ALICE 360 V2

AI-Powered Procurement Oversight
Advanced Pilot with Validated Results

Brazilian Federal Court of Accounts (TCU)

Status: Pilot with **8 organizations**

Proven Impact: 4x productivity gain

Scale: 1.5 million items/month analyzed

LOTA Talks 2025: November 25, 2025

✓ Field Validated   ✓ Positive ROI   ✓ Multi-Organizational

TCU
TRIBUNAL DE CONTAS DA UNIÃO

💰 MATERIALITY AND CONTEXT

Brazilian Public Procurement Landscape

Total Procurement Value (Jul 2023 - Jul 2024)

Brazilian National Public Procurement Portal (PNCP)
(BRL 476.75B at ~5.30 BRL/USD)

At-Risk Procurement

Research suggests **USD $20 billion** potentially at risk for irregularities
(BRL 107.25B)

Potential Annual Savings

If only 5% of at-risk value represents actual overpricing
(BRL 5.36B)

Distribution by Government Level

Level	Percentage	Characteristics
Municipal	44.8%	Highest concentration, diverse contracts
Federal	29.5%	High-value strategic contracts
State	25.7%	Major infrastructure projects

Why AI Matters: Early detection through systems like ALICE 360 has massive potential ROI for oversight institutions worldwide.

TCU
TRIBUNAL DE CONTAS DA UNIÃO

📊 HIGHEST VALUES - LAST 30 DAYS

Federal Procurement - High Materiality Items

Category	Total Value (USD)	Challenge
Virtual Warehouse Services	USD $320M (BRL 1.7B)	Complex IT infrastructure
Engineering Services	USD $245M (BRL 1.3B)	Requires specialized technical knowledge
Specialized Pharmaceuticals	USD $162M (BRL 859M)	Medical expertise needed
Ambulances	USD $157M (BRL 833M)	Technical specifications vary greatly
Aircraft Leasing	USD $116M (BRL 615M)	Highly specialized aviation knowledge

Strategic Insight

High-value items often require **specialized technical knowledge** for adequate analysis. This creates an opportunity for ALICE 360 to act as an **intelligent dispatcher**, routing items to the right specialized audit teams based on technical complexity and materiality.

TCU
TRIBUNAL DE CONTAS DA UNIÃO

💰 PROVEN VALUE

Validated Results in Pilot Program

Productivity Gain

4x

4 investigations in 3h (vs 1 before) - TCU Procurement Audit

💰 **Economic Value: USD 3,255/month savings per auditor**
(Details on next slide)

Organizations in Pilot

6 State/Municipal Courts + 2 TCU audit departments

Success Rate

Our strategic sweet spot for maximum ROI

(Full explanation ahead →)

Categories Created

Automatically by LLM agents

Cost/User/Day

Current version (metadata only)

✓ **Validated Success Cases**

- **Mato Grosso do Sul Court:** Overpricing identified and validated in field inspections
- **TCU Procurement Audit:** Elimination of manual download/extraction/comparison work
- **Positive ROI:** API costs justified by productivity gain and findings

TCU
TRIBUNAL DE CONTAS DA UNIÃO

👥 USER FEEDBACK

Real Experiences from Pilot Organizations

✓ Reported Gains

- **4x productivity increase:** From 1 investigation per 3 hours to 4 investigations in same timeframe (TCU Procurement Audit)
- **Field validation:** Actual overpricing identified and confirmed in inspections (Mato Grosso do Sul Court)
- **Configurable filters:** Each organization sets own materiality threshold
- **Complete transparency:** Explicit justifications for every categorization and similarity decision
- **Work elimination:** No more manual download, extraction, and comparison of procurement documents

💰 Economic Value of Productivity Gain

Auditor Cost per Hour:	Savings per Investigation:	Monthly Savings Estimate:
USD 72.33/h	**USD 162.74**	**USD 3,255**
(BRL 433.98/h at 6.00 BRL/USD) Includes salary, benefits, vacation & 13th salary	2.25 hours saved × USD 72.33/h (BRL 976.45)	20 investigations/month per auditor (BRL 19,529)

Note: Based on Brazilian Federal Court of Accounts auditor total compensation (salary + 40% benefits + vacation/13th salary)

⚠️ Challenges Being Addressed

- **PDF quality:** Variable quality of source documents - some are poor-quality scans
- **Normalization delicacy:** Text normalization is subtle and doesn't always help
- **Cost vs. volume balance:** Finding optimal balance between analysis coverage and API costs
- **Feedback structure:** Moving from informal comments to structured rating system (Google Forms in development)

Continuous Improvement: We're developing a structured feedback system for rating analysis quality and gathering improvement suggestions. User feedback directly influences the Reviewer agent's weekly optimization proposals.

