



CONSEJO INTERNACIONAL
PARA LA LUCHA CONTRA LA CORRUPCIÓN

**LARGE LANGUAGE MODELS IN THE OVERSIGHT OF PUBLIC PROCUREMENT:
WHAT THEY DETECT AND WHAT IS KNOWN ABOUT THE RESULTS. THE
EXPERIENCE OF BRAZIL, KOREA, CHILE, COLOMBIA AND THE EUROPEAN
UNION**

Dzianis Limarau, Ignat Krukovskiy

Consejo Internacional para la Lucha contra la Corrupción, asociación civil
Buenos Aires, 2026
Version 1

DOI: 10.5281/zenodo.23241236

Licensed under CC BY 4.0

<https://cilclegal.org/>

GRNTI 10.07.45, 06.56.31, 20.53.19; UNESCO 5605.99, 5909.01, 1203.04; OECD FOS 5.05 Law, 5.02
Economics and Business, 1.02 Computer and Information Sciences

Recommended citation: Limarau, D., Krukovskiy, I. (2026). Large language models in the oversight of public procurement: what they detect and what is known about the results. Report. Consejo Internacional para la Lucha contra la Corrupción. DOI 10.5281/zenodo.23241236. <https://cilclegal.org/>

Contents

Contents	2
Objective of the Study	3
Intended Readership	3
Executive Summary	3
1. Introduction	5
2. Automated Risk Detection in Procurement: from Red Flags to Language Models	5
3. Data and Method	6
3.1. Selection of Tools	6
3.2. Tool Card	7
3.3. Mapping to the Typology	7
3.4. Performance Indicators	7
3.5. Terminology	7
4. Large Language Model Tools in Five Jurisdictions	7
4.1. Brazil: Alice 360 of the Federal Court of Accounts	8
4.2. Brazil: VigIA of the Court of Accounts of the State of Santa Catarina and PRUMe AI of the Court of Accounts of the State of Amazonas	10
4.3. Korea: Notice Monitoring by the Public Procurement Service	10
4.4. Korea: Review of Contract Documents in Seocho-gu	11
4.5. Chile: Document Analysis by ChileCompra	11
4.6. Italy: Notice Monitoring by the National Anti-Corruption Authority	12
4.7. European Union: the European Anti-Fraud Office and Arachne+	12
4.8. Colombia and Other Cases without a Confirmed Language Model	12
5. Which Mechanisms the Language Models Cover	13
5.1. Functions and Mechanisms	13
5.2. Coverage Matrix	14
5.3. Coincidence with the Boundary of the External Participant's Visibility	16
6. What Is Known about the Results	17
7. Signs beyond the Typology	18
8. Limitations and an Open Question about Red Flags	18
9. Conclusions	19
Sources	20
Appendix. Tools without Language Models in the Same Jurisdictions	23

Objective of the Study

The public authorities of Brazil, the Republic of Korea, Chile, Colombia and the European Union have begun to apply large language models to the review of public procurement. The principal questions of the study: which tools built on language models these authorities apply, what such tools check, which mechanisms of distortion of a competitive procedure they are able to detect, and what has been published about the results of their work.

In this study an oversight tool means a software system by which a public authority, whether a financial control body, a procurement agency, a competition, supervisory or law enforcement authority, automatically detects possible violations in procurement procedures. A large language model (hereafter also LLM) means an artificial intelligence model that processes text in natural language, and an LLM tool means an oversight tool in which at least one function is performed by such a model. Tools of another kind, built on rules, register cross-checks, statistical thresholds and classical machine learning, are described in the appendix for comparison.

The tasks of the study:

1. To compile an inventory of LLM oversight tools for procurement in five jurisdictions as at October 2026. For each tool, to describe its functions, data, alert handling procedure and stage of implementation. The source is the documents of the authorities that apply the tools themselves.
2. To map the functions of each tool to the 30 mechanisms of the typology of the Consejo Internacional para la Lucha contra la Corrupción (hereafter CILC; Limarau, 2026). To establish which stages of the procurement cycle and which clusters of mechanisms the tools cover, and which remain outside their field of view.
3. To calculate, from published data, four performance indicators: coverage, the share of alerts reviewed by a person, the share confirmed and the share that led to measures. To state for which tools these indicators cannot be calculated.
4. To identify the signs that the tools look for and that are absent from the typology, as candidates for new mechanisms.

Intended Readership

The study is addressed to oversight bodies and procurement agencies that are introducing, or planning to introduce, language models for the review of procurement; to researchers of corruption risk who work with the data of electronic procurement systems; and to organisations that assess tools for the governance of artificial intelligence, among them the UN Independent International Scientific Panel on AI, whose Preliminary Report of 1 July 2026 notes that such tools rarely measure their real effectiveness.

No knowledge of the procurement law of individual countries is required of the reader. Each system is described from the open documents of the authority that applies it.

Executive Summary

Language models came to procurement oversight later than rules and statistics. By October 2026 they had not become the principal tool of review in any of the five jurisdictions. The inventory comprises six systems

in which the use of a language model is confirmed by the authority's documents, and two announced projects for which either the technology (Italy) or the procurement function (the European Anti-Fraud Office) is unconfirmed. Of the six confirmed systems, two are in operation, and for only one has the authority published a decision with results.

Data and method. The study rests on the documents of the authorities that apply the tools: the decisions and technical articles of the Federal Court of Accounts of Brazil and of the state courts of accounts, the press releases and tender documentation of the Public Procurement Service of Korea, the publications of the Chilean Public Procurement Directorate (ChileCompra) and the award record in the Chilean procurement system Mercado Público, the directive of the Italian National Anti-Corruption Authority, and the reports of the European Commission and the European Anti-Fraud Office. For each tool a card of six fields has been completed. The functions of the tools have been mapped to the 30 mechanisms of the CILC typology on three values: detects directly, detects indirectly, does not detect. Performance has been calculated from published figures with the denominator stated.

Principal findings. The only LLM tool for which the authority has published a decision with results is Alice 360 of the Federal Court of Accounts of Brazil. The language model in it does not read procurement documentation in search of violations; it performs a preparatory task: it recognises the subject matter of a procurement from imprecise descriptions and cleans the data, after which prices are compared by statistical methods. In the monitoring cycle of 2025-2026 the court sent requests to contracting authorities on 49 procurements, of which 15 were initiated by Alice 360 alerts; in two cases the contracting authority, following the request, reduced the estimated value of the procurement or withdrew it, and the court counts such a reduction as a financial effect of oversight.

The Public Procurement Service of Korea is building a system in which a language model checks notices against 23 legal points; until its launch, notices are checked manually and selectively. In the first quarter of 2026, 15,261 notices were checked and correction requests were sent on 797 of them, the most frequent grounds being breach of the publication period (437 cases) and excessive experience requirements together with regional restrictions (82 cases).

In January 2026 ChileCompra concluded a contract for the analysis of documents by a language model against 20 rules. The system is at the contract performance stage, with launch announced for the fourth quarter of 2026. The Italian National Anti-Corruption Authority and the European Anti-Fraud Office are conducting experiments without published results. The VigIA system of the Court of Accounts of the State of Santa Catarina has, since April 2024, run auditors' questions over the text of every notice through the GPT-4o model; for six months of 2024 the court reported 215 notices on which measures were taken. The PRUMe AI platform of the Court of Accounts of the State of Amazonas remains a research prototype.

