Research Integrity
Analyzer

NAVIGATION
Dashboard
New Analysis
History
Methodology
About Us
Support Us

QUICK STATS
Papers 7
Avg Risk 44

EN
Source Code

Open Source & Free

Methodology

Our transparent methodology for analyzing research integrity

Rule-Based COI Detection Algorithm

Research Integrity Analyzer uses an algorithm based on explicit rules to evaluate conflicts of interest (COI) in scientific papers. The methodology combines semantic extraction through language models with fixed rules and thresholds, following international standards (ICMJE, COPE, WAME, CONSORT, PRISMA, DOAJ).

Objective: Produce a structured estimation of COI risk and editorial credibility using only the paper's text as source, with reproducible and explainable results.

Detection Algorithm Flow

1 Ingestion and Preprocessing

The system extracts and cleans the scientific document content:

- Plain text: title, authors, affiliations, main sections
- Structure detection: Abstract, Introduction, Methods, Results, Discussion, Conclusions
- Key section identification: Conflict of Interest, Funding, Acknowledgements
- Cleaning artifacts and duplicate spaces

2 Semantic Extraction with AI

The language model identifies structured textual facts:

Identifies

- Author and institution names
- Potential funders
- Explicit/implicit COI declarations
- Fragments about funding and sponsors

Detects

- Language patterns (promotional vs critical)
- Presence/absence of limitations
- Companies, foundations, organizations
- Relationships between authors and sponsors

Output: A set of structured textual facts that feed the algorithm rules. The AI doesn't decide "by eye", it only extracts objective information.

3 Rule Application by Dimensions

From the extracted facts, scores 0-100 are calculated for each of the 5 dimensions. These scores don't depend on "intuitions", but on predefined rules based on best practices.

1. Disclosure & Funding Transparency

Transparencia de conflictos declarados y financiación

- → Sin sección COI ni funding en estudio sensible → score 75-90, nivel 'high'
- → COI declarado 'no conflicts' + funding claro → score 20-35, nivel 'low'
- → Funding presente pero sin mención COI → score 40-60, nivel 'medium'
- → Declaraciones vagas → score 60-75, nivel 'high'

2. Funding-Outcome Alignment

Relación entre financiación y resultados

- → Sponsor comercial + resultados muy positivos + sin crítica → score 60-85, 'high'
- → Sponsor público/académico + discusión equilibrada → score 20-40, 'low'
- → Sin sponsor identificable → score 40-55, 'medium'
- → Sponsor + resultados favorables + lenguaje promocional → score 70-90, 'high'

3. Author-Institution-Sponsor Network

Red autores-instituciones-financiadores

- → Varios autores empleados de empresa financiadora → score 70-90, 'high'
- → Afiliaciones académicas diversas sin vínculos comerciales → score 20-40, 'low'
- → Afiliaciones ausentes o genéricas → score 60-80, 'high'
- → Institución única = sponsor → score 55-75, 'high'

4. Journal / Editorial Integrity

Integridad editorial y riesgo de predatory journal

- → Señales de predatory journal → score 70-90, 'high'
- → Indicios de peer review, políticas éticas → score 20-40, 'low'
- → Información insuficiente sobre revista → score 40-60, 'medium'

5. Textual Bias & Reporting Quality

Sesgos de lenguaje y calidad de reporte

- → Lenguaje promocional + sin limitaciones → score 60-80, 'high'
- → Lenguaje sobrio + limitaciones honestas → score 20-40, 'low'
- → Falta de transparencia metodológica (CONSORT/PRISMA) → +10-20 puntos
- → Autocitación excesiva + tono promocional → score 55-75, 'high'

4 Global Score Calculation

The global score and risk level are calculated:

```
overall_score = (dim1 + dim2 + dim3 + dim4 + dim5) / 5
```

Simple average of the 5 dimensions (equal weights: 20% each)

0-33	34-66	67-100
LOW	MEDIUM	HIGH

Consistency: The score → level conversion is fixed and not dynamically modified. If information is missing in a dimension, it's marked in the middle range (45-55) and uncertainty is made explicit.

5 Report Generation

The model writes a report in natural language with fixed structure:

Fixed Structure
Always the same sections

Stable Labels
Consistent levels (low/medium/high)

Explicit Rules
Direct references to applied rules

Reproducibility: The same paper analyzed multiple times produces practically identical results, ensuring stability and reliability.

Privacy & Data Security

What We Store:

- Analysis results and metadata
- Paper titles and author information
- Detected conflicts and recommendations
- URLs of analyzed papers (not file content)

What We Don't Store:

- Full PDF file contents permanently
- Personal user information
- Analysis IP addresses or tracking data
- Any proprietary research data

Role of the AI Model

The AI does not decide risk levels by intuition. Its specific function is:

1. Extract
Information from text: sections, fragments, company mentions, COI declarations

2. Map
Findings to predefined rules and apply thresholds to convert to scores

3. Draft
Report with fixed structure explaining evidence and applied rules

Design focused on: Stability and reproducibility (same paper → identical results) + Transparency (report explains what triggered each score)

Current Algorithm Limitations

- Based solely on the available paper's textual content
- No access to external COI forms, trial registries, or external databases (DOAJ, COPE, ICMJE)
- Predatory journal detection is approximate and based on internal textual signals
- The algorithm indicates COI risk, does not prove its legal existence
- Tool for critical reading and activism against scientific disinformation, not a court of truth

📚 Complete Research Available

Access the complete research document that supports the algorithm methodology, including COI types, international guidelines, textual indicators, and scientific bibliography.

Download Research (PDF) | Technical Algorithm (PDF)

- Methodology validated by international standards
- Complete scientific bibliography
- Total transparency