PODER JUDICIAL
REPUBLICA DE CHILE

BASES TÉCNICAS

1 ANTECEDENTES

El año 2020 el Poder Judicial inició un proceso de modernización de sus Bases Jurisprudenciales con el objeto de proporcionar, tanto a usuarios internos como al público en general, un acceso a las sentencias sistematizadas de todos los tribunales del país, acercando la labor de la judicatura a la ciudadanía. Junto con este objetivo, surgió la necesidad de armonizar la regulación existente en materia de publicidad de sentencias, con los requerimientos del principio general de publicidad, la normativa sobre transparencia y protección de datos personales, con miras a propender a un adecuado acceso a la justicia.

De esta manera, teniendo como orientación que la litigación no puede significar una afectación a la dignidad y libertad de la persona que, por su ejercicio, pueda ver expuestos sus datos personales o sensibles, y con la finalidad de proteger tales condiciones, el Pleno de la Corte Suprema dictó el Acta 44-2022. Esta Acta entró en vigencia el 01 de julio de 2022, surgiendo así la necesidad de anonimizar por parte del Poder Judicial un importante número de fallos para publicarlos en sus bases jurisprudenciales y mantener la actualización de aquellas. Inicialmente esta función se realizó a través de un proveedor externo que proporcionaba servicios de anonimización y un equipo interno que realizaba labores de anonimización y control de calidad.

Con el objeto de adecuar dicha normativa a las actuales necesidades del buscador de fallos, el Pleno del máximo tribunal, aprobó en agosto de 2024, un nuevo Auto Acordado sobre Acceso a Carpetas Electrónicas Judiciales y Buscador De Jurisprudencia Del Poder Judicial (Acta 164-2024), que busca una regulación simplificada que permita entre otros objetivos, mejorar el proceso de anonimización de sentencias.

El nuevo marco regulatorio permite la actualización y sistematización de los deberes institucionales en materia de tratamiento y publicación de la información contenida en sentencias y resoluciones judiciales, en lo relativo a la publicidad y protección de la vida privada de las personas, a través de las plataformas informáticas accesibles para la ciudadanía.

La nueva normativa en materia de publicación de sentencia contenida en el Acta 164-2024, que establece como restricción fundamental la protección de los datos personales, entrará en vigencia en el primer semestre de 2025. Lo anterior, unido a la aprobación por parte del poder Legislativo, el 27 de agosto de 2024, del proyecto de Ley que Regula la protección y el tratamiento de datos personales y crea la Agencia de Protección de Datos personales, contenido en el boletín N°11.144-07 (refundido con el boletín N°11.092-07); torna fundamental para el Poder Judicial contar con una herramienta que, a través del uso de distintas tecnologías, permita llevar a cabo la anonimización de datos personales de forma eficiente y así poder proporcionar bases actualizadas a la ciudadanía con la debida sujeción al nuevo marco regulatorio.

Actualmente la anonimización de sentencias se realiza a través de un proceso manual llevado a cabo por un equipo interno, quienes realizan labores de anonimización y control de calidad.

Los requerimientos y las condiciones de prestación de los servicios establecidas en estas bases son obligatorias, salvo que expresamente se establezca lo contrario, por lo tanto, serán declaradas inadmisibles las ofertas que modifiquen los requerimientos o condiciones mínimas de prestación del servicio.

Durante el año 2024 se han dictado las siguientes sentencias:

PODER JUDICIAL
REPUBLICA DE CHILE

Competencia	2024-01	2024-02	2024-03	2024-04	2024-05	2024-06	2024-07	2024-08	2024-09	2024-10	Total general
1.SUPREMA	3.025	3.218	2.223	2.089	2.557	2.097	2.477	2.303	1.444	1.953	**23.386**
2.APELACIONES	13.021	11.245	12.203	16.479	13.291	11.572	13.662	14.210	12.598	21.457	**139.738**
3.LABORALES	1.729	1.460	1.705	2.002	1.778	1.734	1.948	2.032	1.766	2.066	**18.220**
4.COBRANZAS	132	151	157	179	172	144	158	164	112	170	**1.539**
5.PENAL	7.684	6.650	7.899	9.237	8.487	7.735	8.918	8.892	7.498	9.409	**82.409**
6.FAMILIA	18.979	16.438	17.986	20.730	20.087	17.423	20.214	20.444	17.380	21.484	**191.165**
7.CIVIL	4.681	3.469	3.791	4.456	3.969	3.196	3.898	3.704	2.798	2.754	**36.716**
Total general	49.251	42.631	45.964	55.172	50.341	43.901	51.275	51.749	43.596	59.293	**493.173**

Cada sentencia tiene un número de páginas variable.
* Para el caso de Corte Suprema y Cortes de Apelaciones, esta estadística no incluye las sentencias sobre recursos de protección de ISAPRE.
Estadística completa en: https://numeros.pjud.cl/Inicio.

2 OBJETIVO GENERAL

El objetivo es crear un sistema informático que funcione de manera local en la plataforma del Poder Judicial y que permita la anonimización de datos personales o confidenciales contenidos en sentencias judiciales y otros documentos.

El sistema debe permitir anonimizar la información identificada o identificable de personas naturales o jurídicas involucradas en causas judiciales.

La aplicación debe ser capaz de procesar automáticamente el contenido del texto para anonimizarlas y garantizar la privacidad de los datos.

Además, se pretende complementar la solución con la implementación de un sistema de flujo de trabajo documental (workflow) en el que intervengan operadores humanos antes de la publicación de las sentencias judiciales anonimizadas.

Se hace presente que la empresa Replai SpA efectuó una consultoría cuyo objeto fue evaluar la situación actual del sistema de anonimización, la cual se acompañará al presente proceso para efectos de informativos y de transparencia. En caso de contradicciones entre la información contenida en dicho informe y la establecida en estas bases de licitación, prevalecerá estas bases de licitación.

3 OBJETIVOS ESPECÍFICOS

El sistema informático de anonimización deberá:

A. Anonimizar los documentos que se carguen, especialmente sentencias judiciales.
B. Realizar el procesamiento con OCR (reconocimiento óptico de caracteres), IDP (*Intelligent Document Processing*) u otras tecnologías equivalentes, para reconocer en documentos textos, imágenes y textos contenido en imágenes.
C. Implementar flujos de trabajo de calidad y llevar a cabo el entrenamiento de modelos de aprendizaje para detectar entidades.
D. Obtener luego del procesamiento de anonimización la salida de 3 formatos: archivo pdf, archivo formato docx anonimizados y archivo xml con lista de ocultación.
E. Almacenar los documentos anonimizados en la base de datos del Poder Judicial con estándares de integridad y confidencialidad, lo cual será fundamental para garantizar la seguridad de la información.
F. Permitir el control de versiones del documento anonimizado y permitir la edición del documento anonimizado en el futuro.
G. Integrar este nuevo sistema con el sistema denominado “Mantenedor”, el que administra las asignaciones y control de calidad para el flujo de anonimización. Las características técnicas del Mantenedor están indicadas en el Anexo N° 2. En adelante se referirá a dicho sistema como el “Mantenedor”.
H. Emplear en el proceso de anonimización mecanismos automatizados con herramientas y algoritmos de última generación en *Machine o Deep Learning*, procesamiento de lenguaje natural y de código abierto, que permita identificar los datos PII, que significa *Personal Identifiable Information o* Información Personal Identificable en español, donde estos se refieren a datos que pueden ser utilizados

por sí solos o en combinación con otros para identificar, contactar o localizar a una persona en un contexto determinado.

I. Construir los siguientes módulos:
   a. Reconocedor automático de entidades nombradas.
   b. Gestor de sets de entrenamientos (conjunto de entrenamiento) desde base de datos o archivos.
   c. Gestor de modelos de aprendizaje para anonimización para control de versión, entrenar, reentrenar y publicar uso de modelos (deploy).
   d. Anonimizador masivo de sentencias u otros documentos con la utilización de los modelos publicados.
   e. Editor de post-procesamiento de control de calidad (para corregir marcajes erróneos).
   f. Mantenedor de asignaciones del flujo de trabajo del proceso de anonimización.

J. Identificar como anonimizables y anonimizar los siguientes datos:
   a) Datos relativos a personas naturales:
      - Nombres y apellidos.
      - Números de identificación, como RUN, RUT, pasaporte, DNI (Documento Nacional de Identidad), SSN (Social Security Number), etc.
      - Números de teléfono o móvil.
      - Dirección de la vivienda incluyendo nombre de calle o pasaje y numeración. Queda excluida la comuna y la ciudad (se anonimiza cualquier dato más específico que la comuna, como el condominio, la villa, el barrio, etc.)
      - Código postal
      - Coordenadas GPS
      -Matrícula de vehículos.
      - Números o datos de causas tramitadas ante la administración pública.
      - Fechas de nacimiento (día y mes).
      - Correo electrónico.
      - Nombres de usuario / usuaria (en redes sociales, correos electrónicos, en definitiva, cualquier sistema informático que requiera registro).
      - Números de cuentas bancarias o tarjeta de crédito o débito.
      - Identificador de dispositivo móvil.
      - Dirección IP.
      - Identificadores biométricos (huellas dactilares, firma u otros).
      - Números de informes periciales.
      -Hipervínculos que conduzcan a sitios que contienen datos anonimizables.
      - Fotografías que contienen la imagen de una persona natural.
      - Imágenes que contengan datos PII (*Personal Identifiable Information*) o Información Personal Identificable.
   b) Caratulados de sentencias: en el caso de los caratulados se deben anonimizar aquellos que tengan nombres y/o apellidos de personas naturales.

