



# REMISIÓN Y ANÁLISIS DE ESCRITURAS NOTARIALES PARA LA DETECCIÓN DE FRAUDE FISCAL

***Dirección General de Tributos***

***Departamento de Hacienda, Interior y Administración Pública***

***Instituto Tecnológico de Aragón***

*Convocatoria premios Socinfo Digital Tributos TIC*

*CANDIDATURA: Aplicación de las Tecnologías Emergentes en las Administraciones  
Tributarias*



## CONTENIDO

DESCRIPCIÓN DEL PROYECTO .....	3
Antecedentes .....	3
Objetivo inicial .....	4
Evolución .....	4
Etapa 1 (2020): Primeros pasos .....	4
Etapa 2 (2022): Evolución Deep Learning.....	6
Etapa 3 (2024): Perfeccionamiento con IA Generativa y prueba concepto Copilot .....	7
REPERCUSIÓN PARA EL CIUDADANO Y LAS ADMINISTRACIONES.....	10
Repercusión .....	10
Referencias en prensa .....	11
EQUIPO DE DESARROLLO Y PROVEEDORES .....	12
VALORACIÓN ECONÓMICA .....	14
PLAZOS DE CUMPLIMIENTO .....	15



## DESCRIPCIÓN DEL PROYECTO

### ANTECEDENTES

En virtud de la Orden de Gobierno de Aragón HAP/1226/2017, de 7 de agosto, por la que se regulan las obligaciones formales de los Notarios en el ámbito de los impuestos sobre Transmisiones Patrimoniales y Actos Jurídicos Documentados y sobre Sucesiones y Donaciones, los **Notarios tienen la obligación formal de remitir** a la Dirección General de Tributos de la Comunidad Autónoma de Aragón la siguiente información:

- la **declaración informativa** de los elementos básicos de los documentos notariales por ellos autorizados.
- la remisión de la **copia simple electrónica** de los documentos notariales por ellos autorizados, referentes a actos o contratos que contengan hechos imponible sujetos a los impuestos citados
- la remisión de la copia simple electrónica de los documentos notariales por ellos autorizados, referentes a los impuestos citados, a requerimiento de los órganos tributarios, en determinados supuestos.

En el caso de la declaración informativa tiene un formato estructurado basado en *xml*. Sin embargo, la información recibida en las copias simples autorizadas, por sus propias características, está almacenada en ficheros de texto en formato PDF.

Cuando se completa la firma de la escritura, la notaría remite inmediatamente tanto la declaración informativa como la copia simple electrónica a la Dirección General de Tributos del Gobierno de Aragón. Dicha información se vincula a la autoliquidación del impuesto en el momento en el que el contribuyente completa el trámite de presentación.

Asimismo, la Dirección General de Tributos realiza periódicamente un estudio de las escrituras remitidas por los Notarios para detectar operaciones no declaradas ante la administración tributaria. Esta revisión exige minuciosidad, puesto que existen determinadas cláusulas en las escrituras notariales que condicionan la tributación de los impuestos y deben ser analizadas para determinar posibles casos de fraude fiscal.

El personal de la Dirección General de Tributos se encargaba personalmente de revisar el contenido de las escrituras notariales con el objeto de detectar este tipo de casuísticas.



## OBJETIVO INICIAL

Para conseguir que la interoperabilidad fuera lo más eficiente posible, el objetivo inicial del proyecto era **automatizar la búsqueda de patrones de texto en escrituras para generación de alarmas de posible fraude fiscal aprovechando la potencia de la inteligencia artificial**, así como el desarrollo de una infraestructura centralizada de datos (*data lake*), que almacenara toda la documentación e información relativa a escrituras públicas y fichas-resumen notariales, actas de inspección y autoliquidaciones y documentación complementaria de las mismas.

Es importante destacar que el proyecto tiene como objetivo la generación de alarmas de posible fraude fiscal, no actuaciones automatizadas. Los técnicos tributarios del Gobierno de Aragón son los encargados de analizar y evaluar dichas alarmas, promoviendo las actuaciones de comprobación que consideren necesarias en cada caso.

## EVOLUCIÓN

El proyecto se ha llevado a cabo en varias fases desde su inicio en julio de 2020, hasta su evolución tecnológica actual.