Mechanisms covered. The two operating tools cover five of the 30 mechanisms of the typology: four at the documentation and publication stages, where the language model reads the notice (VigIA), and one at the price justification stage, where inflation of the initial price is detected by a statistical algorithm on data prepared by the model (Alice 360). Counting the systems under construction in Korea, Chile and Seocho-gu, potential coverage reaches 10 mechanisms at five of the ten stages: price justification, documentation, publication, bid submission and performance. No operating tool checks bid submission, bid evaluation, review, performance or payment; bid submission is touched only by the ChileCompra tool under construction, contract performance only by the Seocho-gu system under construction. The cluster of mechanisms "neutralisation of control", for which the share of confirmed complaints in the CILC dataset

is higher than for the other clusters, remains outside the field of view of the operating tools; of those under construction, only the check of mandatory information in the notice by the Korean system touches it. The language model reads the same documents as an external participant in the tender, and at the same time, before bids are submitted, so the boundary of its visibility coincides with the boundary of the participant's visibility in the typology.

Conclusions. No authority publishes the share of well-founded alerts by individual function of the language model. The only published shares relate to tools without a language model and range from 3.4 to 23.5 per cent. The gain from the language model over rules is described only in the TCU decision: the model recognises the subject matter of a procurement from imprecise descriptions, which rules cannot do. The signs that the tools of the five jurisdictions look for, including the tools without a language model in the appendix, yield four candidates for new mechanisms of the typology. To monitor the quality of a language model and of its application, an authority must collect at least four quantities for each function: the number of alerts issued, the number reviewed by a person, the number confirmed and the number that led to measures. Without these data the effectiveness of the language model is indistinguishable from the effectiveness of the rules to which it has been added.

1. Introduction

For two decades automated risk detection in public procurement developed as a system of red flags. A red flag is an observable feature of a procedure that is statistically associated with one violation or another: a single bidder, a short bid submission period or a minimal reduction from the initial contract price. The operating systems are built on this basis: the ALICE robot of the Office of the Comptroller General of Brazil since 2015, the BRIAS bid-rigging indicator analysis system of the Korea Fair Trade Commission since 2006, and the Arachne risk assessment tool of the European Commission since 2013.

Large language models made it possible to read unstructured documents: notices, terms of reference, records. In them the model can find what rules over data fields do not see, for example a requirement drafted for a single supplier. In 2024-2026 the authorities of five jurisdictions began to introduce such functions. The study establishes what exactly they have introduced, what these functions check and what is known about their results from the documents of the authorities themselves.

The CILC methodology considers the 30 most typical mechanisms of distortion of a procurement procedure, distributed across ten stages of the procurement cycle (Limarau, 2026). The typology separates the mechanism, that is the manner of distortion, from the indicator, that is the observable feature of the procedure. For each mechanism it states the observation position: who can notice it, an external participant in the tender, an audit body or an insider. Section 7 of the typology sets the task of matching the catalogue of mechanisms against the fields of open data and establishing which mechanisms lend themselves to automated detection. The present study performs that task for tools built on language models.

The second source of the study is the open CILC dataset on complaints by external reporters about procurement procedures and on the outcomes of their consideration (Limarau, 2026; Krukovskiy, 2026). For each complaint it records whether the complaint was considered on the merits, whether it was found well founded and whether it led to measures. The same four quantities are applied here to the alerts of the tools (section 3.4).

2. Automated Risk Detection in Procurement: from Red Flags to Language Models

The red flag approach has been systematised in the guidance of the OECD and the World Bank. A study for the European Commission tested it on 192 cases in eight states (PwC, Ecorys, 2013). A composite corruption risk index has been calculated on more than 2.8 million contracts from the TED database of European notices (Fazekas, Toth, 2016). A separate branch, the economic screening of bid rigging, detects cover bids from the structure of prices (Porter, Zona, 1993; Imhof, Karagök, Rutz, 2018). A 2022 review by the OECD Competition Committee sums up: single screens do not work in most cases, a set of screens works better, and a flag by itself does not prove a violation (OECD, 2022).

Machine learning on these same indicators is described in academic work. A random forest on Italian data has been tested against investigations and convictions (Decarolis, Giorgiantonio, 2020). Classifiers on screens have been tested on Swiss and Japanese tenders (Huber, Imhof, 2019). In state systems machine learning operates at the Spanish National Commission on Markets and Competition (the BRAVA algorithm) and at the Office of the Comptroller General of the Republic of Colombia (the predictive model of 2025).

The OECD review “Governing with Artificial Intelligence” describes around 200 cases of the application of artificial intelligence in public administration, including in procurement, and singles out ALICE in Brazil and the Observatorio ChileCompra (OECD, 2025). The 2026 review by the Open Contracting Partnership notes that claims about the effect of such systems have not been independently verified and that many projects never leave the pilot stage (OCP, 2026). Neither review separates the functions performed by a language model from the functions performed by rules. That separation is the subject of the present study.

3. Data and Method

3.1. Selection of Tools

The inventory includes systems that are applied by a public authority of one of the five jurisdictions (a federal or national authority, a body of the European Union, or a subnational authority where the tool is documented), that are intended for the oversight of procurement procedures, and in which at least one function is performed by a language model. The inventory is divided into two groups. The first comprises six systems for which both the language model and the procurement function are confirmed by the authority’s documents: a decision, an article in the authority’s own publication, tender documentation or a statement by the authority reported by the press. The second comprises two announced projects for which one of the two conditions is unconfirmed: the Italian National Anti-Corruption Authority reports experiments with artificial intelligence without naming the technology, and the European Anti-Fraud Office reports a generative artificial intelligence pilot without tying it to the review of procurement. The second group is described for completeness and is not included in the calculations of sections 5 and 6. Tools in which the language model serves only as an assistant to staff and has no procurement function (ChatTCU in Brazil, LIA at the Office of the Comptroller General of Brazil, “Qu’en dit la Cour” of the Court of Accounts of France) are not included in the inventory. They are mentioned in section 4 for the purpose of delimitation.

The search was conducted in October 2026. The documents of the authorities (decisions, annual reports, technical articles, tender documentation, press releases), the documents of the OECD, the European Court

of Auditors and the Open Contracting Partnership, and the specialist press in Portuguese, Korean, Spanish, Italian and English were examined. All statements about the functions of the tools rest on the documents of the authorities. Statements from secondary sources are marked.

Section 7 goes beyond the inventory of language models: signs not described in the typology have been collected across all the tools of the five jurisdictions, including the tools without a language model in the appendix. This is a supplementary result of the study, and it rests on a wider range of systems than sections 5 and 6.

3.2. Tool Card

For each tool six fields have been completed: authority and year; data and technique, stating which function is performed by the language model and which by rules or statistics; the list of signs the tool looks for; the alert handling procedure; the published performance indicators; independent assessments. The cards are given in section 4, and for the tools without a language model in the appendix.

3.3. Mapping to the Typology

Each function of a tool has been mapped to the mechanisms of the CILC typology. A function detects a mechanism directly if it is aimed at the trace of that very mechanism: for example, checking qualification requirements for excessiveness is aimed at mechanism SPC-02. A function detects a mechanism indirectly if the indicator it finds, as described in the typology, also relates to that mechanism and not only to the one at which the function is aimed: for example, an artificially shortened submission period (PUB-01) may turn up among the answers to the auditors' questions on a notice, although the questions were drafted to check the requirements for participants. An indirect link is not assigned where the indicator could only coincide with the trace of a mechanism under additional assumptions: thus a price above the market price is not treated as a trace of understatement of the initial price in the expectation of contract amendments (BUD-02) or of cover bids (BID-01), since the typology describes these mechanisms through other indicators. The result is brought together in a "mechanism x tool" matrix (Table 3).

3.4. Performance Indicators

The performance of a tool is assessed by four quantities: coverage, that is the share of procedures that pass through the tool; the share reviewed, that is the share of alerts that a person has reviewed; the share confirmed, that is the share of alerts that the reviewer found well founded; and the share with a consequence, that is the share of confirmed alerts that were followed by a correction, suspension, cancellation or sanction. With these same quantities CILC describes the outcomes of the consideration of complaints by external reporters in its dataset. For each tool it is stated which of the four quantities can be calculated from published data. Indicators that do not match these definitions (the number of requests to contracting authorities, the number of notices with measures, the F1 accuracy of data extraction) are given separately and are not equated with the four quantities. All figures are self-reported. No independent assessment of accuracy has been found for any LLM tool.