K. Excluir de la anonimización los siguientes datos:
   a) Números de identificación judicial, como el Rol, Ruc o Rit.
   b) Datos relativos a organismos e instituciones públicas.
   c) Datos relativos a personas jurídicas públicas y privadas.[1]
   d) Nombres y/o apellidos, siempre y cuando comparezcan en calidad de su cargo o desempeños de sus funciones de: - Ministros/Ministras y jueces/juezas. - Autoridades (Nacionales e internacionales). - Funcionarias/funcionarios públicos. - Abogados/abogadas y procuradores/procuradoras. - Auxiliares de la administración de justicia. - Autoras/autores referidos en el contexto de doctrina citada.

L. Sustituir los datos anonimizados de la siguiente forma:

[1] Existen excepciones establecidas en los cuales si corresponde anonimizar los datos de personas jurídicas.

a) Nombres y apellidos: por otro nombre y/o apellido al azar dentro de una tabla dependiendo del género de la persona (nunca por uno que contenga el propio nombre y apellido real), entre corchetes. Ej. [María]
b) Apodos: por otro apodo al azar dentro de una tabla dependiendo del género de la persona, entre corchetes.
c) Números: secuencialmente, como, por ejemplo: NUM000, NUM001, NUM002, etc.
d) Direcciones: en solo una secuencia todos sus componentes (calle, numeración, portal, piso, villa, etc.) de forma secuencial, como por ejemplo: DIRECCION000, DIRECCION001, DIRECCION002, etc.
e) Personas jurídicas: se sustituye de forma secuencial como por ejemplo: PERSONA_JURÍDICA000, PERSONA_JURÍDICA001, PERSONA_JURÍDICA002, etc.
El sistema deberá permitir la configuración y modificación de los datos sustitutos previamente mencionados.

M. Los módulos contemplados en el literal H., deberá permitir que los/las administradores/as de la plataforma puedan crear nuevos modelos o extractores para identificar o reconocer nuevas entidades nombradas, establecer un tipo de anonimización o crear nuevos tipos o subtipos de tal forma de poder anonimizarlas, sustituir sus datos o excluirlos de la anonimización.

4 REQUERIMIENTOS ESPECÍFICOS

4.1 Descripción de las soluciones integrales requeridas:

La solución desarrollada debe estar conformada con los siguientes módulos:

a. Reconocedor automático de entidades nombradas:

- Los modelos de marcaje de entidades debe soportar la técnica de detección de correferencias entre entidades similares, donde el proceso de entrenamiento debe detectar estas variantes o textos pertenecen a la misma entidad y determinar las entidades principales a las cuales están ligadas.

  El contratista deberá alcanzar la tasa de éxito ofertada para la detección de entidades, como herramienta de medición se utilizará la métrica F1 *Score.*

  Para lo anterior, el contratista deberá proporcionar una muestra procesada en este módulo de 100 fallos por instancia (Corte Suprema, Cortes de apelaciones, Competencia del laboral, Competencia de cobranza laboral, Competencia penal, Competencia de familia, y Competencia civil) que deberán contener el marcaje de entidades contempladas en el formulario anteriormente mencionado.

  El formato de las sentencias con marcaje de entidades deberá ser solo texto anonimizado y archivo xml con marcaje anonimizado y lista de ocultación.

  La Mesa Técnica efectuará un control de calidad de las sentencias entregadas, pudiendo procesar otras para confirmar las métricas. En esta etapa el contratista deberá cumplir con las tasas de éxito comprometidas en la oferta, para así pasar a las siguientes etapas/hitos que establecen las bases de licitación.

- Capacidad para procesar individual o masivamente documentos principalmente en idioma español. Los documentos pueden tener distintos tamaños en bytes, incluir imágenes, logotipos, códigos de barra y QR, enlaces a páginas webs, etc.

- Creación de conjuntos de entrenamiento confiables (golden dataset[2]) para entrenar modelos de marcajes de entidades anonimizables específicos por instancias o por tipo de entidad que se desea detectar. Se refiere a identificar ejemplos de sentencias anonimizadas para el proceso de entrenamiento de nuevos modelos de marcaje de entidades para la anonimización. Por ejemplo: en sentencias civiles establecer un modelo especial para sentencias de cambio de nombre.

  En caso de que ciertas entidades no tengan ejemplos reales en las sentencias entregadas por el Poder Judicial, la empresa deberá crear y proponer datos sintéticos[3] que puedan ser utilizados en las etapas de entrenamiento y validación.

- Implementar un editor para marcar y desmarcar entidades de forma manual para corregir el marcaje automatizado NER (Named Entity Recognition) y estos reemplazos podrán ser masivos dentro del mismo documento, como un reemplazar todos con validación (reemplazo asistido e interactivo). Cada entidad debe ser destacada con un color diferenciado por tipo de entidad. Por ejemplo: Los nombres de personas naturales en un color y las direcciones en otro color.

- Sustitución de entidades: La herramienta debe permitir parametrizar previamente el tipo de sustitución de las entidades detectadas, esta podrá ser mediante uso de Diccionarios preconfigurados compuestos por listas de nombres, apellidos, apodos, calles u otros valores.

- Correferencia: La correferencia consiste en sustituir uniformemente una entidad nombrada de distintas maneras a lo largo de un texto. Este requerimiento es esencial, puesto que es lo que da coherencia al documento. Si se anonimiza erróneamente una entidad a la cual se hace referencia de distintas formas, se pierde el entendimiento del escrito original. El editor debe permitir la visualización de las entidades detectadas a través de la correferencia, de tal forma de detectar rápidamente si existen errores (panel de resumen de entidades).

  La correferencia debe existir a nivel de documentos y de los fallos de la misma causa, esto es para que los datos sustitutos de un fallo de primera instancia sean los mismos que el de las sentencias de instancias superiores.

- El editor debe permitir: Agregar,[4] extender,[5] unificar,[6] eliminar,[7] cambiar[8] las entidades de un documento. Una vez, efectuada cualquiera de estas acciones, la aplicación debe permitir el reprocesamiento automático del documento para aplicar los cambios en todo el texto, por ejemplo si se reemplaza una Entidad-nombre por una Entidad-dirección, debe reprocesarse el documento y el resultado debe contemplar el reemplazo automático en todos los párrafos del documento destacados con la misma Entidad-Nombre por la nueva Entidad-Dirección. Esta funcionalidad debe permitir una previsualización y confirmación individual y colectiva de los cambios antes de ser efectuada.

- La herramienta debe permitir importar un conjunto de documentos (miles) sin marcas de entidades para crear documentos anonimizados (marcados) que sirvan de entrenamiento (golden dataset). Debe soportar multiformato de entrada como los

---

[2] Golden dataset o conjunto de datos de oro es una fuente de datos confiable, bien definida y única que se utiliza para la toma de decisiones, entrenamiento y el análisis de modelos de machine learning.
[3] Los datos sintéticos son una alternativa segura y confiable para el análisis, la experimentación, y el entrenamiento de modelos de IA.
[4] Incorporar entidades no detectadas.
[5] Ampliar el marcaje del texto perteneciente a la entidad.
[6] Unir dos o más entidades en una nueva.
[7] Desmarcar una entidad detectada erróneamente.
[8] Cambiar el tipo de entidad asociada, por ejemplo, si el sistema detectó un nombre que corresponde a una dirección, debe permitirse corregir la entidad.

siguientes: pdf, doc, docx, rtf, txt. La salida (exportar) de estos documentos debe ser un formato portable como Xml, Json u otros utilizados en herramientas de lenguaje natural.

- Control de calidad. Debe contemplar una funcionalidad que permita el control de calidad del proceso de anonimización de tal forma de verificar que el texto anonimizado cumpla con los protocolos de anonimización institucional. Esto puede ser mediante un panel de comparación de ambos documentos antes y después del marcaje (anonimización) integrado con editor para proceder a los cambios.

- El editor desarrollado debe integrarse con los flujos actuales y futuros de asignación, control de calidad que realiza el mantenedor para carga de sentencias. Debe existir un control de versión del documento mediante gitlab locales, svc locales u otro que permita comparar versiones y administrar la versión original, modificadas (versiones intermedia) y final.

