

PROPIUESTA TÉCNICA

Licitación: Contratación del servicio de análisis de documentos mediante modelos de Inteligencia Artificial para la detección de anomalías en procesos de compra pública.

Cliente: Dirección de Compras y Contratación Pública (ChileCompra).

Fecha: 13 de Agosto de 2025

1. Análisis del Proyecto

1.1 Entendimiento del Problema

La Dirección ChileCompra enfrenta un desafío de **escalabilidad y cobertura** en su labor de fiscalización. Actualmente, el Departamento Observatorio realiza un monitoreo manual sobre una muestra reducida (aprox. 4.500 procesos anuales), lo que representa menos del 0,5% del universo total de transacciones. El "dolor" principal radica en la existencia de una gran cantidad de información no estructurada (archivos PDF, bases de licitación, resoluciones) que escapa a las reglas de negocio automatizadas actuales.

El cambio propuesto es estratégico, no solo técnico. Al implementar una solución basada en **Inteligencia Artificial Generativa (LLMs)**, ChileCompra busca transitar de un modelo reactivo y muestral a un modelo **proactivo y masivo**, capaz de procesar 200.000 documentos anuales. Esto permitirá elevar los estándares de **probidad y transparencia**, detectando anomalías complejas que requieren interpretación semántica y contextual, mitigando riesgos de corrupción y prácticas contrarias a la libre competencia.

1.2 Análisis de Requerimientos

- **Requerimientos Funcionales:**

- Ingesta y clasificación automática de documentos (PDFs, imágenes) desde el repositorio cloud.
- Extracción de texto mediante **OCR avanzado** (Optical Character Recognition).
- Segmentación semántica (*Chunking*) y generación de *Embeddings*.
- Evaluación de **20 reglas de negocio/hipótesis de incumplimiento** utilizando LLMs.
- Generación de datos estructurados (SQL) y categorización de hallazgos (scoring de riesgo).

- **Requerimientos Técnicos (Stack Tecnológico Sugerido - GCP):**

- **Infraestructura:** Google Cloud Platform (Mandatorio).
- **Procesamiento & ETL:** Cloud Functions, Cloud Run, Dataflow.
- **IA & ML:** Vertex AI (para gestión de modelos), Document AI (OCR), Gemini Pro / PaLM 2 (LLMs).
- **Base de Datos:** Cloud SQL (PostgreSQL) para metadatos y Vector Search para embeddings.
- **Lenguaje:** Python (Frameworks: LangChain, Pandas).
- **Repositorio:** Artifact Registry, Cloud Storage.

1.3 Análisis de Riesgos y Estrategia de Mitigación

Riesgo	Tipo	Impacto	Estrategia de Mitigación y Contingencia
Alucinaciones del Modelo (Falsos Positivos)	Tecnológico	Alto	Implementación de arquitectura RAG (Retrieval-Augmented Generation) para anclar respuestas en el contexto documental. Uso de <i>Prompt Engineering</i> avanzado (Chain-of-Thought) y validación humana en el loop (HITL) durante el entrenamiento.
Calidad deficiente de documentos origen (Escaneos antiguos)	Tecnológico	Medio	Uso de motores OCR de última generación con pre-procesamiento de imagen (desruido, binarización). Implementación de un umbral de confianza; documentos ilegibles se marcan para revisión manual.
Consumo excesivo de créditos GCP	Gestión	Medio	Monitoreo granular de costos con alertas de presupuesto en GCP. Optimización de tokens en los prompts y uso de modelos "distilled" o más ligeros para tareas de menor complejidad cognitiva.
Sesgo en la interpretación normativa	Normativo	Alto	Curación estricta del set de datos de entrenamiento/contexto con dictámenes de Contraloría vigentes. Pruebas de regresión específicas para asegurar neutralidad en la evaluación.