3.5. Terminology

The terms follow the CILC typology: mechanism, risk indicator, observation position (external participant, audit body, insider). A sign of restricted competition means an indicator that an external participant can see in the documentation. An alert is a message that the tool issues to a person about one procedure. The

confirmation rate is the share of complaints found well founded after consideration on the merits (Krukovskiy, 2026). Here it is applied to alerts.

4. Large Language Model Tools in Five Jurisdictions

Table 1 brings the inventory together: six systems with a confirmed language model and two announced projects without such confirmation. A card for each tool follows.

Table 1. LLM oversight tools for procurement, October 2026

Jurisdiction	Tool and authority	Function of the language model	Confirmation	Stage	Published results
Brazil	Alice 360, Federal Court of Accounts (TCU)	Recognition of the subject matter of a procurement from imprecise descriptions, data cleaning before price comparison	Court decision	Operation, 2025-2026 cycle	Acórdão 1145/2026: 15 of 49 requests to contracting authorities, 2 with a reduction in the value of the procurement
Brazil	VigIA, Court of Accounts of the State of Santa Catarina (TCE-SC)	Answers to auditors' questions on the text of a notice before publication (GPT-4o)	Article in the court's journal	Operation since April 2024	7,711 notices analysed, 215 with measures, April to October 2024
Brazil	PRUMe AI, Court of Accounts of the State of Amazonas (TCE-AM)	Extraction of fields and clauses from contracts and notices	Academic article by court staff	Prototype	F1 0.86 on an annotated subsample of 20 documents
Korea	Notice monitoring, Public Procurement Service (PPS)	Checking of notices against 23 legal points and for consistency	Tender documentation, press release	Development, launch announced for the second half of 2026	Before launch: manual selective checking, 797 correction requests from 15,261 notices in the first quarter of 2026
Korea	Review of contract documents, Seocho-gu District Office (Seoul)	Checking for completeness, consistency and compliance with rules	Statement by the district office in the press	Development, deadline September 2026	none
Chile	Document analysis, ChileCompra	20 rules over unstructured documents; triage of confidential complaints	Tender documentation, ChileCompra publications	Contract performance, launch announced for the fourth quarter of 2026	none
Italy	Notice monitoring, National Anti-Corruption Authority (ANAC)	Detection of anti-competitive conditions in notices	Unconfirmed: technology not named	Experiment (2026 directive)	none
EU	GETI with generative AI, European Anti-Fraud Office (OLAF)	Integration of risk indicators into a generative AI tool	Unconfirmed: procurement function not named	Pilot, 2025	none

4.1. Brazil: Alice 360 of the Federal Court of Accounts

Authority and year. The Federal Court of Accounts of Brazil (Tribunal de Contas da União, hereafter TCU), its digital audit and procurement audit units. Alice 360 continues the line of the ALICE tool, whose code TCU received from the Office of the Comptroller General in 2016. The version built on a language model was introduced in the cycle of continuous monitoring of federal procurement from June 2025 to March 2026. The outcome of the cycle was approved by the decision of the TCU plenary Acórdão 1145/2026 of 6 May 2026 (case TC 009.250/2025-5).

Data and technique. The first version of the tool was a query screen. An auditor entered a description of the subject matter, the unit of supply and a statistical method (interquartile range, z-score and others), the system searched the database of concluded contracts and returned deviations from the market price (point 78.1.1 of the decision). After access was obtained to language model platforms, “especially Claude”, the tool, in the words of the decision, “began to use AI really intensively” for the recognition and analysis of procurement subject matter. It now generates daily reports on procurements and items with possible price inflation (point 78.1.2). The language model helps “to clean and process the data and incorrect information” and recognises the subject matter “even with imprecise descriptions” (points 75.2 and 11 of the rapporteur’s opinion). The statistical part: the current version uses a trimmed mean, the Alice 360 panel uses the median after removal of outliers (points 78.1.3, 78.1.5). The data are taken from the federal procurement system Compras.gov.br and the National Public Procurement Portal (PNCP). Reading of notices and terms of reference by the language model is not described in the decision; nor do other TCU documents attribute such a function to Alice 360.

Signs. The price of the procurement item deviating from the market price. Extension to the links of bidders with officials and to bids with signs of infeasibility is planned (point 78.3.1).

Alert handling. The Alice 360 panel, including the panel on medicines, has been available to auditors since October 2025. An alert becomes the ground for a request to the contracting authority (diligência) and, on confirmation, for a binding order to correct. The Platform of Alerts, Risks and Typologies (PARTS) is to deliver alerts to the working tools of the auditors of TCU and of the partner state courts (point 78.4).

Results. In the 2025-2026 cycle the court sent requests to contracting authorities on 49 procurements. For 15 of them the ground was an alert from the Alice 360 panel. In two cases the contracting authority, following the request, reduced the estimated value of the procurement or withdrew it. TCU counts such a reduction as a financial effect of oversight (benefício financeiro). In two further cases a reduction is expected (points 78.1.4, 82). Across all 49 procurements the estimated value after the requests was republished at R\$105.6 million below the original. The amounts for the two procurements detected by Alice 360 are not given in the readable part of the decision. The only named example of a withdrawal: item 233 (lenalidomide) in electronic tender 90044/2025 of the Navy Command (point 64.1). The number of alerts and the share confirmed are not given in the decision. The decision notes that the development and testing of Alice 360 took time from the team that would otherwise have gone to the analysis of procurements (point 78.2). The unit’s proposal to recommend continuation of the project was not adopted by the plenary, which confined itself to an instruction to coordinate technological initiatives (point 9.2 of the operative part).

Data quality. The decision records 502 federal procurements with an estimated value above R\$100 million whose actual value proved to be at least one hundred times lower. Inconsistency of the data, in the words of the decision, “regularly led the team to false positives” (points 73, 74).

Independent assessments. There are no independent assessments. An article by TCU auditors on the PARTS platform (Santos et al., 2026) describes the transition from machine learning to generative artificial intelligence as a way of reducing the need for specialists in model training, but contains no accuracy indicators.

4.2. Brazil: VigIA of the Court of Accounts of the State of Santa Catarina and PRUMe AI of the Court of Accounts of the State of Amazonas

The VigIA system. Authority: the Court of Accounts of the State of Santa Catarina (Tribunal de Contas do Estado de Santa Catarina, hereafter TCE-SC). It has operated since April 2024 with the scientific support of the Federal University of Santa Catarina. The technique is described in an article by court staff in the court's own journal (Albuquerque, Wensing, Joppi, 2024, p. 156): audit questions are run automatically, without human involvement, over every incoming notice through integration with the generative model GPT-4o. The system receives notices before their publication through the e-Sfinge system, around 250 a day; the questions are registered by auditors. In the first two months, according to the same article, more than 2,000 procedures were processed and more than 10,000 questions executed. From 18 April to 8 October 2024, 33,507 notices were received, 7,711 analysed and 63,445 model answers generated. Measures were taken on 215 notices: correction, withdrawal, annulment or suspension, for a value of around R\$2 billion (TCE-SC, 2024). Coverage: 23 per cent of the notices received. The share of notices with measures among those analysed: 2.8 per cent; this is not the share of confirmed alerts, since the number of alerts and the number of answers reviewed are not published. The authors of the article note that the construction of accuracy metrics remains an unsolved task.