---

### ETAPA 1 (2020): PRIMEROS PASOS

La etapa 1 se inició mediante una colaboración entre la **Dirección General de Tributos** y el [Instituto Tecnológico de Aragón](#).

El **Instituto Tecnológico de Aragón (ITA)** es un Centro Tecnológico con personalidad jurídica propia, sin ánimo de lucro y cuyos fines son de interés general, legalmente constituido a iniciativa del Gobierno de Aragón en 1984 y presidido por el Departamento de Economía, Empleo e Industria del Gobierno de Aragón.

La colaboración se inició en un Marco de Cooperación cuyo objetivo era mejorar los procesos internos, la gestión de la información y la prestación de servicios públicos como



medida de impacto ante la necesidad existente de agilizar la eficiencia de las administraciones públicas y sobre todo tras la crisis del COVID19.

Este Marco de Cooperación se denomina **Unidades Mixtas de Innovación**, porque describe perfectamente la esencia de la iniciativa, crear grupos multidisciplinares entre el personal funcional y experto de la Administración Pública y equipos de tecnólogos de ITA referentes y especializados en tecnologías de Big Data, Inteligencia artificial, Open data, Blockchain, IoT, Data analytics y simulación de procesos complejos. ITA ofrece el conocimiento de su personal y una metodología de trabajo, avalada por muchos años de experiencia y casos de éxito, centrada en el rápido desarrollo de prototipos que permiten identificar las necesidades concretas del servicio y plantear las opciones tecnológicas más eficientes, económicas y de vanguardia, asegurando la viabilidad de su futura evolución y escalabilidad, y todo ello con la neutralidad tecnológica que proporciona el carácter completamente público.

La finalidad de estas unidades es identificar y analizar la información, los procesos y los servicios que se realizan en los diferentes departamentos con los que se trabaje para desarrollar rápidamente prototipos, a modo de mínimos productos viables (MVP), que permitan una rápida respuesta de la Administración ante los retos cotidianos y los que se avecinan.

Uno de los aspectos significativos del proyecto era, por supuesto, garantizar la protección de en el tratamiento y acceso a los datos personales, especialmente recogido en el contrato de colaboración.

Concretamente, durante esta etapa se desarrolló un prototipo de una herramienta de Búsqueda de patrones de texto en escrituras para generación de alarmas de posible fraude, que permite analizar la viabilidad técnica de lograr la categorización y búsqueda de patrones de texto en escrituras y la generación de alarmas para detectar posibles fraudes en las transmisiones de bienes, en particular de inmuebles, así como en otro tipo de operaciones sometidas a diferente tributación en los impuestos sobre Sucesiones y Donaciones y sobre Transmisiones Patrimoniales y Actos Jurídicos Documentados.

Para poder desarrollar dicho prototipo se desarrolló una infraestructura centralizada de datos, también denominado *data lake*, que almacena toda la documentación e información relativa a escrituras públicas y fichas-resumen notariales, autoliquidaciones y documentación complementaria de las mismas.



---

## ETAPA 2 (2022): EVOLUCIÓN DEEP LEARNING

La etapa 2, iniciada en junio de 2022, tenía por objetivo proporcionar soporte con la finalidad de que la Dirección General de Tributos pudiera asentar los supuestos casos de fraude implementados durante la etapa anterior, desarrollar una prueba de concepto o prototipo para introducir tecnologías de Inteligencia Artificial en el sistema (*Deep Learning*), y desarrollar nuevos supuestos de fraude que fueran de interés para la Dirección General de Tributos durante el segundo semestre del año 2022.

Este objetivo principal se desglosaba en los siguientes objetivos específicos:

- Validar la viabilidad de un modelo inteligente de análisis automático de texto en escrituras, y de esta forma, poder extraer información relevante y de valor para la Dirección General de Tributos.
- Analizar semánticamente los documentos proporcionados, localizando patrones de texto y extrayendo hechos imponderables de los documentos en base a las reglas que se definan para solventar los nuevos supuestos de fraude que se concreten.
- Generar una prueba de concepto, que permitiera a través de técnicas de *Deep Learning* (aprendizaje profundo) un clasificador automático y detector de riesgo en base a las reglas actualmente obtenidas durante el primer año del proyecto.
- Seguir trabajando en un repositorio único y estándar de información, que almacene los documentos proporcionados en el marco de este proyecto y desplegar en la infraestructura del Gobierno de Aragón la información para poder gestionarla y realizar el procesamiento del conjunto de documentos de Tributos seleccionados para este proyecto.
- Dar soporte a la Dirección General de Tributos, para que pueda afianzar los supuestos casos de fraude implementados durante la etapa previa.