2. Propuesta de Solución

2.1 Detalle de Solución Técnica (Metodología)

Proponemos una metodología híbrida que combina la estructura de fases (Waterfall) requerida por los entregables con la agilidad de **Scrum** para el desarrollo de los modelos. La arquitectura se basará en un patrón **ETL + LLM Pipeline**:

- **Fase de Discovery & Diseño:** Levantamiento detallado de las 20 reglas con el equipo del Observatorio. Definición de la arquitectura en GCP.
- **Ingesta & Pre-procesamiento (ETL):** Pipeline automatizado que detecta nuevos archivos, aplica OCR y almacena texto limpio en formato Markdown.
- **Motor de Inferencia (Core AI):**
 - *Vectorización:* Transformación de texto a vectores para búsqueda semántica.
 - *Evaluación:* El LLM recibe el contexto (fragmento del documento) y la regla normativa (Prompt) para determinar cumplimiento.
- **Post-procesamiento:** Estructuración de la salida en JSON/SQL y carga en el Data Warehouse de ChileCompra.
- **QA & Trazabilidad:** Uso de matrices de trazabilidad para asegurar que cada regla normativa esté cubierta por un set de pruebas unitarias y de integración.

2.2 Etapas del Proyecto

El proyecto se ejecutará en **12 meses**, alineado estrictamente con los hitos de pago y entregables de las bases:

- **Etapa 1 (Mes 1-2):** Kick-off, Diseño de Arquitectura, Definición de las 20 reglas y Metodología.
- **Etapa 2 (Mes 3-4):** Implementación del Flujo Base (OCR/ETL) y validación de infraestructura.
- **Etapa 3 (Mes 5-6):** Desarrollo e implementación del primer lote de 5 reglas (Prioridad Alta).
- **Etapa 4 (Mes 7-8):** Desarrollo e implementación del segundo lote de 7 reglas.
- **Etapa 5 (Mes 9-10):** Desarrollo e implementación del tercer lote de 8 reglas.
- **Etapa 6 (Mes 11-12):** Consolidación, Transferencia Tecnológica, Cierre y paso a Garantía.

2.3 Cronograma Resumido

Fase / Hito	Duración	Descripción Clave	Hito de Validación
Inicio y Metodología	Mes 1-2	Plan de trabajo, Fichas de reglas, Diseño de Prompts.	Informe N°1 Aprobado
Infraestructura OCR/ETL	Mes 3-4	Despliegue en GCP, Pipeline de OCR funcional, Markdown output.	Informe N°2 Aprobado
Iteración 1 (5 Reglas)	Mes 5-6	Implementación reglas 1-5, Análisis de cobertura y precisión.	Informe N°3 Aprobado
Iteración 2 (7 Reglas)	Mes 7-8	Implementación reglas 6-12, Ajuste de modelos.	Informe N°4 Aprobado
Iteración 3 (8 Reglas)	Mes 9-10	Implementación reglas 13-20, Dashboard de resultados.	Informe N°5 Aprobado
Cierre y Transferencia	Mes 11-12	Documentación final, Capacitación, Paso a Producción.	Informe N°6 Aprobado

2.4 Definición de Entregables

Entregable	Descripción	Criterios de Aceptación	Responsable
Plan de Proyecto y Metodología	Cronograma detallado, fichas de las 20 reglas y diseño de prompts.	Aprobación de contraparte técnica. Cronograma en MS Project.	Jefe de Proyecto
Pipeline OCR/ETL	Código fuente y despliegue del motor de extracción de texto.	Procesamiento exitoso de muestra de documentos. Output en Markdown.	Ingeniero de Datos / Científico de Datos
Informes de Avance (Reglas)	Resultados estadísticos, SQL dumps con "flags" de anomalías.	Validación de precisión (Recall/Precision) sobre dataset de prueba.	Líder Técnico
Manuales y Transferencia	Documentación de arquitectura, operación y guías de extensión.	Documentación completa en formato Wiki/PDF. Capacitación realizada.	Equipo Completo

3. Descripción del Equipo de Trabajo

Modelo de Gobierno

Se establecerá un modelo de gobierno ágil. El **Jefe de Proyecto (PM)** actuará como punto único de contacto con el Administrador de Contrato de ChileCompra. Se realizarán reuniones de seguimiento quincenales (Sprint Reviews) y comunicación continua mediante herramientas colaborativas (Jira/Teams). Se gestionará un **Backlog** priorizado de riesgos y dependencias.