The PRUMe AI platform. Authority: the Court of Accounts of the State of Amazonas (Tribunal de Contas do Estado do Amazonas, hereafter TCE-AM). The platform is described in an academic article (Bezerra, Lopes, 2025) as an external control system built on language models. The model extracts data from documents by the RAG method (retrieval-augmented generation) and records where each value was taken from. The sample of the study: 150 real documents (notices, contracts and contract amendments, reports and opinions), of which 55 per cent are digital PDFs and 45 per cent scans. Triage time fell from 21.4 to 7.9 minutes per document, full analysis time from 39.2 to 17.8 minutes. Extraction accuracy (F1) was calculated on an expert-annotated subsample of 20 documents: 0.86 for contract fields and 0.82 for clauses. The specific model is not named in the article. Production operation is not confirmed. This is the only tool in the inventory with published accuracy metrics, and they relate to data extraction on 20 documents, not to the detection of violations.

4.3. Korea: Notice Monitoring by the Public Procurement Service

Authority and year. The Public Procurement Service (hereafter PPS). The artificial intelligence-based system for monitoring self-conducted procurement is being developed under a programme of the Ministry of the Interior and Safety. Production launch is announced for the second half of 2026 (ETNews, 2026). Legal basis: an amendment to the Electronic Procurement Act, which entered into force on 23 December 2025 and gave PPS the power to require contracting authorities to correct notices in the event of a manifest breach of the law.

Data and technique. The input is notices, specifications and terms of reference from 16 electronic procurement systems, around 2,500 a day. The language model checks the text for probable breaches of the State Contracts Act and the Local Government Contracts Act and assigns a risk class by type of breach. A

member of staff double-checks the alert, PPS sends a correction request to the contracting authority, and the outcome is returned to the training data. The composition of the check is disclosed in the open competition for the creation of the model (DACON, August to October 2026, organised by the Ministry of the Interior and Safety and the National Information Society Agency, with PPS as the client). Competitors must identify breaches under 24 points, of which 23 relate to the rules on state and local government contracts, on support for small business products and on software, and one checks the consistency of the notice text with the values registered in the system. The model is fixed for all competitors (Gemma 4 26B with bge-m3 embeddings, fine-tuning prohibited), the data consist of 20,000 unlabelled and 200 labelled notices with a test set of 1,853 notices, and quality is assessed by the mean F1 over the 24 points (DACON, 2026). The model of the production version is not named.

Signs. Errors in admission conditions, excessive experience requirements, recognition of the experience of certain organisations only, wrongly applied regional restrictions, breach of the publication period, exceeding the limits for bid evaluation criteria, absence of mandatory information.

Alert handling. Classification by risk, a single check by a member of staff, a correction request to the contracting authority.

Results. Until the launch of the system, notices are checked manually and selectively. In the first quarter of 2026, 15,261 notices were checked and correction requests were sent on 797, that is on 5.2 per cent. Grounds: 437 for breach of the publication period, 113 for errors in applying the small business support law, 82 for excessive experience requirements and regional restrictions, 65 for exceeding the limits for evaluation criteria (PPS, 2026). From January to May 2026 around 30,000 notices were checked and around 1,250 requests sent, of which 1,207 were accepted by contracting authorities (96.6 per cent). These figures describe the manual checking that the language model is to replace. For the model they serve as a baseline: the distribution of breaches by type and the share of requests accepted.

4.4. Korea: Review of Contract Documents in Seocho-gu

Since April 2026 the Seocho-gu District Office of the city of Seoul has been developing a system for the review of contract documents. A language model with retrieval over sources is hosted on the district's own infrastructure. The system checks the consistency of data, the presence of mandatory clauses, compliance with laws and internal rules, and the soundness of calculations. For each point it issues a finding of “compliant”, “non-compliant” or “warning” with a reference to the source. The deadline for completing development is September 2026 (Herald Economy, 2026). Entry into operation is not confirmed as at October 2026. The tool belongs to the contract performance stage, which the other LLM tools do not cover.

4.5. Chile: Document Analysis by ChileCompra

Authority and year. The Chilean Public Procurement Directorate (ChileCompra), Observatorio department. Legal basis: Law 19.886 as amended by Law 21.634 (2023). It empowered ChileCompra to conduct automated cross-checks and to apply artificial intelligence, and obliged it to refer materials on violations to the Comptroller General, the public prosecutor's office and the competition authority.

Contract. Tender 869591-2-LP25 “Análisis de Documentos mediante modelos de IA” was published on 14 August 2025; of ten bidders, the contract was awarded on 9 January 2026 to the company Sociedad de Profesionales X-Consulting Limitada for CLP 160 million (Mercado Público record). The subject of the contract: an automated procedure for detecting anomalies under 20 rules of non-compliance with procurement rules. The procedure is applied to the unstructured text documents of Mercado Público,

historical and current, around 200,000 processes a year. The contract provides for the transfer of competences to the Observatorio team. In December 2025 the project underwent an algorithmic impact assessment under the GobLab methodology of Adolfo Ibáñez University and was classed at the high impact level (ChileCompra, 2025).

Functions of the language model according to ChileCompra publications. In publications of different months ChileCompra describes four functions: analysis of tender terms, annexes and administrative acts for inconsistency, unlawful restrictions and breaches of the logic of the rules (November 2025); analysis of attached PDF documents with text recognition for possible signs of collusion and price inflation (July 2026); analysis of tender terms, technical specifications and contracts to detect 20 new types of violation (September 2026); triage of confidential complaints (November 2025).

Stage. As at October 2026 the tool is at the contract performance stage. ChileCompra describes it as a pilot and expects deployment in the fourth quarter of 2026 (ChileCompra, July 2026). There are no results. For comparison: the operating rule-based mechanism of the Observatorio applies 15 rules to all Mercado Público procedures. From August to November 2025 it issued more than 6,199 alerts, and on 54 per cent of them the authorities took measures (appendix, Table A1).

4.6. Italy: Notice Monitoring by the National Anti-Corruption Authority

On 26 May 2026 the Italian National Anti-Corruption Authority (Autorità Nazionale Anticorruzione, hereafter ANAC) adopted a programme directive on supervisory activity for 2026. The directive provides for the creation, “including through experiments with artificial intelligence”, of a structured system for monitoring notices. The purpose of the monitoring: to detect anti-competitive conditions and gross breaches of sectoral rules so that ANAC can contest notices under article 220 of the Public Contracts Code (ANAC, 2026). The technology, the contractor, the timetable and the results are not stated in the directive. The ANAC annual report for 2025 contains no indicators. The use of a language model is not confirmed by ANAC documents, so the project is assigned to the second group of the inventory and is not included in the calculations of sections 5 and 6.

4.7. European Union: the European Anti-Fraud Office and Arachne+

The European Anti-Fraud Office (hereafter OLAF). The information is taken from the European Commission’s report on the implementation of the Anti-Fraud Strategy (SWD(2026) 246). OLAF’s analytical tool, named GETI, searches structured and unstructured data simultaneously in different languages. It has been supplemented with a set of risk indicators for the proactive detection of irregularities. By May 2026 six analytical reports had been prepared. During 2025 work proceeded on a pilot to integrate the risk indicators into a tool built on generative artificial intelligence. The results of the pilot have not been published. The OLAF Report 2025 reports new artificial intelligence tools for three tasks: identifying fraud-related reports in large data sets, anonymising reports and managing case documents. The technique is not stated, and the connection with procurement is not disclosed. None of the documents links the generative pilot with the review of procurement procedures, so the project is assigned to the second group of the inventory and is not included in the calculations of sections 5 and 6.

The Arachne+ tool. This is the European Commission’s tool for risk assessment in projects financed from EU funds. It entered into operation on 27 April 2026 and is built on 78 predefined indicators. According to the Commission’s description, the tool “does not take automated decisions”. The application of artificial intelligence is assigned to the third stage of development and to the agenda of the expert group for 2026

(European Commission, 2026). Arachne+ does not use a language model, so the tool is described in the appendix.