TCU
TRIBUNAL DE CONTAS DA UNIÃO

⚠️ AI CHALLENGES: MARKET LESSONS

Learning from Industry's AI Missteps

Real Examples of AI Failures

- **Amazon (2018):** Recruiting AI biased against women - trained on historical data reflecting gender bias
- **Google Photos (2015):** Classified Black individuals as "gorillas" - inadequate training data diversity
- **Microsoft Tay (2016):** Chatbot turned racist in 24h - no adequate safeguards
- **Healthcare AIs:** Multiple cases of diagnostic bias against minority groups

⚠️ Common Pitfalls

- Blind trust in AI outputs
- No validation mechanisms
- Hidden biases in training data
- Lack of human oversight
- No explainability requirements
- Insufficient testing before deployment

✓ ALICE 360 Safeguards

- Multi-agent validation (agents check each other)
- Mandatory human approval for category changes
- Explicit justifications for every decision
- No single point of AI authority
- Progressive rollout with constant monitoring
- Real user feedback integrated continuously

TCU
TRIBUNAL DE CONTAS DA UNIÃO

🎯 THE SWEET SPOT: 70-75% SUCCESS DELIVERS MAXIMUM VALUE

Expected Value of Perfect Information (EVPI) - Douglas Hubbard

*"**Smart information** beats **perfect information**.*
*The right level of certainty delivers **maximum value at optimal cost**."*

📈 The EVPI Principle

Key insight from "How to Measure Anything":

- **ECI (Expected Cost of Information):** Cost to acquire information
- **EVI (Expected Value of Information):** Value gained from information
- **EVPI (Expected Value of Perfect Information):** Maximum theoretical value

⭐ **Strategic Optimization: 70-75% is where maximum value meets optimal cost!**
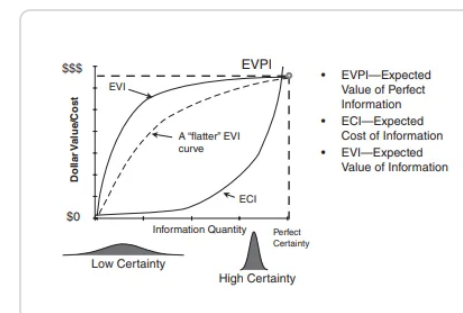
EVPI curve: As information quality increases, value increases rapidly at first, then levels off. Perfect information costs exponentially more for diminishing returns.

✓ **ALICE 360 Reality**

✨ **70-75% Success Rate**
Our Strategic Sweet Spot for Maximum ROI

- 🎯 **Optimal value delivery:** USD 3,255/month savings per auditor
- 💰 **Outstanding ROI:** 32-81x return on investment
- ✅ **Field-proven results:** Real overpricing identified and confirmed
- ⚡ **Strategic choice:** More precision would cost exponentially more for minimal gains

🌟 **Smart AI Implementation**

We've achieved the optimal balance. Our 70-75% success rate isn't a limitation—it's a **strategic optimization**. Following Hubbard's EVPI principle, this delivers **maximum practical value** (USD 3,255/month savings, **32-81x ROI**) at optimal cost. Chasing 95-100% would cost exponentially more while adding **minimal additional value**. This is **smart AI deployment**—focused on real-world impact, not theoretical perfection.

🎯 VALIDATED STRATEGY

"Accuracy Before Precision"

*"Better to be **in the right region** than to be **precise in the wrong place**"*

✓ Our Pragmatic Approach

- First: Get close to the target (accuracy)
- Then: Refine precision iteratively
- Generates immediate value for auditors
- Continuous improvement based on feedback
- SAIs already identifying real irregularities
- Positive ROI validates the approach

⚠️ Avoided Pitfall

- Waiting for perfect precision before launch
- Building theoretical solution without validation
- No user feedback loops
- Analysis paralysis delaying value delivery
- Overengineering before proving concept
- Ignoring practical needs of auditors

Result: Audit organizations are finding actual overpricing that justifies investment, even without perfect statistical precision. The system delivers value TODAY while continuously improving.