Equipo Propuesto

Rol	Perfil / Skills Clave	Experiencia	Dedicación
Líder Técnico	Arquitecto Cloud, Python, IA/ML. Maestría en Cs. Computación.	+5 años liderando proyectos complejos. Exp. en GCP.	100%
Gestor de Proyectos (PM)	Certificación PMP / Scrum Master. Cloud Digital Leader.	+5 años en gestión de proyectos TI públicos/privados.	100%
Ingeniero de Datos (DevOps)	GCP Data Engineer, Terraform, Kubernetes, CI/CD pipelines.	+3 años en automatización y flujos ETL en nube.	100%
Científico de Datos (CV/LLM)	Experto en NLP, OCR (Tesseract/Vision AI), Prompt Engineering.	+3 años en modelos generativos y visión por computador.	100%

4. Competencias del Proveedor (Credenciales)

[NOMBRE EMPRESA] cuenta con una sólida trayectoria en la modernización del estado y transformación digital. Hemos ejecutado exitosamente proyectos de **Misión Crítica** para el sector público, especializándonos en el procesamiento masivo de datos no estructurados y la automatización de procesos decisionales.

Nuestra capacidad técnica se sustenta en una arquitectura empresarial moderna, con amplia experiencia en el diseño de microservicios y soluciones **Cloud Native** sobre Google Cloud Platform. Somos **Google Cloud Premier Partners**, lo que garantiza acceso a soporte avanzado y dominio de las herramientas Vertex AI y Document AI requeridas para este proyecto.

Dominamos metodologías de gestión híbridas (Agile/Scrum y PMI), asegurando entregas iterativas de valor sin perder el control presupuestario y temporal. Nuestros procesos están certificados bajo normas **ISO 9001 (Calidad)** e **ISO 27001 (Seguridad de la Información)**, garantizando la confidencialidad e integridad de los datos sensibles de ChileCompra.

5. Soporte y Garantía

De acuerdo con las bases, ofrecemos una **Garantía Técnica de 6 meses** contados desde la recepción conforme del último entregable (Informe N°6). Esta garantía incluye:

- Corrección de incidencias (bugs) imputables al desarrollo (código, configuración, prompts).
- Solución de problemas de rendimiento o degradación del servicio.
- Ajustes por "alucinaciones" recurrentes detectadas en producción que no cumplan con los umbrales de calidad definidos.
- Atención mediante mesa de ayuda con SLA definidos para incidencias críticas (invalidantes) y no críticas.

6. Resumen Ejecutivo

[NOMBRE EMPRESA] presenta esta propuesta para convertirse en el aliado tecnológico de ChileCompra en su evolución hacia la fiscalización inteligente. Nuestra oferta combina la potencia de los **Grandes Modelos de Lenguaje (LLM)** con la robustez de la infraestructura **Google Cloud**, diseñada específicamente para manejar el volumen y complejidad de la contratación pública chilena.

Nos diferenciamos por un enfoque consultivo que no solo implementa tecnología, sino que transfiere capacidades al equipo del Observatorio. Aportamos agilidad, experiencia comprobada en proyectos de IA Generativa y un compromiso irrestricto con la seguridad y la normativa vigente. Nuestra solución es modular, escalable y auditável, asegurando que la inversión pública se traduzca en una mayor eficiencia y transparencia para el Estado.

Anexo: Experiencia Relevante

Cliente	Proyecto	Descripción Breve
Poder Judicial	Automatización de Análisis de Sentencias	Implementación de modelos NLP para clasificar y extraer entidades de 500.000 sentencias históricas. Uso de OCR y anonimización de datos.
Servicio de Impuestos Internos	Detección de Anomalías Tributarias	Desarrollo de plataforma Big Data en GCP para identificar patrones de evasión mediante análisis de grafos y machine learning.
Ministerio de Salud	Digitalización de Expedientes Clínicos	Pipeline de ingesta y estructuración de documentos médicos no estructurados utilizando Vision AI y almacenamiento seguro en nube.