4.8. Colombia and Other Cases without a Confirmed Language Model

In Colombia no procurement oversight tool uses a language model, judging by the documents of the authorities themselves. The predictive model of the Office of the Comptroller General of the Republic of Colombia (2025) is built on decision trees and a random forest. The HELIOS system of the Ministry of National Defence (August 2026) is described as machine learning on the structured data of the SECOP procurement system. The early warning system of the procurement agency Colombia Compra Eficiente (2025) consists of four rules. The advanced analytics centre of the Office of the Procurator General (Procuraduría) (May 2026) was created to monitor elections. Circular 023 of the Office of the Comptroller General of 4 December 2025 permits staff to use artificial intelligence only as an auxiliary means and prohibits basing conclusions on it alone.

In Brazil the ALICE robot of the Office of the Comptroller General runs on regular expressions, SQL cross-checks and machine learning classifiers. It does not use a language model. The same office's generative platform LIA (2025) is unconnected with the review of procurement. In Korea the BRIAS system of the Korea Fair Trade Commission remains a system of weighted rules. Its 2026 modernisation adds vector similarity search by product group, but not a language model (KFTC, 2026). The BRAVA tool of the Spanish National Commission on Markets and Competition, the model of the Lithuanian Special Investigation Service (2026) and the tool of the Hungarian Integrity Authority are built on machine learning over structured data.

5. Which Mechanisms the Language Models Cover

5.1. Functions and Mechanisms

Table 2 breaks the work of each LLM tool down into separate functions and ties each function to the mechanisms of the typology.

Table 2. Functions of language models in oversight tools and the mechanisms of the CILC typology

Tool	Status	Function of the language model	Who detects the violation	Mechanism, directly	Mechanism, indirectly	Performance of the function
Alice 360, TCU	operating	Recognition of the subject matter of a procurement and data cleaning for price comparison	statistical algorithm on data prepared by the model	BUD-01 (inflation of the initial price)		Not published by function; 15 of 49 requests to contracting authorities, 2 with a reduction in value
Alice 360, TCU	planned	Links of bidders with officials; infeasible bids	not disclosed	BID-02, EVL-04		not implemented
VigIA, TCE-SC	operating	Answers to auditors' questions on a notice before publication	language model	SPC-01, SPC-02, SPC-05 (by the nature of the questions)	PUB-01	215 notices with measures out of 7,711 (2.8 per cent); alerts and confirmations not published

Tool	Status	Function of the language model	Who detects the violation	Mechanism, directly	Mechanism, indirectly	Performance of the function
PPS, Korea	under construction	Admission conditions and experience requirements	language model	SPC-02	SPC-01	Manual checking baseline: 82 of 797 requests in a quarter
PPS, Korea	under construction	Regional restrictions	language model	SPC-02		included in the same 82
PPS, Korea	under construction	Publication period	language model	PUB-01		437 of 797
PPS, Korea	under construction	Limits for bid evaluation criteria	language model	SPC-03		65 of 797
PPS, Korea	under construction	Application of the small business support law	language model	SPC-02		113 of 797
PPS, Korea	under construction	Consistency of the text with registered values	language model	PUB-02	SPC-05	not published
ChileCompra	under construction	Inconsistency of tender terms and acts, unlawful restrictions	language model	SPC-05, SPC-01, SPC-02		not implemented
ChileCompra	under construction	Signs of collusion and price inflation in PDFs	language model	BID-01, BUD-01		not implemented
ChileCompra	under construction	Triage of confidential complaints	language model	none (handling of submissions, not detection)		not implemented
Seocho, Seoul	under construction	Completeness and consistency of contract documents, soundness of calculations	language model	IMP-01	IMP-02	not implemented
ANAC	second group	Anti-competitive conditions in notices	not disclosed	SPC-01, SPC-02	PUB-01	experiment; not included in the calculations
OLAF	second group	Risk indicators in a generative tool	not disclosed	not disclosed		pilot; not included in the calculations

5.2. Coverage Matrix

Table 3 shows which of the 30 mechanisms are covered by the tools of the first group with a known set of functions: Alice 360 of TCU, VigIA, PPS, ChileCompra and the Seocho-gu system. The notation distinguishes who detects the violation and the state of the tool: the letters D and I denote direct and indirect detection by the language model in an operating tool, S detection by a statistical algorithm on data prepared by the language model, D* and I* direct and indirect detection by the language model in a tool under construction, and the entry “planned” a function that is announced but not developed.

Table 3. Coverage of the mechanisms of the CILC typology by LLM tools

Code	Mechanism	TCU	VigIA	PPS	ChileCompra	Seocho
PLN-01	Fictitious or inflated need					

Code	Mechanism	TCU	VigIA	PPS	ChileCompra	Seocho
PLN-02	Agreement of requirements with the future winner					
BUD-01	Inflation of the initial price	S			D*	
BUD-02	Understatement of the price in the expectation of contract amendments					
PRC-01	Unjustified single-source procurement					
PRC-02	Splitting of the volume					
PRC-03	Abuse of urgency					
SPC-01	Requirements drafted around a predetermined bidder		D	I*	D*	
SPC-02	Excessive qualification requirements		D	D*	D*	
SPC-03	Criteria with a high weight and indeterminate scoring			D*		
SPC-04	Onerous security					
SPC-05	Indeterminate requirements for the content of a bid		D	I*	D*	
PUB-01	Artificially shortened submission period		I	D*		
PUB-02	Publication without the mandatory identifiers			D*		
PUB-03	Amendment of the documentation at the end of the period					
BID-01	Cover bids				D*	
BID-02	Concealed affiliation	planned				
EVL-01	Discriminatory rejection of bids					
EVL-02	Rejection without reasons stated					
EVL-03	Award of points in breach of the procedure					
EVL-04	Conflict of interest within the commission	planned				
RVW-01	Closure of the challenge period before the results					
RVW-02	Return of a complaint on formal grounds					
RVW-03	Refusal at the second instance					
RVW-04	Contract before review is completed					
IMP-01	Contract amendments increasing the price					D*
IMP-02	Acceptance of works not performed					I*
IMP-03	Non-application of sanctions to the supplier					

Code	Mechanism	TCU	VigIA	PPS	ChileCompra	Seocho
PAY-01	Payment against fictitious documents					
PAY-02	Sale of the order of payment					

Actual coverage is provided by the two operating tools: five mechanisms out of 30. The language model detects three of them directly (SPC-01, SPC-02, SPC-05 at VigIA) and one indirectly (PUB-01). The fifth, inflation of the initial price (BUD-01), is detected at Alice 360 by the statistical price comparison algorithm; the language model only prepares the data for it, and its contribution to detection has not been measured separately. An upward deviation of the price from the market price has not been assigned to the traces of understatement of the price (BUD-02) or of cover bids (BID-01): the typology describes these mechanisms through other indicators, price growth through contract amendments and the structure of bids, which Alice 360 does not analyse. Potential coverage, counting the systems under construction in Korea, Chile and Seocho-gu, amounts to 10 mechanisms at five stages, 9 of them directly. This is an estimate of announced functions, not an achieved result.

Across the clusters, potential coverage is uneven. Restriction of access (cluster C2) is covered almost completely: SPC-01, SPC-02 and PUB-01 are covered, SPC-04 and PUB-03 are not. Of the cluster distortion of choice (C4) the tools see only the documentation, SPC-03 and SPC-05; the bid evaluation stage remains outside their field. Simulation of competition (C3) is not covered by the operating tools; signs of collusion in documents are announced only for the ChileCompra tool under construction. Removal of the benefit beyond the competitive phase (C6) is touched at two points: the initial price (BUD-01, the operating Alice 360) and contract performance (IMP-01 and IMP-02, only the Seocho-gu system under construction). Avoidance of competition (C1) is covered by nothing. Of the cluster neutralisation of control (C5) the PPS system under construction covers only publication without mandatory information (PUB-02); the review stage is not covered by any tool.