Tecnológicamente, el proyecto apostaba en esta etapa por introducir **técnicas de Inteligencia Artificial como *Deep Learning***.

En los últimos años, *Machine Learning* (aprendizaje automático) se ha convertido en el foco estratégico de la innovación digital en el ámbito empresarial y está transformando el modo en que operan, automatizando procesos y permitiendo obtener *insights* en tiempo real.

A su vez, esta tecnología posee estrechos lazos con otro campo de estudio derivado de la Inteligencia Artificial (AI por sus siglas en inglés): el procesamiento de lenguajes



naturales (NLP por sus siglas en inglés). La combinación entre NLP y Machine Learning permite construir modelos capaces de aprender a interpretar el lenguaje humano. Una de las áreas más interesantes vinculadas con esto es el *text analysis*: analizar texto de forma automática. Es en este contexto en el que se enmarcan diferentes técnicas y modelos que permiten la extracción y clasificación de la información.

Adicionalmente, el **Deep Learning** es una técnica de aprendizaje automático que enseña a los ordenadores a hacer lo que resulta natural para las personas: aprender mediante ejemplos. Esta tecnología se presenta como clave para aprender a realizar tareas de clasificación directamente a partir de texto, imágenes o sonido. Los modelos, que pueden obtener en ocasiones precisiones superiores al rendimiento humano, se entrenan mediante un amplio conjunto de datos etiquetados y arquitecturas de redes neuronales que contienen muchas capas.

Utilizar estas técnicas para analizar escrituras permite añadir una nueva dimensión a los datos. Hace posible interpretar automáticamente acerca de qué aspectos o temáticas están hablando (**topic detection**) y qué términos específicos aparecen con frecuencia en relación con dichas temáticas (**keyword extraction**). Con los resultados de estos análisis, se obtienen *insights* que pueden mejorar drásticamente la identificación de posibles fraudes y la toma de decisiones. El objetivo de la prueba de concepto, era la clasificación automática de las escrituras en base a los documentos anotados en la base de datos. Asimismo, se tuvieron en cuenta técnicas de negación de frase para la solución ya existente.

La inclusión de técnicas de Inteligencia Artificial aumentó los requerimientos respecto al tratamiento de datos personales, materializado a través de una Evaluación de Impacto, previa al arranque del proyecto.

En este caso, se materializó mediante la figura de la [encomienda de gestión de la Dirección General de Tributos al Instituto Tecnológico de Aragón](#).

---

#### ETAPA 3 (2024): PERFECCIONAMIENTO CON IA GENERATIVA Y PRUEBA CONCEPTO COPILOT

La etapa 3 propone una evolución sustancial en el motor de comprensión de textos que se utiliza actualmente. El objetivo es adoptar y generar un modelo basado en la nueva tecnología de Large Language Models (LLMs) (se utilizaría un modelo abierto, que ITA está desarrollando en diferentes proyectos de investigación, para mantener la confidencialidad y el análisis de riesgos ya existentes), el cual consta de billones de parámetros. Esta



actualización representa un salto considerable respecto al modelo BERT existente, que se basa en la tecnología de Transformers y cuenta con 345 millones de parámetros.

La meta es avanzar hacia un modelo de mayor envergadura que ofrezca una capacidad superior en la comprensión de las escrituras. Para ello, el modelo será sometido a un entrenamiento intensivo con una diversidad de escrituras. Este enfoque permitirá optimizar la comprensión del lenguaje jurídico y, a la vez, afinar la capacidad de análisis de la herramienta.

Con una mayor comprensión de las sutilezas y la complejidad del lenguaje jurídico, el modelo mejorado estará en condiciones de reducir de manera significativa la ocurrencia de falsos positivos. Además, su sofisticación permitirá disminuir la necesidad de introducir numerosos ejemplos para su correcto funcionamiento.