TCU
TRIBUNAL DE CONTAS DA UNIÃO

🔁 EVOLUTION: ALICE 360 V1 → V2

From Pattern Matching to Multi-Agent Intelligence

📋 Version 1.0 (2021-2023)

- **Pattern Matching:** Based on regular expressions and fixed rules

- **Manual Categories:** Created and maintained by humans

- **Limited Flexibility:** New patterns required code changes

- **No Self-Learning:** System didn't improve over time

- **High Maintenance:** Constant manual updates needed

Deprecated

🤖 Version 2.0 (2024-2025)

- **Multi-Agent Architecture:** Specialized LLM agents with checks and balances

- **Automatic Categorization:** Creates categories autonomously with justifications

- **Self-Learning:** Improves through weekly reviews and feedback

- **Adaptive:** Handles new procurement types automatically

- **Network Effect:** Each organization enriches the system for all

In Production - Pilot

✓ **Why V2 Changes Everything**

The shift to multi-agent architecture fundamentally changed the system's capabilities. Instead of rigid rules, we now have **intelligent agents that understand context**, **validate each other's work**, and **explain their reasoning**. This makes the system trustworthy enough for government audit use.

TCU
TRIBUNAL DE CONTAS DA UNIÃO

🎯 PROCUREMENT AUDIT: STRATEGIC DISPATCHER

Optimizing Specialized Audit Resources

1   ### ALICE 360 Analysis

Identifies technical category, materiality value, complexity, and degree of specialization required

2   ### Procurement Audit Decision

Keep: Items not requiring very specific technical expertise (pharmaceuticals, admin services, consumables)
Forward: Items requiring highly specialized knowledge to specialized departments

3   ### Specialized Audits Receive

IT Audit: Complex technology infrastructure
Urban Audit: Civil works and engineering
Health Audit: Medical equipment and procedures

✓ Strategic Value

This approach **optimizes everyone's time**. Specialized audit teams receive only cases that truly need their expertise, pre-filtered by materiality and technical complexity. Procurement Audit can handle general cases efficiently while ensuring specialized cases get the right expertise. **Applicable to any SAI with specialized audit departments.**

⚖️ DEMOCRACY ANALOGY: CHECKS AND BALANCES

Why Multi-Agent Architecture Builds Trust

> *"Just as **democracy strengthens** when no power is absolute, AI systems become **trustworthy** when everyone validates, everyone corrects, everyone learns."*

🏛️ Democratic Government

- **Executive:** Proposes and implements
- **Legislative:** Creates and reviews laws
- **Judiciary:** Validates and corrects
- **Result:** No single authority, mutual validation, accountability

🤖 ALICE 360 Multi-Agent

- **Categorizer:** Proposes categories and dimensions
- **Analyzer:** Compares and identifies discrepancies
- **Reviewer:** Validates and proposes improvements
- **Result:** No agent has final authority, mandatory human approval

Critical Safeguard

The Reviewer agent's proposals for category changes require **mandatory human approval**. This ensures that AI remains a powerful tool that **augments human judgment** rather than replacing it. Transparency and accountability are built into the architecture.

🤖 MULTI-AGENT ARCHITECTURE

Built-in Checks and Balances

🏷️ CATEGORIZER

- Analyzes item descriptions
- Defines relevant dimensions
- Assigns weights by price impact
- Generates structured JSON

🔍 ANALYZER

- Compares dimensions between items
- Calculates similarity scores
- Justifies accepts/rejects
- Identifies price discrepancies

⚖️ REVIEWER

- Reviews categories weekly
- Improves similar items filtering
- Refines dimensions/weights
- Proposes optimizations

Critical Design Principle

No agent has final authority. The Categorizer proposes, the Analyzer validates, the Reviewer suggests improvements - but **humans approve** significant changes. This creates a system where AI power is checked by AI skepticism and human oversight.

TCU
TRIBUNAL DE CONTAS DA UNIÃO

🎯 INTELLIGENT FILTERING AGENT

First-Stage Optimization Before Main Pipeline

1   Semantic Analysis

Understands full context, identifies synonyms, normalizes variations automatically

2   Automatic Categorization

Uses predefined categories, creates new ones when needed, provides confidence scores

3   Strategic Keyword Extraction

Primary keywords (weight 3), secondary keywords (weight 1), technical boost terms

4   YAML-Based Filtering

Uses category-specific YAML parameters and catalog to filter similar items (SQL optimization in V3)

92%	**~2s**	**90%**
Categorization Accuracy	Processing Time/Item	Relevant Results

✓ Strategic Impact

This agent **reduces API costs by ~60%** by strategically identifying the most relevant items before heavy processing, while simultaneously **increasing analysis precision** by focusing resources where they matter most.