5.3. Coincidence with the Boundary of the External Participant's Visibility

The typology divides the procurement cycle into three segments by observation position. The middle segment, from documentation to review, is visible to an external participant in the tender. LLM tools, operating and under construction, read the same documents as the participant, and at the same time, before bids are submitted. Their coverage therefore falls within the same middle segment, and within its first half. The participant also sees the bid evaluation and review stages, but the tools do not read them: the records of commissions and the decisions of review bodies are not among their input data. The initial and final segments, visible to an audit body and to an insider, are almost untouched by LLM tools. In the initial segment there is one point, initial price justification (BUD-01), which the typology assigns to audit and partly to the external participant; Alice 360 covers it, and covers it by statistics rather than by reading documents. In the final segment only the Seocho-gu system under construction reads contract documents and thereby enters the field of audit.

The language model has not widened the set of mechanisms available for observation compared with the external participant. It has changed the scale of observation: the participant reads individual procedures, the model reads all the notices in the flow. The cluster “neutralisation of control” (C5), for which the share of confirmed complaints in the CILC dataset is higher than for the other clusters, remains outside the field of view of the operating tools, and of those under construction only the check of mandatory information in the

notice (PUB-02) by the Korean system touches it. No tool checks challenge periods, the return of complaints or the actions of the second instance.

6. What Is Known about the Results

Table 4 shows which performance indicators can be calculated from published data.

Table 4. Performance indicators of LLM tools from published data

Tool	Coverage	Share reviewed	Share confirmed	Share with a consequence	Other published indicators
Alice 360, TCU	Procurements from R\$10 million, excluding construction and complex IT; share of the flow not published	not published	not published	not published	15 of 49 requests to contracting authorities in the cycle initiated by alerts; 2 of the 15 requests with a reduction in value or withdrawal of the procurement
VigIA, TCE-SC	7,711 of 33,507 notices received (23 per cent)	not published	not published	not published	215 notices with measures out of 7,711 analysed (2.8 per cent)
PPS, Korea (manual checking before launch)	15,261 notices in a quarter; share of the flow not published	not applicable: no alerts, checking is manual	not applicable	not applicable	797 correction requests (5.2 per cent of those checked); 96.6 per cent of requests accepted by contracting authorities (January to May)
ChileCompra	100 per cent of procedures announced	not implemented	not implemented	not implemented	algorithmic impact assessment (GobLab UAI) conducted, results not disclosed
Seocho, Seoul		not implemented			none

The PRUMe AI accuracy metrics (F1 0.86 and 0.82) are not included in the table: they measure the quality of data extraction on an annotated subsample of 20 documents and relate to none of the four quantities of section 3.4. The two tools of the second group, ANAC and OLAF, publish no indicators.

None of the four quantities of section 3.4 can be calculated from published data for any LLM tool; coverage is known only for VigIA. The share of confirmed alerts is the key indicator, and not one authority applying an LLM tool has published it. The published figures of TCU and TCE-SC characterise something else: the share of requests to contracting authorities initiated by alerts, and the share of analysed notices on which measures were taken. The only published shares relate to a tool without a language model, ALICE of the Office of the Comptroller General of Brazil: 23.5 per cent of well-founded alerts on a sample of 17 (Dantas, Martins, 2022) and 3.4 per cent on a sample of 29 (Ministry of Defence of Brazil, 2026). At PPS the share of correction requests accepted, 96.6 per cent, relates to manual checking. It shows the agreement of the contracting authority, not confirmation by an independent body. After the launch of the language model the same indicator will become a measure of its accuracy on one condition only: that PPS publishes the number of alerts before review by staff.

The gain from the language model over rules is described only in the TCU decision, and described qualitatively: the model recognises the subject matter of a procurement “even with imprecise descriptions”, which rules over fields cannot do, and inconsistency of the data had previously “regularly led to false positives”. There is no estimate of the gain in numbers. The TCU plenary did not include in the operative part the proposal to recommend continuation of the project and confined itself to an instruction to coordinate

technological initiatives. This is the only case in the inventory in which an authority has named the cost of implementation: the team's time taken from the analysis of procurements.

7. Signs beyond the Typology

The signs that the tools of the five jurisdictions look for mostly fit within the 30 mechanisms of the typology. As stated in section 3.1, all tools are considered here, including the tools without a language model in the appendix; the four candidates for new mechanisms come precisely from them. Several further signs prove to be indicators of mechanisms already described.

Table 5. Signs sought by the tools and not described in the CILC typology

Sign	Tool	Relation to the typology	Proposal
Participation of a supplier entered in a register of sanctioned suppliers (the CEIS and CNEP registers in Brazil; analogues at HELIOS and the Minas Gerais prosecutor's office)	ALICE CGU, ALICE TCU, HELIOS, MPMG	not described	mechanism of stage VII: admission of a debarred supplier through a related person or by inaction of the commission
Donations by bidders to election campaigns, party affiliation of partners, services to campaigns	MPMG, trails T14-T16	not described	mechanism of stage I or VII: procurement as a reward for political support
Double financing of one project from different funds	Arachne	not described	mechanism of stage X: paying for the same thing twice
Forged insurance policies during performance	DIARI, Colombia	IMP in part	mechanism of stage IX: fictitious performance security
A firm without signs of activity (no electricity connection, no employees, registration on the eve of the tender)	MPMG, ALICE TCU, PPS	indicator of BID-02	indicator
One IP address shared by several bidders, identical bid texts	CADE Cérebro	indicator of BID-01	indicator
Award price at 97-99 per cent of the estimate	Sherlock, BRIAS	indicator of BID-01, BUD-01	indicator
Breach of the notice publication period	PPS	PUB-01 directly	already described

Of the four candidates, three belong to segments the external participant does not see (stages I, IX and X), and one to the bid evaluation stage. This is exactly what the typology predicts: the tools of audit bodies and prosecutors see what the participant does not see, and the signs they look for enrich precisely those segments. Inclusion of the candidates in the catalogue requires testing on the materials of investigations and is deferred to the next version of the typology.

8. Limitations and an Open Question about Red Flags

The inventory is built from open documents and is limited to what the authorities report about their tools. An authority may use a language model without naming it in press releases; thus the VigIA model is named only in the article by court staff, not in the court's official communications. The converse is also true: the

words “artificial intelligence” in a press release do not always mean a language model. Such cases (ALICE of the Office of the Comptroller General, the model of the Office of the Comptroller General of Colombia, Arachne+) are excluded from the core. The position as at October 2026 will change: ChileCompra and PPS have announced launches for the end of the year, and ANAC and OLAF are conducting experiments.

Performance has been calculated from self-reported figures, and their denominators differ. TCU counts checks, TCE-SC notices with measures, PPS correction requests. The tools cannot be compared with one another on these figures, and the study does not do so.

The mapping of functions to mechanisms was performed by one researcher from the descriptions of the functions. Agreement between independent coders was not assessed. Where the description of a function is incomplete (OLAF, ANAC), no mechanisms have been assigned.

Open question. All the tools in the inventory, with and without a language model, issue an alert on binary logic: a rule fired or it did not, a threshold was exceeded or it was not. The BRIAS collusion score and the PPS risk class are ordinal scales, but at the output they again turn into a decision to review or not. Academic work instead builds probabilistic estimates (Decarolis, Giorgiantonio, 2020; Huber, Imhof, 2019). The CILC confirmation rate offers a way of calibrating such estimates: the share of alerts found well founded serves as an empirical estimate of the probability of a violation given the sign. The estimate is conditional: it relates only to the alerts selected for review and depends on the criteria by which the reviewer finds an alert well founded. Without random sampling of alerts for review and fixed confirmation criteria it cannot be transferred to all procedures with the same sign. Would it not be more accurate to describe the risk of a procedure by a calibrated probability with a confidence interval than by a set of red flags? That question goes beyond the scope of the present study and is posed here as the task of a separate work.