El impacto de esta mejora no se limita a un aumento de la precisión en la detección de fraudes. También se espera que genere un ahorro considerable de tiempo y recursos, al mejorar la eficiencia del proceso de análisis y reducir la cantidad de falsos positivos que deben ser revisados manualmente. Así, la herramienta mejorada proporcionará un soporte más eficaz y fiable para la detección de fraudes en las escrituras.

En ese contexto, la etapa 3 del proyecto ha arrancado en el primer trimestre de 2024 teniendo como objetivo mejorar el modelo inteligente de análisis automático de texto en escrituras desarrollado en iteraciones anteriores generando un modelo abierto basado en la nueva tecnología con IA Generativa, desarrollar una prueba de concepto de Co-Piloto para la detección de Fraude.

Este objetivo principal se desglosa en los siguientes objetivos específicos:

- Mejorar el modelo inteligente de análisis automático de texto en escrituras desarrollado en iteraciones anteriores generando un modelo abierto basado en la nueva tecnología Transformers.
- Desarrollar una prueba de concepto de **Co-Piloto** para la Detección de Fraude.
- Analizar los resultados de la herramienta actual de detección de fraude actual y corrección de errores.

El desarrollo de un prototipo Co-Piloto para la detección de fraude en escrituras es un proyecto ambicioso que combina tecnologías avanzadas como el procesamiento de lenguaje natural y la inteligencia artificial con la necesidad de colaboración y manejo de documentos legales. Implica la creación de una herramienta de asistencia de Inteligencia Artificial que puede colaborar con los usuarios para identificar y prevenir fraudes en escrituras.





Basado en el modelo inteligente de análisis automático de texto en escrituras desarrollado en iteraciones anteriores, el alcance y requisitos que el prototipo Co-Piloto debería enriquecer y alcanzar son los siguientes:

- Identificación de fraudes: el prototipo debe ser capaz de analizar escrituras en busca de nuevos indicios de fraude a través de la asistencia del asistente Co-Piloto.
- Interfaz de usuario colaborativa y amigable: debe permitir a los usuarios interactuar de manera natural con el sistema. La interfaz debe ser intuitiva y fácil de usar.
- Procesamiento de lenguaje natural (NLP): el prototipo debe integrar y utilizar técnicas de NLP avanzadas para comprender y analizar el contenido de los documentos.
- Capacidad de aprendizaje automático: el prototipo debe incluir capacidades de aprendizaje automático para mejorar con el tiempo su capacidad de detección de fraudes.
- Potencia de procesamiento: se requerirá un sistema con suficiente potencia de procesamiento para manejar el análisis de documentos
- Recopilación de datos: se necesitará un conjunto de datos amplio y variado de documentos para entrenar y generar un nuevo modelo de detección de fraude basado en tecnologías *Transformers*.
- Escalabilidad: el prototipo debe ser escalable para manejar un volumen creciente de documentos y usuarios.

Asimismo, ha requerido una revisión y adaptación de la Evaluación de Impacto en materia de protección de datos personales para dar cobertura al nuevo escenario tecnológico.

Como en la etapa anterior, el proyecto se ha materializado a través de una nueva [encomienda de gestión al Instituto Tecnológico de Aragón](#).



## REPERCUSIÓN PARA EL CIUDADANO Y LAS ADMINISTRACIONES

### REPERCUSIÓN

En este caso, el salto a nivel de eficiencia en la gestión tributaria de la Comunidad Autónoma de Aragón ha sido considerable, pasando de una lectura individualizada de las escrituras notariales por parte del personal técnico a la revisión de las alarmas generadas por la herramienta. Este hecho ha permitido **incrementar el volumen de escrituras analizadas, precisando las causas específicas de fraude**. Es preciso destacar el salto en versatilidad introducido en la etapa 2, que permite al usuario de perfil tributario diseñar sus propias búsquedas sin la asistencia de especialistas de perfil informático. Pasando de un sistema basado en búsquedas tasadas previamente configuradas por el equipo multidisciplinar, a mecanismos inteligentes en manos de personal de la Administración. Lo que posibilita que, ante una nueva situación derivada de una nueva normativa o un criterio jurisprudencial reciente, el personal de perfil tributario pueda elaborar de forma autónoma una nueva búsqueda y, por tanto, se genere una **rápida adaptación a la nueva realidad existente**.