- Integración con Mantenedor de enriquecimiento actual mediante la conexión a base de datos Oracle para identificar según perfil e identificación de usuario su actual asignación de sentencias anonimizables ya premarcadas con los modelos de anonimización. Lo cual se podrá realizar a través de lectura/escritura directa en las tablas de la base de datos o mediante API o servicio desarrollado por el proveedor.

  Los perfiles identificados por este sistema son: Administrador informático, administrador jurídico, supervisor de control de calidad, control de calidad, anonimizador, visita.

Perfiles asociados a la anonimización

Admin. Informático	Informático del Centro Documental que realiza la mantención del sistema y cambios en la configuración del mismo en base a los requerimientos del equipo jurídico.
Admin. Jurídico	Abogado que tiene acceso a todos los mantenedores del sistema y módulos de informe, con el fin de hacer el seguimiento del trabajo de los abogados anonimizadores, control de calidad y enriquecimiento.
Supervisor control de calidad (QC) Anonimización	Abogado que valida el control de calidad de las anonimización hechas por personal interno.
Funcionario control de calidad (QC) Anonimización	Abogado y/o técnico jurídico que realiza el control de calidad de las anonimizaciones realizadas
Anonimizador	Abogados y/o Técnicos jurídicos que poseen los permisos para crear y cargar en el sistema sentencias anonimizadas.
Visita	Permite observar el uso de plataforma en caso de una demostración y/o capacitación

b. Gestor de modelos de aprendizaje automático:

- Debe contemplar la creación de nuevos modelos de aprendizaje automático de marcaje de entidades. La herramienta debe permitir su actualización y optimización (reentrenamiento) a lo largo del tiempo para mejorar la anonimización. La empresa deberá crear en conjunto con el equipo de la unidad encargada los modelos iniciales de marcaje de entidades para el proceso de anonimización y capacitar al equipo para continuar con la actualización y mejora de éstos. Debe considerar un modelo por cada competencia y/o tipo de anonimización y/o entidad dependiendo de su complejidad.

- Debe permitir la carga de un conjunto de entrenamiento de documentos (dataset) pre-marcados en el editor de anonimización de tal forma de reutilizar conjuntos de entrenamiento anteriores para mejorarlos y corregirlos (buscar el Golden dataset).

- Debe permitir crear, entrenar, re-entrenar modelos de marcaje y optimizar sus hiperparámetros a través de buenas prácticas mediante algoritmos o herramientas de muestreo (grid search, random search, optuna, etc).

- Disponibilizar un conjunto de algoritmos para entrenamiento de modelos de aprendizaje en marcaje o reconocimiento de entidades nombradas para anonimización (BERT, CRF, SVM, CNN, RNN, etc.). También se podrán desarrollar y utilizar modelos LLM[9] locales desarrollados con el corpus de sentencias del Poder Judicial. Estos modelos serán de código abierto y se deberán indicar los requerimientos en el informe de dimensionamiento.

- Administrar las versiones de los modelos de marcaje para anonimización creados y registrar sus datos más relevantes como nombre, metadata_model_id, fecha de creación, tipo de modelo (ejemplo: Logistic Regression), ruta de modelo (ejemplo: model_v1), metadata_model_parameters (ejemplo: random_state: 42, C: 0.5, tol: 0.004, n_jobs: -1) , tiempo entrenamiento (segundos), métricas (ejemplo: accuracy, f1 score, precision, recall, etc.). Además de otros metadatos como descripción del modelo, (ejemplo: LR con Doc2Vec y filtro de entidades de nombres de personas), propietario o constructor, ruta de logs, parámetros de test (ejemplo: 85 training 15 testing), archivo data_source (ejemplo: dataset_v1.jsonl), observaciones de preprocesamiento, parámetros del modelo (ejemplo: fecha, vector_size, window, min_count, epochs, etc.) y training_data_size (ejemplo: 40355 bytes). Idealmente registrar todo esto en base de datos para registros históricos.

- Informe de comparación y optimización de modelos a través de registro de versiones y métricas obtenidas, para determinar qué modelo es mejor dado sus métricas obtenidas y los mejores hiperparámetros que permitieron su obtención para posteriormente realizar un despliegue del mejor modelo a producción. El despliegue deberá ser mediante API para lo cual deberá implementarse una colección de servicios que permitan pasar a producción una versión del modelo y probarlo masivamente, monitorear y obtener métricas.

- Los modelos de anonimización iniciales serán validados en dos etapas. La primera mediante métricas obtenidas a partir del proceso de entrenamiento y validación propios del algoritmo de aprendizaje empleado (métricas F1 *Score, Recall, Accuracy*, etc), utilizando el conjunto de sentencias entregado al proveedor (60%). Posteriormente, en la etapa de desarrollo, en ambiente de *testing,* se efectuará una validación con sentencias del proceso diario, que no hayan sido procesadas por el modelo ni el proveedor (40%). Este control de calidad del modelo será validado por la Mesa Técnica.

  Para efectuar el proceso de validación la empresa deberá tener desarrollados y disponibles los servicios web (endpoint) de anonimización por cada instancia, que retornen estas dos salidas: solo texto anonimizado y archivo xml con marcaje anonimizado y lista de ocultación.

  El volumen de datos requerido para entrenar, anonimizar y validar el F1 score mínimo / tasa de éxito de los modelos de anonimización será un set de sentencias judiciales conforme a lo siguiente:

[9] Los modelos de lenguaje de gran tamaño, también conocidos como LLM, son modelos de aprendizaje profundo muy grandes que se preentrenan con grandes cantidades de datos. El transformador subyacente es un conjunto de redes neuronales que consta de un codificador y un decodificador con capacidades de autoatención.

Competencia	N° de sentencias	ENTRENAMIENTO Y VALIDACIÓN INTERNA*	VALIDACIÓN F1 score mínimo / tasa mínima de éxito * En ambiente de testing
		60%	40%
1.SUPREMA	2.317	1.390	927
2.APELACIONES	10.394	6.236	4.158
3.LABORALES	1.564	938	626
4.COBRANZA	14	8	6
5.PENAL	15.393	9.236	6.157
6.FAMILIA	68.799	41.279	27.520
7.CIVILES	4.095	2.457	1.638
Total general	**102.576**	**61.546**	**41.030**

* Las cantidades de sentencias indicadas son aproximadas y referenciales.

** El volumen de sentencias reservadas para la validación no será proporcionado al contratista, ya que será utilizado para la validación interna del Poder Judicial.

*** Los porcentajes indicados para entrenamiento podrán aumentar, para efectos de mejorar las métricas de los modelos iniciales debido a casuísticas.

Para pasar a la etapa/hitos siguientes la tasa de éxito del/los modelos de anonimización deberá ser a lo menos de un 80%. Si existen distintos modelos, los resultados se promediarán para efectos de su medición. Además, deberá cumplir con la tasa mínima de éxito ofertada en los modelos de detección de entidades.

- Exportar modelo final compatible (con mejores hiperparámetros y métricas) a otras plataformas de Python, de tal forma de guardar, almacenar y empaquetar estos modelos para su uso futuro e implementación en producción. Estas prácticas son necesarias por las razones siguientes: a) Copia de seguridad: un modelo entrenado se puede guardar como copia de seguridad en caso de que los datos originales se dañen o destruyan. b) Reutilización y reproducción: la creación de modelos de aprendizaje automático requiere mucho tiempo. Para ahorrar costos y tiempo, resulta esencial que el modelo obtenga los mismos resultados cada vez que lo ejecute. Guardar y almacenar su modelo de la manera correcta se encarga de esto. c) Implementación: al implementar un modelo entrenado en un entorno del mundo real, es necesario empaquetarlo para facilitar la implementación.

- Automatización de técnicas de optimización en procesos de entrenamiento para obtener mejores tiempos de respuesta y precisión. La herramienta de anonimización deberá incorporar técnicas automatizadas de optimización durante los procesos de entrenamiento de modelos, con el fin de mejorar los tiempos de respuesta y la precisión. Estas optimizaciones incluirán el ajuste automático de hiperparámetros, el uso de optimizadores avanzados como Adam, y la posibilidad de ejecutar entrenamientos distribuidos. Asimismo, deberá contar con técnicas como Early Stopping y Transfer Learning para acelerar el proceso, garantizando la eficiencia del modelo. Además, se deberá implementar un sistema de preprocesamiento automatizado y pipelines de evaluación continua para asegurar la generalización del modelo, y se deberá facilitar su despliegue mediante estándares como ONNX[10].

[10] Open Neural Network Exchange (ONNX) proporciona un formato uniforme diseñado para representar cualquier framework de aprendizaje automático

- Publicación final de mejores modelos como servicio web con mecanismo de despliegue semiautomático para disponibilizar modelos de anonimización una vez validado el paso a producción en endpoint de api[11] y además que permita la integración con el editor de marcaje.

c. Anonimizador masivo

- Servicio de anonimización masiva mediante servicio web (api) que soporte el procesamiento masivo de documentos. Este servicio deberá estar preparado para el procesamiento diario de miles de documentos desde las distintas competencias (multiflujos) que lo requieran a través de mecanismos de encolamiento de tareas, cache y con control de tiempos de latencias máximos.

