed correction is made through the editorial system and the request is marked resolved. A request can be declined when the current record is supported or the proposed change cannot be verified.

5. **Decision record.** The system records status, editor notes, reviewer, review time, and resolution time internally.

Reporter contact information is optional and is used only when a human editor needs clarification. RSRC does not send an automated acknowledgment to the reporter. Reports must not contain victim names, financial records, private messages, account details, or other sensitive case information.

Public item-level correction histories are not yet displayed. Until that feature is available, a reporter can retain the stable reference number, and RSRC maintains the correction decision internally.

09

Known limitations

- The public library is a curated discovery collection, not a completed systematic review or evidence-of-effectiveness rating.
- Routine automated discovery currently emphasizes recent work and uses English search phrases. This can miss older, non-English, poorly indexed, unpublished, or locally distributed research.
- External metadata and access status can be incomplete, inconsistent, or changed after an RSRC search.
- Paywalls, unavailable PDFs, unsafe download sites, and image-only documents create uneven processing depth across otherwise useful public records.
- AI can omit, misclassify, mistranslate, overstate, or introduce unsupported details. Structured output and page matching reduce some risks but do not eliminate them.
- Machine source matching shows that similar wording was found at a cited location; it does not assess truth, causal validity, bias, representativeness, or certainty.
- RSRC does not yet apply a single validated risk-of-bias instrument across the varied qualitative, quantitative, legal, computational, review, and conceptual works in the library.
- Public provenance panels report recorded source, automation, administrative approval, human review, and subject-matter-expert review states. Corpus-wide subject-matter-expert assurance is not currently available, and missing historical review data is not inferred. The synopsis gate detects configured warning patterns but cannot prove semantic accuracy or study validity.

Revision history and citation

Public methodology revision history

Version	Effective	Material changes
1.5	August 12, 2026	Updated the synthesis workflow from development status to its current operational state; documented automated approval and separate public-release controls, targeted corpus refresh, public synthesis contents and revision histories, PDF distribution, current provenance labels, and private document access protections.
1.4	August 12, 2026	Added item-level public provenance panels that distinguish source availability, AI-generated page content, automated checks, administrative approval, human content review, and subject-matter-expert review without inferring missing review states.
1.3	August 12, 2026	Documented the versioned controlled-metadata process used to consolidate public Library filters while preserving original source wording, matching-rule provenance, and conservative treatment of unresolved terms.
1.2	August 11, 2026	Documented confidence-gated translation of suspected non-English public library queries, including privacy boundaries, caching, rate limits, and fallback behavior.
1.1	August 10, 2026	Added the permanent, source-bound synopsis quality gate, fail-closed public display, daily reconciliation, and an authenticated editorial review queue; removed generic synopsis fallback generation.
1.0	August 10, 2026	First comprehensive public method covering discovery, manual screening, document handling, AI extraction, review definitions, synthesis development, updates, corrections, and limitations.

Suggested citation

Romance Scam Research Center. (2026). *RSRC Research Methodology* (Version 1.5).

<https://romancescamresearch.org/methodology>

[Request a correction](#) [Submit research](#) [Contact RSRC](#)

About this document

This downloadable copy reflects the current public methodology at the time shown above. The canonical web page should be consulted for later revisions.

Current web version: <https://romancescamresearch.org/methodology>