9. Conclusions

1. As at October 2026, language models in the procurement oversight of the five jurisdictions are at the implementation stage. The inventory contains six tools with a confirmed language model and two announced projects without such confirmation. Two tools are operating: for Alice 360 the court has published a decision with results, for VigIA the court has published the number of notices with measures and named the model in an article by its staff. The remaining four are in development, contract performance or prototype.
2. The only function of a language model on which an authority has reported results in a decision is not the reading of documentation in search of violations but the recognition of the subject matter of a procurement from imprecise descriptions; the violation itself is detected by a statistical algorithm. Reading of notices in search of excessive requirements and conditions drafted for a single supplier operates in one Brazilian state (VigIA) and is under construction in Korea and Chile.
3. The operating LLM tools cover five of the 30 mechanisms of the typology, three of them directly by the language model. Counting the systems under construction, potential coverage amounts to 10 mechanisms at five stages. With the exception of initial price justification (BUD-01, which the typology assigns to audit and partly to the external participant) and contract performance (IMP-01 and IMP-02 at the Seocho-gu system under construction), all of them lie in the segment visible to the external participant, and mainly at the documentation stage. The planning, choice of procurement method, bid evaluation, review and payment stages are not covered by any tool. Of the cluster “neutralisation of

control” the PPS system under construction covers one mechanism out of five. The language model has widened the scale of observation, not its boundaries.

4. No authority publishes the share of well-founded alerts by function of the language model. Without this indicator the gain from the model over the rules to which it has been added cannot be measured. To monitor the quality of a language model and of its application, an authority must collect at least four quantities for each function: the number of alerts issued, the number reviewed by a person, the number confirmed and the number that led to measures. The authorities have such data. PPS already publishes the distribution of correction requests by type of breach for manual checking.
5. The signs that the tools of the five jurisdictions look for, including those without a language model, and that are absent from the typology yield four candidates for new mechanisms. Three of them lie in segments invisible to the external participant, which accords with the structure of the typology by observation position and sets the content of its next version.
6. The binary logic of red flags is common to all the tools in the inventory. It leaves open the question of a transition to calibrated probabilities. The confirmation rate is proposed as a means of such calibration, on condition of random selection of alerts for review and fixed confirmation criteria.

Sources

Sources are grouped by section of the study.

CILC Publications (Sections 1, 3, 5, 6, 8)

Limarau, D. Mechanisms of Corruption in Public Procurement: A Typology by Stage of the Procurement Cycle. Version 1.3. CILC, 2026. DOI 10.5281/zenodo.21939393.

Krukovskiy, I. Cost Assessment of Procurement with Confirmed Violations at the Notice Publication and Bid Submission Stage. CILC, 2026. DOI 10.5281/zenodo.22237028.

Limarau, D. A Dead-End Branch of Appeal: How the Supervisory Authority Ignores the Signals of an External Reporter. CILC, 2026. DOI 10.5281/zenodo.22819817.

Surveys and Academic Works (Section 2)

OECD. Governing with Artificial Intelligence: The State of Play and Way Forward in Core Government Functions. OECD Publishing, 2025. DOI 10.1787/795de142-en.

OECD. Data Screening Tools for Competition Investigations. DAF/COMP/WP3(2022)5, 2022.

Open Contracting Partnership. AI in public procurement: between experimentation and impact. 9 July 2026.

PwC, Ecorys. Identifying and reducing corruption in public procurement in the EU. European Commission, 2013.

Fazekas, M., Toth, I. J. From corruption to state capture: a new analytical framework. Political Research Quarterly, 2016.

Porter, R., Zona, J. D. Detection of bid rigging in procurement auctions. Journal of Political Economy, 1993.

Imhof, D., Karagök, Y., Rutz, S. Screening for bid rigging: does it work? Journal of Competition Law and Economics, 2018.

Huber, M., Imhof, D. Machine learning with screens for detecting bid-rigging cartels. International Journal of Industrial Organization, 2019.

Decarolis, F., Giorgiantonio, C. Corruption red flags in public procurement: new evidence from Italian calls for tenders. *Banca d'Italia, Questioni di Economia e Finanza* 544, 2020.

Brazil (Sections 4.1, 4.2, 4.8)

Tribunal de Contas da União. Acórdão 1145/2026-Plenário de 6 de maio de 2026, TC 009.250/2025-5, relator Ministro Jorge Oliveira. Acompanhamento. Edição 2025/2026 da fiscalização contínua de aquisições públicas federais. <https://portal.tcu.gov.br/imprensa/noticias/fiscalizacao-do-tcu-acompanha-compras-publicas-federais-entre-2025-e-2026>

Santos, R. F., Santos, P. G., Marins, F., Hans, E., Bezerra, S., Terêncio, F. A., Feitoza, A. Integrando a Análise de Dados ao Dia a dia do Auditor com a Plataforma de Alertas, Riscos e Tipologias do TCU. Seminário Internacional O Futuro da Auditoria Pública, Revista TCU, 2026.

Albuquerque, A. M. de, Wensing, I. M., Joppi Filho, N. L. Inteligência Artificial Generativa (IAG): perspectivas inéditas no controle público. *Revista do Tribunal de Contas do Estado de Santa Catarina*, ano 2, n. 3, p. 150-164, maio/out. 2024. DOI 10.52028/tce-sc.v02.i03.ART08.SC.

Tribunal de Contas do Estado de Santa Catarina. Inteligência artificial criada pelo TCE/SC possibilita retificação em 215 editais de licitação, com previsão de investimentos de R\$ 2 bilhões. 9 de outubro de 2024. <https://www.tcesc.tc.br/inteligencia-artificial-criada-pelo-tcesc-possibilita-retificacao-em-215-editais-de-licitacao-com>

Bezerra, A. de S., Lopes, L. C. Controle externo baseado em LLM: avaliação empírica do PRUMe AI. *Revista Aracê*, v. 7, n. 9, e8409, 2025. DOI 10.56238/arev7n9-271.

Borges, A. O., Belga, J. L., Gazzola, P. O. L., Morais, A. B. ALICE: Análise automatizada de contratações públicas para prevenção de irregularidades e combate à corrupção. *Cadernos Técnicos CGU*, 2025. https://revista.cgu.gov.br/Cadernos_CGU/article/download/867/552/5714

Dantas, F., Martins, R. Efetividade dos alertas da ferramenta ALICE. Universidade Católica de Brasília, 2022.

Ministério da Defesa, Secretaria de Controle Interno. Relatório final de consultoria sobre a ferramenta Alice. 22 de janeiro de 2026. <https://www.gov.br/defesa/pt-br/acesso-a-informacao/auditorias-1/acoes-de-consultoria-1/arquivos/relatorios/2025/2-relatorio-final-alice-8615445.pdf>

Controladoria-Geral da União. Alice: Analisador de Licitações, Contratos e Editais. <https://www.gov.br/cgu/pt-br/assuntos/auditoria-e-fiscalizacao/alice>

Portal Tela. IA da CGU emitiu 35 mil alertas de risco em licitações federais. 30 de março de 2026. <https://www.portaltela.com/noticias/politica/2026/03/30/ia-da-cgu-emitiu-35-mil-alertas-de-risco-em-licitacoes-federais/>

Conselho Administrativo de Defesa Econômica (CADE). Data screening tools for competition investigations, Note by Brazil. OECD, DAF/COMP/WP3/WD(2022)36, 2022. [https://one.oecd.org/document/DAF/COMP/WP3/WD\(2022\)36/en/pdf](https://one.oecd.org/document/DAF/COMP/WP3/WD(2022)36/en/pdf)

Ministério Público de Minas Gerais, Universidade Federal de Minas Gerais. Painel Trilhas de Licitações. Anais do Simpósio Brasileiro de Banco de Dados (SBBDD), 2023. https://sol.sbc.org.br/index.php/sbbdd_estendido/article/download/44196/43959/

Korea (Sections 4.3, 4.4, 4.8)

ETNews. 조달청, AI로 자체입찰 전수 점검, 공공조달 공정성 높인다. 6 July 2026.