  La latencia de la API debe ser inferior a 500 milisegundos para la obtención de datos específicos que no se encuentren en el texto de una sentencia, sino que datos que ya fueron procesados. Esto es necesario para que la API pueda ser utilizada en aplicaciones que no requieren un rendimiento en tiempo real, pero que sí requieren una respuesta rápida, como buscadores o aplicaciones móviles.

  El tiempo de latencia para el procesamiento masivo de sentencias debe ser razonable (deseable menor a 120 segundos por sentencia). El rendimiento de la API debe ser capaz de manejar un volumen de tráfico de hasta 100 solicitudes por segundo. Esto es necesario para que la API pueda ser utilizada en aplicaciones de tamaño medio.

- Sistema de registros o bitácora de tareas realizadas, con registro de errores, panel de monitoreo del anonimizador masivo identificando la tarea realizada y metadatos del documento o sentencia procesada como crr_documento_id, texto, entidades detectadas, correferencias detectadas, etc.

- Detectar, configurar tipos de tratamiento y realizar la anonimización de imágenes. Esto permitirá automatizar casos en que se detecten documentos con imágenes que contengan datos anonimizables para aplicar cambios preconfigurados por defecto para imágenes, eliminar imágenes, sustituir imagen con una de color gris o anonimizar los datos sensibles en la imagen (enmascaramiento o supresión) y/o transcribir texto de la imagen (o aplicar OCR) y anonimizar.

- Detectar, configurar tipos de tratamiento y realizar la anonimización de datos contenidos en tablas. Esto permitirá automatizar casos en que se detecten tablas con estas características para anonimizar los datos sensibles sin perder la ubicación y estructura de la tabla.

- Configurar tipos de técnicas disponibles de anonimización para las entidades detectadas que pueden ser distintas entre ellas, esto deberá configurarse previamente. Sin perjuicio de considerar como base las técnicas necesarias para realizar la sustitución de la forma indicada en el literal l, del punto N° 3 de este documento. Algunas de éstas pueden ser: aleatorización, sustitución, enmascaramiento, entre otras, como por ejemplo:

Técnica de Anonimización	**Definición**	**Ejemplo**

[11] Application Programming Interface.

Aleatorización	Reemplaza valores sensibles con valores aleatorios, manteniendo la integridad estadística del conjunto de datos.	Cambiar los códigos postales reales de los clientes por códigos postales ficticios.
Sustitución	Reemplaza los datos originales por datos ficticios o generados artificialmente.	Sustituir direcciones de correos electrónicos reales por direcciones de correo electrónico aleatorias.
Enmascaramiento	Oculta todos los datos originales, reemplazándolos con caracteres genéricos o marcas que impidan la identificación del texto. * esta técnica sólo se permite para la anonimización de imágenes	Por ejemplo: el imputado [redacted]; La víctima XXYY

- Configurar los tipos de formatos requeridos de los archivos de salida, estos deben ser PDF y DOCX anonimizados, XML con lista de ocultación, MS Word, TXT y ARX Data Anonymization Format (DAF).[12]

d. Post-procesamiento control de calidad

- Editor de corrección de marcas erróneas de anonimización. El editor deberá permitir abrir documentos ya anonimizados en etapas de control de calidad, de tal forma de realizar cambios/ajustes posteriores al documento, para luego validar y pasar a producción, realizando los cambios de estados necesarios en las bases de datos del mantenedor.

- Integración con Mantenedor actual mediante conexión a base de datos Oracle , para identificar según el perfil e identificación de usuario su actual asignación de control de calidad.

- Integrar estados de validación, el editor deberá actualizar los estados de paso a producción o rechazo de anonimización, agregar comentarios e identificar el tipo de error detectado y mantener o visualizar todos los datos necesarios para el proceso de control de calidad y publicación.

- Bitácora de éxito de modelos (versus correcciones realizadas), deberá existir un panel que resuma las sentencias procesadas en flujo de anonimización por modelo y su porcentaje de éxito y fallas, además de identificar entidades exitosas o no, por tipo de anonimización, por competencia, por usuario que participó en validación, etc. Todo esto de tal forma de permitir evaluar algún cambio futuro a los modelos actuales y determinar qué mejoras deben realizarse.

e. Módulo para flujo de anonimización

[12] El ARX Data Anonymization Format (DAF) es un formato específicamente diseñado para almacenar datos anonimizados de manera estructurada y compatible con herramientas de anonimización.

- Módulo mantenedor de anonimización: Administra el flujo de sentencias por competencia con marcas de anonimización en sus respectivos sistemas de tramitación y asignar al equipo interno las tareas (sentencias/documentos) de anonimización y control de calidad. El sistema por desarrollar deberá integrarse y funcionar en total compatibilidad con esta plataforma, en caso que sea necesario el proveedor deberá considerar una modificación en la interfaz de usuario (MVC13) del Mantenedor. Este sistema fue desarrollado en PHP, Laravel Framework y jQuery.

Todas las librerías o componentes de la arquitectura de software deben ser de código abierto, certificadas por instituciones o fundaciones.

Las empresas deberán proponer en el hito N° 5, un plan interno de mantenimiento de las librerías o componentes de la arquitectura de software por parte de la Corporación.

La solución implementada no debe considerar el pago de licencias comerciales.

4.2 Métricas de ética en la aplicación de la inteligencia artificial

El contratista deberá adherir a la aplicación de los siguientes principios y entregará en la reunión de coordinación N° 1 un protocolo que indique como se implementarán estos principios en el sistema a desarrollar:

1. Proporcionalidad e inocuidad
2. Seguridad y protección
3. Equidad y no discriminación
4. Derecho a la intimidad y protección de datos
5. Transparencia y explicabilidad
6. Supervisión y decisión humanas
7. Responsabilidad y rendición de cuentas[14]

5 EQUIPO DE TRABAJO

Cantidad	Rol	Profesión y/o Técnico Nivel Superior, según se indique	Descripción	categoría
1	Jefe/a de proyecto	Ingeniero informático, ingeniero civil o en ejecución en informática; en computación; en computación e informática. Ingeniero Civil Industrial, Ingeniero en Computación o profesión equivalente de al menos 8 semestres.	Responsable de liderar el proyecto, establecer objetivos, asignar tareas y supervisar el progreso velando por el cumplimiento de los plazos de trabajo establecidos en la planificación.	Equipo líder

[13] Model-View-Controller.
[14] Unesco, Recomendaciones sobre ética de la inteligencia artificial, adoptada el 23 de noviembre de 2021. Disponible online.

PODER JUDICIAL
REPUBLICA DE CHILE

1	Arquitecto/a de software	Ingeniero informático, ingeniero civil o en ejecución en informática; en computación; en computación e informática. Ingeniero Civil Industrial, Ingeniero en Computación o profesión equivalente de al menos 8 semestres.	Responsable de la definición de la arquitectura de los servicios a desarrollar.	
1	Científico/a de datos	Ingeniero civil en informática, ingeniero civil matemático, ingeniero matemático, licenciatura en matemáticas, licenciatura en ciencias con mención en matemáticas, profesión equivalente en el ámbito de las matemáticas y científico de datos	Responsable de analizar los datos y desarrollar el modelo de procesamiento de lenguaje natural para la anonimización de las sentencias judiciales.	
El proveedor deberá considerar la para el desarrollo de la solución la cantidad de desarrolladores y tester que estime pertinente para el correcto avance del proyecto y el cumplimiento de los estándares de calidad requeridos en las bases y ofrecidos por el proveedor	Desarrollador/a de software Senior y Tester	El proveedor determinará los requerimientos exigidos para estos profesionales.	Responsable de implementar la aplicación de anonimización, integrando los modelos de *Deep o machine learning* y aplicación de las técnicas de procesamiento de lenguaje natural	

Dado que todas las coordinaciones técnicas, reuniones y reportes se efectuarán en idioma español, los profesionales del equipo líder deberán acreditar dominio funcional del idioma español, hablado y escrito. - La acreditación del dominio del idioma español podrá realizarse mediante cualquiera de los siguientes mecanismos: a) Certificado de estudios formales cursados en español, emitido por una institución de enseñanza secundaria o superior. b) Certificado oficial de dominio del idioma español, como DELE (Instituto Cervantes) o equivalente, nivel B2 o superior. c) Declaración jurada firmada por el oferente, adjuntando resumen curricular de cada integrante clave, con experiencia laboral mínima de un año en contextos hispanohablantes (formato libre)

6 METODOLOGÍA Y FORMA DE TRABAJO

El contratista deberá trabajar utilizando una metodología ágil, y proveer las herramientas que permitan llevar el control de las actividades del proyecto y su estado de avance, ajustándose al cumplimiento de los plazos establecidos y ofertados.