⚙️ PROCESSING PIPELINE

From Document to Insight

📄 Phase 1: Intelligent OCR

Dual pipeline - Tesseract for economy, GPT Vision for difficult PDFs. Success rate: 70-75%

🏷️ Phase 2: Automatic Categorization

Creator agent analyzes descriptions and creates categories with relevant dimensions

🔍 Phase 3: Vector Search

LanceDB indexed with PNCP metadata. Next: on-demand document enrichment for N similar items

📊 Phase 4: Similarity Analysis

Analyzer agent compares dimensions, calculates scores. Configurable threshold with explicit justifications

📈 Phase 5: Interactive Report

Overview with all contracts, detailed analyses for items above threshold, full drill-down capability

🔄 Phase 6: Self-Learning

Reviewer agent reviews categories weekly. Structured feedback in development. Future ML training with LLM data

TCU
TRIBUNAL DE CONTAS DA UNIÃO

💰 COST CONTROL

Configurable Balance Between Coverage and Investment

📊 Lower Threshold Approach

Example: USD $2,000 (BRL 10,000)

- **More items analyzed**

- Greater coverage of procurement

- Finds more potential irregularities

- Higher API cost

- More alerts for team to investigate

  Best for: Organizations with larger budgets and investigation capacity

🎯 Higher Threshold Approach

Example: USD $190,000 (BRL 1,000,000)

- **Lower API cost**

- Focus on high materiality

- Fewer but more significant alerts

- Less coverage but higher impact

- Manageable workload

  Best for: Smaller SAIs or limited investigation teams

Key Principles

- **Each SAI controls their balance** between sample size and cost
- **Independent containers** - each organization uses own API keys
- **Full cost transparency** - real-time monitoring and reporting
- **Network effect advantage** - shared categories reduce everyone's costs

TCU
TRIBUNAL DE CONTAS DA UNIÃO

🌐 NETWORK EFFECT

Collaboration Multiplies Value

*"Each SAI that joins the system **enriches the database for everyone else**"*

1

SAI 1 Creates Pharmaceutical Categories

Defines dimensions for drugs: active ingredient, dosage, presentation → Everyone benefits

2

SAI 2 Creates IT Equipment Categories

Defines dimensions for computers: processor, RAM, storage → Enriched database for all

3

SAI 3 Finds Categories Already Ready

Reduced cost, faster deployment, immediate value from others' work

Organizations in Pilot

Shared Categories

Value Grows Exponentially

For international collaboration: Shared categories across SAIs worldwide could create a global procurement knowledge base, making AI-powered oversight more affordable and effective for all participants.

🚀 **NEXT EVOLUTION**

Strategic On-Demand Enrichment

Current Limitation

Our vector database doesn't yet have full document extraction for all items. Current analysis uses **PNCP metadata only**, which already delivers significant value.

1 **Extract Complete Description**

Get full details from the item of interest flagged by the system

2 **Search for N Similar Items**

Find 30 similar items per flagged item (still using metadata for search)

3 **Fetch Complete Documents**

Extract full procurement documents for those N similar items

4 **Detailed Comparison**

Deep analysis with all dimensions, prices, specifications, and contexts

Trade-offs

- **Volume:** 10 items × 30 similar = 300 documents
- **Cost:** Increases proportionally
- **Time:** Parallel processing via containers
- **Control:** SAIs define how many similar items

Strategic Benefits

- **Not exhaustive:** Strategic, targeted enrichment
- **Higher quality:** Much better context for analysis
- **Proven value:** Metadata-only already works well
- **Configurable:** Each SAI chooses depth vs. cost

TCU
TRIBUNAL DE CONTAS DA UNIÃO

🎯 PROJECT MATURITY

Ready for Progressive Expansion

✅ **Architecture Validated**

Multi-agent system proven in production with 8 organizations. Core design is solid and scalable.

Completed

✅ **ROI Demonstrated**

4x productivity gain validated. Actual overpricing identified in field. API costs justified by results.

Completed

✅ **Network Effect Proven**

~50 shared categories created collaboratively. Each new organization reduces costs for all.

Completed

🔄 **Feedback System Development**

Moving from informal feedback to structured Google Forms rating system for continuous improvement.

In Progress

📋 **On-Demand Enrichment**

Next phase: fetch complete documents for N similar items per alert to multiply analysis quality.

Planned - 2025

✓ **System Is Ready**

ALICE 360 V2 has moved beyond proof-of-concept. The architecture is proven, ROI is validated, and the system is generating **actionable value today** while continuously improving. Ready for progressive expansion to analyze 1.5 million items monthly and potentially collaborate with SAIs worldwide.