DACON. 나라장터 자체입찰 공고 법령 위반사항 모니터링 AI 경진대회. Ministry of the Interior and Safety, National Information Society Agency, Public Procurement Service, 2026.

Public Procurement Service (PPS). Results of the check of self-conducted procurement for breaches of legislation in the first quarter of 2026 (as reported by 뉴스서울 on 27 April 2026 and 정보통신신문 on 1 June 2026).

Herald Economy. 서초구, AI 기반 계약서류 검토시스템 구축. 7 May 2026.

Korea Fair Trade Commission. Competition and Corruption in Public Procurement, Note by Korea. OECD, DAF/COMP/WD(2026)77, 9 June 2026. [https://one.oecd.org/document/DAF/COMP/WD\(2026\)77/en/pdf](https://one.oecd.org/document/DAF/COMP/WD(2026)77/en/pdf)

Korea Fair Trade Commission. Data screening tools for competition investigations, Note by Korea. OECD, DAF/COMP/WP3/WD(2022)31, 2022. [https://one.oecd.org/document/DAF/COMP/WP3/WD\(2022\)31/en/pdf](https://one.oecd.org/document/DAF/COMP/WP3/WD(2022)31/en/pdf)

Chile (Section 4.5)

Mercado Público. Licitación 869591-2-LP25 “Análisis de Documentos mediante modelos de IA”: award record (Resolución 16-B of 9 January 2026).

ChileCompra. ChileCompra lanza licitación inédita para aplicar inteligencia artificial en compras públicas. 20 August 2025; ChileCompra expone avances en uso de datos e inteligencia artificial. 28 November 2025; ChileCompra aplicó herramienta para el uso ético de IA en su proyecto de detección de anomalías. 31 December 2025; Cuenta Pública Participativa 2026. 27 July 2026; ChileCompra presenta en instancia internacional impulsada por el Banco Mundial. 13 July 2026; ChileCompra comparte avances en inteligencia artificial en encuentro G7/MMGP. September 2026.

ChileCompra. ChileCompra incrementa en 207% la detección de irregularidades con cruce masivo de datos y tecnologías de inteligencia analítica. December 2025. <https://www.chilecompra.cl/2025/12/chilecompra-incrementa-en-207-la-deteccion-de-irregularidades-con-cruce-masivo-de-datos-y-tecnologias-de-inteligencia-analitica/>

ChileCompra. ChileCompra detecta potenciales conflictos de interés por \$3.452 millones tras cruce de datos masivo. January 2026. <https://www.chilecompra.cl/2026/01/chilecompra-detecta-potenciales-conflictos-de-interes-por-3-452-millones-tras-cruce-de-datos-masivo/>

Italy and the European Union (Sections 4.6, 4.7)

ANAC. Direttiva programmatica sull'attività di vigilanza, anno 2026. 26 May 2026.

European Commission. Commission Staff Working Document SWD(2026) 246: monitoring of the Commission Anti-Fraud Strategy action plan, state of play 31 May 2026.

European Commission, Anti-Fraud Knowledge Centre. What is Arachne? (Arachne+), 2026.

OLAF. The OLAF Report 2025. 20 April 2026.

European Court of Auditors. Special report 06/2026. 2026. https://www.eca.europa.eu/ECAPublications/SR-2026-06/SR-2026-06_EN.pdf

Comisión Nacional de los Mercados y la Competencia. Memoria 2024. https://www.cnmc.es/sites/default/files/editor_contenidos/CNMC/Memorias/Memoria_CNMC_2024_.pdf

Specialiųjų tyrimų tarnyba (STT). Pilot of explainable machine learning for the detection of risks in procurement. February 2026. <https://stt.lt/en/news/7481/>

Colombia (Section 4.8, Table A1)

Contraloría General de la República de Colombia. Modelo Analítico Predictivo en Contratación, 19 May 2025; Circular 023 de 4 de diciembre de 2025.

Contraloría General de la República de Colombia. Comunicado de prensa sobre la DIARI. September 2024. <https://ambitojuridico.com/sites/default/files/2024-09/COMUNICADO-LaDiariha-2024%20%282%29.pdf>

Cambio Colombia. El cerebro digital de la Contraloría que protege los recursos públicos. October 2025. <https://cambio colombia.com/contenido-especial/articulo/2025/10/el-cerebro-digital-de-la-contraloria-que-protege-los-recursos-publicos>

Superintendencia de Industria y Comercio. Herramientas de investigación: Sherlock. Presentation for UNCTAD. https://unctad.org/system/files/non-official-document/tdrbpconf10_unrc_present_investigative_tech_rusunque_sp.pdf

Ministerio de Defensa Nacional de Colombia. Sistema HELIOS (as reported by Vanguardia and Cambio on 6 August 2026).

Appendix. Tools without Language Models in the Same Jurisdictions

Table A1. Procurement oversight tools built on rules, cross-checks and classical machine learning

Tool, authority, year	Technique	Signs	Clusters of typology mechanisms	Published results
ALICE, Office of the Comptroller General (CGU), Brazil, 2015	regular expressions over notice texts, SQL cross-checks against registers, price ranges, classifiers	restrictive requirements, specifications drafted for a supplier, inconsistency of documents, deadlines, price inflation, sanctioned suppliers	C2, C4, C6	2024: 161,000 processes, 212 audits, R\$1.25 billion of effect; January 2023 to February 2026: 35,000 alerts, 729 audits; share well founded 23.5 per cent (17 alerts, 2021) and 3.4 per cent (29 alerts, 2025)
Projeto Cérebro, competition authority CADE, Brazil, 2014	Newcomb-Benford tests, rotation entropy, shared IP address, text similarity	cover bids, rotation	C3	one documented case (2018-2021)
Painel Trilhas de Licitações, Public Prosecutor's Office of the State of Minas Gerais (MPMG), 2023	graphs, 20 trails	single bidder, shared contact details, links with officials, donations	C3, C4, C6	MRR 0.210 against 0.035 for random selection
BRIAS, Korea Fair Trade Commission (KFTC), 2006	weighted rules, collusion score	price ratio, number of bidders, simultaneous submission, repeated consortia	C3	2018-2021: 26 cases, 9 sanctions; 2007-2017: 10,036 flags, 7 investigations
Observatorio ChileCompra (rules), 2024-2025	15 rules, cross-checks against the SIAPER register of public officials	splitting, official with a stake in a supplier, conflict of interest, formal defects	C1, C3, C4, C5	2025: more than 13,000 notifications, 54 per cent with measures taken, 206 referrals to the authorities
DIARI, Office of the Comptroller General, Colombia, 2025	random forest	delays, non-performance, forged policies, direct contracts	C1, C6	946 alerts (2025), share confirmed not published

Tool, authority, year	Technique	Signs	Clusters of typology mechanisms	Published results
Sherlock, Superintendence of Industry and Commerce (SIC), Colombia	three rules over SECOP	price at 97 per cent, repeated wins, single bidder	C3	no cases from the tool
HELIOS, Ministry of National Defence of Colombia, 2026	machine learning over structured data	single bidder, rapid awards, extensions, sanctions	C2, C3, C6	none
Arachne and Arachne+, European Commission, 2013 and 2026	107 and 78 indicators, cross-checks against the Orbis database	concentration, links, contract amendments, double financing	C3, C4, C6	no outcome statistics; the European Court of Auditors (report 06/2026) did not assess effectiveness
BRAVA, National Commission on Markets and Competition (CNMC), Spain	machine learning classifier	price patterns	C3	one case; accuracy not published
Model of the Special Investigation Service (STT), Lithuania, 2026	explainable machine learning	not disclosed	not determined	pilot