La gestión y control de todos los requerimientos le corresponderá al Jefe/a de Proyecto que se designe.
El horario de prestación del servicio será en horario hábil de Santiago de Chile.
La metodología y forma de trabajo, es parte integrante de estas bases, y contemplan todas las etapas propias del ciclo de desarrollo de software, como son: toma de requerimientos, diseño, desarrollo, control de calidad, capacitación y puesta en producción, conforme a lo indicado a continuación.

El paso de una etapa a otra se formalizará en un acta suscrita por los el/la Jefe/a de Proyecto y el/la secretario/a de la Mesa Técnica de ambas partes.

Cada una de estas etapas se detallan a continuación:

a) **Reunión de coordinación N° 1**. Se realizará una reunión de lanzamiento del proyecto una vez firmado el contrato donde el contratista entregará la documentación requerida en el numeral 37 de las bases administrativas, Durante esta reunión inicial se establecerá un calendario de reuniones periódicas, para realizar el seguimiento y la evaluación del estado de avance del proyecto. En esta reunión deberán asistir todos los profesionales presentados por la empresa junto a su oferta. El contratista levantará un acta de cada una de las reuniones que se efectúen, la cual será remitida a la Mesa Técnica dentro del plazo de 2 días para su revisión y observaciones.
Una vez aprobada la documentación se suscribirá el acta de inicio del proyecto, el cual dará comienzo al plazo de desarrollo.

b) **Generación de requerimientos**: El jefe de proyecto de la empresa y la Mesa Técnica revisan los requerimientos contenidos en el presente documento para el desarrollo del software.

c) **Diseño de interfaces de usuarios para cada uno de los módulos.** El jefe de proyecto de la empresa y la Mesa Técnica establecerán el diseño de interfaz (wireframe o maquetas) de cada uno de los módulos a implementar y sus funcionalidades internas. Además, definirán las salidas del sistema a nivel de archivos, reportes y tableros.

d) **Diseño de modelos de aprendizaje**: El jefe de proyecto de la empresa y la Mesa Técnica definirán las entidades, correferencias y relaciones a extraer. Además, se deberá proponer el diseño de un editor que pueda integrarse en los modelos para anonimizar sentencias y el Mantenedor.

e) **Desarrollo**: La empresa adjudicada deberá llevar adelante el desarrollo del sistema en sus tres ambientes:

- Ambiente de desarrollo: Ambiente en el cual se realiza la integración y despliegue de los desarrollos y para su primera etapa de control de calidad, éste es suministrado y administrado por el proveedor.
- Ambiente de prueba: Ambiente en el cual se utiliza para realizar pruebas de control de calidad sobre los sistemas informáticos en horarios de servicio, el cual es suministrado y administrado por el proveedor.

Dentro de los primeros 6 meses contados desde la suscripción del acta de inicio, el contratista deberá entregar un informe de dimensionamiento de hardware para la etapa de producción inicial y estimación de crecimiento futuro. El dimensionamiento deberá formularse en términos de resultados y requisitos funcionales más que de características de diseño o descriptivas; y basarse en normas internacionales, cuando las haya, o, a falta de ellas, en reglamentaciones técnicas nacionales, normas nacionales reconocidas. No deberá haber ningún requisito o referencia respecto de una marca o nombre comercial, patente, diseño o tipo, origen específico, productor o proveedor, a menos que no exista una manera suficientemente precisa o comprensible de describir los requisitos y siempre que expresiones tales como "o equivalente" se incluyan en el dimensionamiento. Esta plataforma deberá incorporar mecanismos de redundancia, alta disponibilidad y recuperación ante desastres, asegurando que los servicios permanezcan operativos incluso en caso de fallos en el sistema o interrupciones, permitiendo una continuidad en el funcionamiento y acceso a los datos de manera confiable y eficiente.

- Ambiente de producción: Ambiente en el cual reside el sistema informático que es utilizado por los usuarios, el cual es suministrado y administrado por la Corporación. Deberán acordarse e implementar además ambientes de desarrollo y pruebas,

Al ocupar una metodología de desarrollo ágil el proveedor deberá efectuar entregas periódicas de los productos desarrollados contemplando todas las etapas y ceremonias consideradas en este tipo de metodología.
Cada ambiente deberá tener un control de versión el cual permitirá, a través de la herramienta *Gitlab* coordinar y validar los pasos a producción de cada una de las ramas de desarrollo.
El desarrollo final deberá estar acorde al cumplimiento de los objetivos y requerimientos específicos indicados en el numeral 3 y 4 de este documento. Para pasar a la siguiente etapa/hito la empresa deberá cumplir con los requerimientos anteriores, además deberá ir haciendo entrega de los códigos fuentes.

f) **Creación de modelos:** El proveedor deberá crear los modelos iniciales de aprendizaje para la anonimización para cada competencia. Estos modelos deberán quedar operativos y validados en conjunto con la Mesa Técnica. Este proceso es iterativo, hasta lograr la tasa de éxito mínima exigida (80%) (para mayores detalles ver literal b) del sección 4.1 punto séptimo de las presentes bases)y la tasa mínima de éxito ofertada en los modelos de detección de entidades, que será verificado por la Mesa Técnica en el Hito N° 2.

Para pasar a la siguiente etapa la empresa deberá corregir las observaciones que se hayan formulado, lo que deberá contemplar dentro de los plazos de su carta Gantt.

g) **Control de calidad:**
El contratista deberá entregar un prototipo o primera versión del sistema de anonimización para efectuar los siguientes controles de calidad:

-Informático: Esta tarea considera realizar, en un ambiente de pruebas, todas las validaciones necesarias que certifiquen el correcto comportamiento de la

funcionalidad desarrollada. Es objetivo también de esta etapa identificar errores, mejoras y faltas de funcionalidad, las cuales deben ser identificadas, documentadas y entregadas al contratista para su corrección y/o desarrollo faltante. Esta etapa es responsabilidad del equipo designado para esta tarea por la Mesa Técnica.

-Jurídico: Esta tarea considera realizar, en un ambiente de pruebas, todas las validaciones de los modelos de anonimización y las funcionalidades que deberá desempeñar el equipo de anonimización y control de calidad. Esta etapa es responsabilidad del equipo designado para esta tarea por la Mesa Técnica.

El Secretario/a de la Mesa Técnica informará a la empresa los resultados de los controles de calidad.

Para pasar a la siguiente etapa la empresa deberá corregir las observaciones que se hayan formulado, lo que deberá contemplar dentro de los plazos de su carta Gantt.

Para concluir esta etapa se debe dar solución a todos los errores detectados y ajustes solicitados que tengan relación con los requerimientos iniciales, y deberá alcanzar las tasas de éxito requeridas en cada hito.

h) **Marcha blanca y ajustes en ambiente de prueba:** Se contempla una etapa denominada “marcha blanca”. Tiene por objeto que el proveedor entregue una versión final previa al paso a producción la cual será probada bajo un entorno controlado con una capacidad de procesamiento menor y un grupo de usuarios específicos, por ejemplo: aplicada a una competencia.

Para concluir esta etapa se debe dar solución a todos los errores detectados y ajustes solicitados que tengan relación con los requerimientos iniciales, y deberá alcanzar las tasas mínimas requeridas. De lo anterior, se dejará constancia en un acta que será suscrita por el Secretario/a de la Mesa Técnica y el jefe/a de proyecto. **Este periodo no deberá extenderse por más de 2 meses.**

El periodo contempla las siguientes subetapas en forma secuencial: dos semanas de anonimización y verificación de la tasa mínima de éxito requerida, dos semanas de análisis efectuado por la Mesa Técnica, dos semanas de ajustes y dos semanas de validación final.

Para pasar a la siguiente etapa la empresa deberá corregir las observaciones que se hayan formulado, lo que deberá contemplar dentro de los plazos de la Carta Gantt.

i) **Marcha blanca en producción y ajustes:** Tiene por objeto probar el sistema en un ambiente productivo, con usuarios reales, con el objeto de evaluar la carga de trabajo y su respuesta satisfactoria, sin que se presenten inconvenientes en el uso del sistema. En caso de presentarse inconvenientes deben ser corregidos en esta etapa y no en la etapa de garantía.

Para concluir esta etapa se debe dar solución a todos los errores detectados y ajustes solicitados que tengan relación con los requerimientos iniciales, y deberá alcanzar la tasa de éxito mínima y la ofertada en los modelos de detección de entidades. De lo anterior, se dejará constancia en un acta que será suscrita por el Secretario/a de la Mesa Técnica y el jefe/a de proyecto. Este periodo no se extenderá más allá de 2 meses.

El periodo contempla las siguientes subetapas en forma secuencial: dos semanas de anonimización y verificación de la tasa de éxito requerida y, dos semanas de análisis efectuado por la Mesa Técnica, dos semanas de ajustes y dos semanas de validación final.

Una vez concluida esta etapa se entenderá concluido el Hito N° 5, y se dará inicio al Hito N° 6.

j) **Puesta en producción final (Sistema de anonimización final integrado a sistema Mantenedor de anonimización)**: La puesta en producción final, se ejecutará toda vez que la etapa de control de calidad informático y jurídico valide las funcionalidades requeridas. Será responsabilidad de la empresa proveedora, con el apoyo la Mesa Técnica, el paso de producción final del sistema desarrollado y sus componentes, integrando el software desarrollado.

k) **Documentación y código fuente**: El proveedor deberá entregar antes de la puesta en producción, a conformidad de la Mesa Técnica lo siguiente:
- Manual del usuario, video tutoriales o sesiones de capacitación y grabar las sesiones de transferencia del conocimiento.
- Especificación de mecanismos de autenticación/autorización de usuarios y niveles de acceso a sistemas.
- Documento de Modelamiento de Datos (modelo y diccionario de datos por módulos del sistema)
- Informe técnico de administración de modelos de aprendizaje, optimización, reentrenamiento y publicación para asegurar su continuidad.
- Manual del administrador de la solución. Este deberá considerar los aspectos técnicos de código fuente, el funcionamiento del sistema, bases de datos, de errores frecuentes y su solución.
- Entrega de código fuente comentado.

l) **Capacitación**: En esta etapa se deberá efectuar el plan de capacitación y transferencia de conocimiento ofertado y se deberán entregar los documentos ahí requeridos. Las capacitaciones se deberán efectuar en los días siguientes del paso a producción.

La gestión y control de todos los requerimientos generados se efectuará mediante una carta Gantt, la cual serán actualizadas con la información entregada por cada uno de los participantes de cada etapa del desarrollo. La responsabilidad de su actualización corresponderá al jefe/a de proyectos de la empresa.

7 GARANTÍA TÉCNICA

El desarrollo tendrá una garantía mínima de 6 meses contados desde la puesta en producción final. El oferente podrá ofrecer un plazo superior. El costo de esta garantía está incluido en el precio total ofertado.

Si durante dicho periodo se presentan errores, fallas en la solución desarrollada, será responsabilidad del contratista efectuar las modificaciones y correcciones necesarias, de acuerdo con las indicaciones de la Corporación.

La garantía técnica se administrará por un sistema de tickets o mecanismo similar que permita el registro, control y seguimiento del reporte.

La garantía técnica deberá estar disponible para el ingreso de requerimientos de lunes a viernes excepto festivos, en horario de 8:00 a 18:30 hrs.

Tener presente que los requerimientos y condiciones de prestación de los servicios son obligatorios, salvo que expresamente se establezca lo contrario, por lo tanto, serán declaradas inadmisibles las ofertas que modifiquen los requerimientos o condiciones mínimas de prestación del servicio.

7.1 Tiempos de respuesta, diagnóstico y solución

Una vez generado el reporte el Contratista deberá responderlo en los siguientes plazos:

a) Respuesta inicial: 1 hora en horario continuo desde que se efectuó el reporte.

b) Diagnóstico: 3 horas continúas considerando horario y días hábiles y no hábiles, sin perjuicio de lo anterior a solicitud del contratista y en casos justificados la Corporación podrá aceptar períodos de hasta 6 horas en horario hábil.

c) Solución: A falta de acuerdo entre las partes, los requerimientos deberán solucionarse en el plazo máximo de 12 horas continúas considerando horario y días hábiles y no hábiles. Este plazo se inicia una vez comunicado el diagnóstico. A solicitud previa al vencimiento del plazo del contratista y en casos justificados la Corporación podrá conceder plazos superiores. Además, en caso de que se requiera, la Corporación podrá exigir a la empresa un plan de trabajo para la solución del problema. Dicho plan deberá entregarse el día hábil siguiente de efectuada la solicitud. Éste será analizado por la Corporación en relación con el tiempo de ejecución de las tareas, plazo de solución del problema y los recursos asignados.

Los requerimientos efectuados por la Corporación fuera de horario hábil se entenderán efectuados a las 8:00 hrs. del día hábil siguiente.

Serán incidencias graves:

- Sistema dejó de funcionar.
- Perdida de información
- Vulneración del sistema o fuga de información
- Sistema presenta errores bloqueantes que impiden su uso.
- No permite subir o ingresar información.
- Informes con errores.
- Sistema o funcionalidad no opera o no genera movimientos.
- Problemas de autenticación y acceso al sistema y sus opciones.

7.2 Mantenimiento preventivo

Durante el periodo de garantía técnica la empresa deberá prestar mensualmente el servicio de mantención preventiva para:

- Actualizaciones de software: Mantener actualizadas las librerías que utiliza el sistema de anonimización para garantizar que tengan las últimas correcciones de seguridad y mejoras de rendimiento.

- Respaldo de datos: Realizar copias de seguridad periódicas de los datos importantes para evitar la pérdida de información en caso de fallos en el sistema. En conjunto con la Mesa Técnica se definirá una estrategia para respaldo completo, diferencial e incremental de los datos y los aplicativos.

- Optimización del sistema: Limpiar archivos temporales y realizar otras tareas de optimización para mejorar el rendimiento del sistema.

- Monitoreo de seguridad: Utilizar software de seguridad, como antivirus y firewalls, y realizar análisis de seguridad periódicos en el código fuente para detectar y prevenir posibles amenazas (sonarQube y similares), el que será provisto por la Corporación.

- Optimización de red: Configurar y mantener la red informática para garantizar un flujo de datos eficiente y seguro dentro del sistema.

Estas son las actividades mínimas para desarrollar, el oferente podrá ofrecer en su oferta actividades complementarias de mantenimiento preventivo.

Estas actividades deberán coordinarse con el Departamento de Informática y Computación de la Corporación para las licencias de sistema operativo, base de datos y antivirus. Las licencias de sistema operativo, base de datos y antivirus serán proveídas por la Corporación.

7.3 Exclusión de garantía

El contratista no se encuentra obligado a:

a) Proveer soportes para dispositivos y software que no son provistos por éste o problemas con los sistemas que son causados por tales aplicaciones.

b) Proveer soporte cuando el software ha sido modificado sin autorización del contratista.

c) Proveer soporte por daños provocados por incendio, virus, alzas de voltaje u otros eventos que escapen del control del contratista, salvo que fueren imputables al mismo, a sus dependientes o agentes.

8 ACTIVIDADES PERMANENTES

8.1 Reuniones e Informes

a) Entrega de un informe o estado de avance de la gestión realizada, por cada sprint o etapa, a la Mesa Técnica o cuando ésta lo solicite.
b) Reuniones de trabajo con otros usuarios o profesionales de otras unidades que determine la Mesa Técnica, para la planificación, coordinación, estudios o cualquier actividad inserta en el ámbito que trata la presente licitación.
c) Reportes de diseño o avances del sistema (Sprint, Planning, Daily, Retrospective).
d) Reportes periódicos de los modelos de aprendizaje con gráficas de tablas de confusión, histogramas

Corresponderá al jefe de proyecto de la empresa tomar apuntes, documentar, hacer seguimiento de compromisos y generar actas de reuniones.

8.2 Desarrollo del sistema

a) Diseño y desarrollo del frontend de la aplicación, según lo especificado en el presente proceso. Análisis y diseño de la aplicación, generando documentación del modelo de procesos, modelo de datos, casos de uso, diagramas de secuencia, diagramas de clases, en los casos que sea necesario y/o la Corporación los requiera.
b) Desarrollo y construcción del backend de la aplicación, según sea requerido, de acuerdo lo especificado en el presente proceso.
c) Ambiente de prueba. El procedimiento de entrega de productos deberá ser en un ambiente de prueba, donde se garantiza su correcto funcionamiento. Cada producto será entregado al secretario/a de la Mesa Técnica o personal designado, quienes verificarán el cumplimiento de los requerimientos dentro del plazo acordado en la Carta Gantt. El adjudicatario deberá corregir las observaciones en el plazo acordado en la Carta Gantt. Una vez aprobados en el ambiente de prueba, se definirá el paso a producción, aunque la Mesa Técnica puede solicitar reingresar el producto para correcciones o mejoras de las funcionalidades ya establecidas.
d) Control de Calidad: Definición y ejecución del Plan de Prueba que contiene documentación de evidencia de pruebas, pruebas de regresión y pruebas

de rendimiento y/o estrés. Se requiere, además, que el oferente disponga de herramientas para seguimiento de bugs/errores.

e) Soporte post-instalación durante la ejecución del contrato. En caso de que el producto entregado fallare en el ambiente de producción se aplicará el procedimiento establecido para la garantía técnica.

8.3 Arquitectura del sistema

8.3.1 Multicapas

El sistema deberá operar bajo una arquitectura multicapa, que considera la división de la aplicación entre tres capas básicas:

- Interfaz usuario: provee la interfaz gráfica que permite a los usuarios del sistema hacer *uso* de los servicios de información en la forma más cómoda e intuitiva posible. No se preocupa de cómo resolver el problema.
- Lógica de la aplicación: provee un conjunto de servicios (procedimientos, métodos y funciones) que permiten administrar la lógica del problema a resolver (distribuida entre el servidor de aplicaciones y la base de datos) y que son invocados desde la interfaz usuaria según lo requerido.
- Administración de los datos: almacenamiento de la información requerida por los usuarios del sistema.
- Auditoria de las transacciones: el sistema deberá contener un registro adecuado de las transacciones, de tal manera que permita obtener cualquier auditoría respecto de éstas de una forma ágil y oportuna.

La arquitectura multicapa va más allá de la programación modular ya que no sólo consiste, en dividir el software en varios módulos, sino más bien es la separación física de estas tres capas, de esta manera existe un módulo invocador: aquel que requiere el servicio, y un módulo invocado: aquel que provee el servicio, requiriendo de la tercera capa para acceder a los datos. Se debe considerar que dentro de la arquitectura multicapa existen muchas variaciones que dependen básicamente de la ubicación física de las capas, la política definida del Departamento de Informática y Computación es desarrollar los sistemas bajo una arquitectura multicapa, donde la Interfaz Usuario es vía Web (browser), la lógica de la aplicación podrá radicar ya sea en un servidor de aplicaciones como en un servidor de Base de Datos, y el almacenamiento de éstos en la Base de Datos.

Esta modalidad tiene las siguientes características:

Toda la Lógica de la Aplicación está ubicada físicamente dentro del servidor (de aplicaciones y/o de base de datos) implementada por métodos, funciones, transacciones, procedimientos almacenados, triggers, declaración de integridad referencial, restricciones en la base de datos, y otros elementos de ésta. Cuando la Lógica de la Aplicación está en el servidor (de aplicaciones y/o de base de datos) permite que los servicios se construyan libres de contexto, es decir, operan independientes de qué servicios fueron llamados previamente, permitiendo que la aplicación tenga tiempos de respuestas adecuados para la carga de trabajo. No necesita conocer qué Interfaz Usuario muestra los datos e implica además que el acceso a los datos se efectúa a través del servidor de aplicaciones evitando que en la interfaz existan instrucciones directas a dicha Base de Datos. Así mismo, la Interfaz Gráfica Web tiene el mínimo indispensable de la Lógica de la Aplicación, haciéndolo más pequeño, liviano y fácil de distribuir, teniendo sólo la lógica de navegación de pantallas, presentación de datos y validaciones físicas de los datos.

8.3.2 Interfaz web

La interfaz WEB con que trabajará el usuario, debe estar implementada con tecnología de última generación, siendo deseable que utilice como base la arquitectura Single-Page Application y que además garantice:

- Compatibilidad con los navegadores de última generación, Google Chrome, Mozilla Firefox y Safari (macOS), Microsoft Edge.
- Compatibilidad con HTML 5.
- Disminuir la cantidad de Clicks para ingresar contenido y para acceder al contenido. Orientada a mejorar la experiencia de usuario, evitando los refrescos de pantalla y evitando pérdida de información en caso de digitar algún dato incorrecto.
- Cumpla estándares de accesibilidad y usabilidad.
- Sistema responsivo para cualquier tipo de pantalla de escritorio utilizada por equipo de anonimización.

Para que un nuevo sistema se integre correctamente con el motor corporativo de base de datos Oracle , se deben cumplir dos tipos de requisitos: requisitos generales y requisitos específicos.

8.3.3 Requisitos generales

Sistema operativo: El sistema operativo del nuevo sistema debe ser compatible con Oracle . Algunas plataformas compatibles son:

- Windows: Windows Server.
- Linux: Red Hat Enterprise Linux.
- Solaris: Oracle Solaris.

En todos los casos deberán considerarse versiones estables y actualizadas.

Hardware: El hardware sugerido para el sistema de anonimización por la empresa, en el dimensionamiento que efectúe, debe cumplir con los requisitos mínimos para el controlador/cliente de Oracle . Esto varía según la carga de trabajo y la configuración, pero en general, se necesita un procesador multinúcleo, suficiente memoria RAM y almacenamiento de alta velocidad. El sistema debe estar conectado a una red que pueda soportar el tráfico de Oracle . Esto implica tener suficiente ancho de banda y baja latencia.

Software:
Cliente Oracle: Se debe instalar un cliente Oracle compatible en el nuevo sistema. Este proporciona las herramientas para conectarse a la base de datos Oracle .
Bibliotecas de Oracle: Las bibliotecas de Oracle necesarias para la integración deben instalarse en el nuevo sistema. Estas bibliotecas proporcionan las funciones para interactuar con la base de datos Oracle .
Software de terceros: Cualquier software de terceros utilizado para la integración debe ser compatible con Oracle . Esto incluye herramientas de desarrollo, middleware y aplicaciones.

Seguridad:

- Autenticación: El nuevo sistema debe poder autenticarse con la base de datos Oracle . Esto puede hacerse mediante contraseñas, tokens de seguridad u otros métodos compatibles.
- Autorización: El nuevo sistema debe estar autorizado para acceder a los recursos de la base de datos Oracle . Esto puede hacerse mediante roles, privilegios y otros mecanismos de control de acceso.
- Cifrado: Se debe utilizar cifrado para proteger los datos en reposo y en tránsito entre el nuevo sistema y la base de datos Oracle .

8.3.4 Requisitos específicos:

Además de los requisitos generales, la integración también puede tener requisitos específicos dependiendo del caso de uso. Estos requisitos pueden incluir:

- Compatibilidad de formatos de datos: El nuevo sistema debe ser compatible con los formatos de datos utilizados por Oracle . Esto incluye formatos de datos para tablas, índices, vistas y otros objetos de la base de datos.

- Compatibilidad de protocolos: El nuevo sistema debe ser compatible con los protocolos utilizados por Oracle para comunicarse con la base de datos. Esto incluye protocolos como HTTP, HTTPS y Oracle Net.

- Dependencias: Es importante identificar todas las dependencias entre el nuevo sistema y la base de datos Oracle . Esto incluye dependencias de datos, software y hardware.

- Pruebas: Es fundamental realizar pruebas exhaustivas de la integración entre el nuevo sistema y la base de datos Oracle . Esto ayudará a garantizar que la integración funcione correctamente y cumpla con todos los requisitos.

8.3.5 Librerías de machine learning

Para asegurar un uso seguro y correcto de las librerías de Machine Learning en el nuevo sistema, se deben considerar diversos aspectos técnicos, de seguridad y organizacionales. A continuación, se detallan los requisitos fundamentales en cada categoría:

8.3.5.1 Aspectos técnicos:

- Compatibilidad: El sistema debe cumplir con los requisitos de hardware y software especificados por las librerías de Machine Learning que se pretenden utilizar. Esto incluye contar con la versión adecuada del sistema operativo, la arquitectura de procesador compatible, las librerías dependientes y las herramientas de desarrollo necesarias.

- Rendimiento: El sistema debe poseer la capacidad de procesamiento y memoria RAM suficientes para manejar el entrenamiento y la ejecución de los modelos de Machine Learning. La carga computacional puede variar significativamente según la complejidad de los modelos y el volumen de datos.

- Estabilidad: El sistema debe ser estable y confiable para evitar fallos o interrupciones durante el proceso de entrenamiento o ejecución de los modelos de Machine Learning. Esto implica contar con medidas de protección contra caídas del sistema, redundancia de datos y mecanismos de recuperación en caso de errores.

- Escalabilidad: El sistema debe ser escalable para adaptarse al crecimiento en el volumen de datos, la complejidad de los modelos y la demanda computacional. Esto permite incorporar nuevos datos, entrenar modelos más grandes y atender a un mayor número de usuarios sin afectar el rendimiento o la estabilidad del sistema.

8.3.5.2 Aspectos de seguridad:

- Autenticación y autorización: El sistema debe implementar mecanismos robustos de autenticación y autorización para controlar el acceso a los datos, modelos y recursos de Machine Learning. Esto incluye la identificación y verificación de usuarios, la asignación de roles y permisos, y la protección contra accesos no autorizados.

- Protección de datos: Los datos utilizados para entrenar y ejecutar los modelos de Machine Learning deben estar protegidos contra accesos no autorizados,

divulgación accidental o alteración intencional. Esto implica implementar medidas como cifrado de datos, control de acceso basado en roles, auditoría de actividades y protocolos de seguridad para la transferencia de datos.

- Gestión de vulnerabilidades: El sistema debe contar con mecanismos para identificar, evaluar y mitigar las vulnerabilidades de seguridad asociadas a las librerías de Machine Learning y al entorno de software en general. Esto incluye la actualización regular de las librerías, la aplicación de parches de seguridad y la realización de pruebas de penetración para detectar vulnerabilidades.

- Monitoreo y registro de actividades: El sistema debe implementar mecanismos para monitorear y registrar las actividades relacionadas con el uso de las librerías de Machine Learning. Esto permite detectar comportamientos anormales, identificar posibles ataques y facilitar la investigación en caso de incidentes de seguridad.

8.3.5.3 Aspectos organizacionales:

- Definición de políticas: La empresa debe aplicar políticas claras que regulen el uso de las librerías de Machine Learning, incluyendo aspectos como la selección de librerías, el manejo de datos, la protección de la privacidad, la responsabilidad por los resultados de los modelos y la ética en el uso de la tecnología.

- Capacitación del personal: El personal PJUD involucrado en el desarrollo y uso de las librerías de Machine Learning debe recibir capacitación adecuada sobre los aspectos técnicos, de seguridad y éticos de la tecnología. Esto permitirá tomar decisiones informadas, utilizar las herramientas de manera responsable y mitigar riesgos potenciales.

- Gestión de riesgos: La empresa debe implementar un proceso de gestión de riesgos para identificar, evaluar y mitigar los riesgos asociados al uso de las librerías de Machine Learning. Esto incluye considerar aspectos como la confiabilidad de los datos, el potencial de sesgos en los modelos, el impacto en la toma de decisiones y las posibles consecuencias legales o éticas.

- Supervisión y auditoría: La empresa debe establecer mecanismos de supervisión y auditoría para verificar el cumplimiento de las políticas, identificar desviaciones y garantizar el uso responsable de las librerías de Machine Learning. Esto puede incluir revisiones periódicas, auditorías externas y la implementación de mecanismos de denuncia.

Al cumplir con estos requisitos técnicos, de seguridad y organizacionales, la empresa garantizará un uso seguro, correcto y responsable de las librerías de Machine Learning en sus sistemas, maximizando los beneficios de esta tecnología mientras se minimizan los riesgos asociados.

8.4 Viabilidad a largo plazo

El oferente deberá contemplar:

- El uso de programas en su última versión estable y actualizada no más allá de 3 años. El lenguaje y las herramientas utilizadas deben tener una antigüedad igual o superior a 5 años con mantención o actualizaciones permanentes.

- Que el sistema quede documentado de tal forma de asegurar su mantención futura por la Corporación.

- El código fuente deberá quedar respaldado en repositorios locales del Poder Judicial para el control de versiones como Gitlab u otro, de tal forma de conocer la evolución de éste a lo largo del desarrollo del proyecto.

- El código fuente deberá cumplir con los estándares de modularidad, [15] legibilidad y mantenibilidad.[16]

- La información debe quedar en formatos abiertos (libre uso) de tal forma que permita su migración a otras herramientas.

- Que el sistema sea escalable, de tal forma de tener comportamientos eficientes sin disminuir su rendimiento al aumentar la cantidad de datos a procesar. Se deberán evitar estructura de datos innecesariamente grandes o algoritmos ineficientes, minimizando siempre el uso de recursos como memoria y CPU/GPU.

- Garantizar la seguridad del sistema. El sistema deberá estar preparado para ataques intrusivos, de tal forma de validar las entradas (validación de datos de entrada, bloqueo de ataques de inyección SQL, bloqueo de ataque cross site scripting xss, etc.), administrar las sesiones con tiempo de expiración, controlar el acceso mediante cuentas de autenticación con validación segura o cifrada, protección de los datos mediante mecanismos de implementar el mínimo privilegio mediante perfiles.

- Uso de protocolos de seguridad (TLS1.3 o superior/HTTPS) en la implementación de servicios Web y en la aplicación de anonimización.

- Medidas de seguridad: el aplicativo debe incluir control de acceso, cifrado de contraseñas y sesiones, gestión de errores y registros de transacciones (logs), respaldo y recuperación de datos.

8.5 Dimensionamiento de la plataforma tecnológica y de operación

El oferente deberá dimensionar la plataforma de hardware y software requerido para la operación del sistema.

> Dentro de los primeros 6 meses contados desde la suscripción del acta de inicio, el contratista deberá entregar un informe de dimensionamiento de hardware para la etapa de producción inicial y de crecimiento futuro. El dimensionamiento deberá formularse en términos de resultados y requisitos funcionales más que de características de diseño o descriptivas; y basarse en normas internacionales, cuando las haya, o, a falta de ellas, en reglamentaciones técnicas nacionales, normas nacionales reconocidas. No deberá haber ningún requisito o referencia respecto de una marca o nombre comercial, patente, diseño o tipo, origen específico, productor o proveedor, a menos que no exista una manera suficientemente precisa o comprensible de describir los requisitos y siempre que expresiones tales como "o equivalente" se incluyan en el dimensionamiento. Esta plataforma deberá incorporar mecanismos de redundancia, alta disponibilidad y recuperación ante desastres, asegurando que los servicios permanezcan operativos incluso en caso de fallos en el sistema o interrupciones, permitiendo una continuidad en el funcionamiento y acceso a los datos de manera confiable y eficiente.

---

[15] Modularidad: dividir el código en módulos cohesivos y de responsabilidad única, con bajo acoplamiento entre ellos para minimizar los cambios. Se deben utilizar abstracciones adecuadas para separar la implementación de la interfaz y facilitar la reutilización del código.

[16] Legibilidad y mantenibilidad del código: utilizar nombres descriptivos, agregar comentarios claros y concisos, mantener un formato consistente según las convenciones de estilo como PEP 8, y evitar la duplicación de código mediante la escritura de funciones reutilizables y técnicas de refactorización.

Para efectos del diseño de la arquitectura, se deberán considerar aspectos de redundancia y carga del sistema, de manera de garantizar el rendimiento y la alta disponibilidad del servicio. Para el dimensionamiento de la infraestructura, se debe tener en cuenta el rendimiento de la solución, a partir de las especificaciones de carga y desempeño que considere la transferencia de datos, número de usuarios (100 usuarios aprox. a nivel nacional) y nivel de concurrencia.

8.6 Ambientes en que debe ser desplegada la solución

La solución debe ser instalada en 3 ambientes, cada uno con su propia base de datos y servidor de aplicaciones, con el siguiente propósito:

- Ambiente de desarrollo: para realizar los cambios y ajustes en el sistema de anonimización (programación modelo vista controlador) y obtener versiones que quedarán disponibles en el ambiente de prueba.
- Ambiente de prueba: para tener un ambiente estable que permita realizar pruebas y uso controlado por parte de la Mesa Técnica o usuarios que la Mesa determine, de tal forma de validar el correcto funcionamiento de las nuevas funcionalidades del sistema. El ambiente de prueba valida la versión para un paso a producción.
- Ambiente de producción: para que quede disponible a los usuarios finales la última versión del sistema de anonimización con un comportamiento estable y actualizado con las últimas funcionalidades validadas y disponibles para su uso, quedando conectado a la base de datos en ambiente de producción.

8.7 Características adicionales de la solución

8.7.1 Tablero de monitoreo del proceso de anonimización

El sistema debe poseer un tablero de monitoreo que muestre indicadores de gestión:

a) Número de sentencias anonimizables; sentencias anonimizadas; sentencias asignadas por funcionario.

b) Tiempo de anonimización (productividad) por funcionario.

c) Sentencias asignadas a control de calidad, tasa de error por tipo de evento.

El sistema debe generar avisos vía correo electrónico, parametrizables, para notificar asignaciones y control de calidad.

8.7.2 Alertas

El aplicativo debe generar alertas visuales al momento de operar con el sistema, para ir informando y advirtiendo al usuario los cambios que se van realizando en la información. Además, advertir si un parámetro fue ingresado incorrectamente en su formato o se omite el ingreso de uno de ellos al ser obligatorio.

El sistema debe alertar a los administradores de errores o fallas del sistema mediante correo electrónico y un tablero de monitoreo al interior del sistema para el perfil administrador.

8.7.3 Bitácora de errores y funcionamiento

La implementación del nuevo sistema debe proveer de un sistema de bitácora o registro de errores y funcionamiento. De tal forma de registrar y enviar comentarios o errores a los administradores técnicos.

Este registro debe permitir las labores de:

- Identificar fallas reportadas del aplicativo y gestionar su rápida corrección con soporte técnico/garantía del sistema.

- Identificar todas las transacciones realizadas en el sistema.

- Detectar las mantenciones de datos en caso de falla del sistema.

- Identificar mejoras del aplicativo solicitadas por usuarios y que sean pertinentes.

- Apoyo en las unidades de negocios de CAPJ, para la toma de decisiones.

- Entregar información/reportes del sistema que no es posible mediante la interfaz web (usuario) de la aplicación.

- Control de calidad de nuevas versiones y mejoras del sistema.

8.8 Métricas de Seguridad

El consultor adhiere a la aplicación de los siguientes mecanismos de seguridad:

1. Uso de protocolos de seguridad (TLS 1.3 o superior/HTTPS) en la implementación de servicios Web y en la aplicación de anonimización.

2. Medidas de seguridad: el aplicativo debe incluir control de acceso, cifrado de contraseñas y sesiones, gestión de errores y registros de transacciones (logs), respaldo y recuperación de datos.

3. Normativa Chilena (Decreto 83 publicado el 12 de Enero del 2005, del Ministerio Secretaría General de La Presidencia, que Aprueba Norma Técnica para los Órganos de la Administración del Estado sobre Seguridad y Confidencialidad de los Documentos electrónicos).