En el caso de la ciudadanía, la aplicación de Inteligencia Artificial en el análisis de las escrituras públicas ya ha permitido detectar comportamientos fraudulentos de colaboradores sociales en el ámbito de la presentación de los impuestos mencionados. Por otro lado, resulta fácil adivinar que el siguiente paso que se contempla en la utilización y desarrollo de la comprensión semántica de las escrituras que proporciona la A.I., ha de ser la elaboración y ofrecimiento a los ciudadanos de **borradores de autoliquidaciones** que faciliten el cumplimiento voluntario de sus obligaciones tributarias. De manera que el perfeccionamiento progresivo de las búsquedas de hechos imponible y de otras circunstancias con trascendencia tributaria contenidas en una escritura pública, utilizadas ahora para su contraste con la autoliquidación tributaria presentada, pueda ser igualmente utilizado, una vez determinado que se alcanza un nivel óptimo de acierto, para **proponer al ciudadano** su plasmación en la autoliquidación que ha de presentar en periodo voluntario.



## REFERENCIAS EN PRENSA

- <https://www.ita.es/actualidad/noticias/aragon-aplica-la-inteligencia-artificial-para-la-deteccion-de-fraude-fiscal/>
- <https://www.heraldo.es/noticias/aragon/2024/01/11/aragon-aplica-inteligencia-artificial-deteccion-fraude-fiscal-1703156.html>
- <https://www.elperiodicodearagon.com/aragon/2024/01/11/aragon-aplica-inteligencia-artificial-deteccion-96765883.html>
- <https://cronicaaragon.es/aragon-implementa-ia-para-combatir-el-fraude-fiscal>
- [https://www.telecinco.es/noticias/ciencia-y-tecnologia/20240111/inteligencia-artificial-gobierno-aragon-combatir-fraude-fiscal\\_18\\_011410103.html](https://www.telecinco.es/noticias/ciencia-y-tecnologia/20240111/inteligencia-artificial-gobierno-aragon-combatir-fraude-fiscal_18_011410103.html)
- [https://www.ondacero.es/emisoras/aragon/audios-podcast/zaragoza-mas-de-uno/mas-uno-zaragoza-12012024\\_2024011265a14bf367d53e0001e53642.html](https://www.ondacero.es/emisoras/aragon/audios-podcast/zaragoza-mas-de-uno/mas-uno-zaragoza-12012024_2024011265a14bf367d53e0001e53642.html)



## EQUIPO DE DESARROLLO Y PROVEEDORES

El proveedor de servicio es el Instituto Tecnológico de Aragón. El **Instituto Tecnológico de Aragón (ITA)** es un Centro Tecnológico con personalidad jurídica propia, sin ánimo de lucro y cuyos fines son de interés general, legalmente constituido a iniciativa del Gobierno de Aragón en 1984 y presidido por el Departamento de Economía, Empleo e Industria del Gobierno de Aragón que cuenta así mismo con el reconocimiento del Ministerio de Ciencia, Innovación y Universidades como Oficina de Transferencia Tecnológica.

El equipo de desarrollo de la solución está formado por los técnicos del Instituto Tecnológico de Aragón, junto con la colaboración del personal del Servicio de Administración Tributaria y del Servicio de Informática de la Dirección General de Tributos del Gobierno de Aragón:

- **Rafael del Hoyo** recibió el MSc. En la Universidad de Zaragoza se licenció en Física y se doctoró en Ingeniería Electrónica por la universidad de Zaragoza. Durante los últimos 20 años ha estado trabajando en el Instituto Tecnológico de Aragón en la aplicación de la Inteligencia artificial a diferentes sectores, en especial el industrial y el sector TIC en general. Durante este tiempo ha ejercido diferentes cargos en el Instituto, como responsable del grupo de Telemática, Coordinador de Tecnologías TIC y actualmente responsable de la línea del plan tecnológico de Inteligencia Artificial, Sistemas Cognitivos y Big Data. Durante este tiempo ha compaginado su trabajo con la docencia, en primer lugar, en la Universidad de Zaragoza, dentro del Área de Ingeniería Telemática y el Departamento de Informática e Ingeniería de Sistemas (DIIS), por 8 años. Desde el 2010 hasta la actualidad es profesor de la Universidad de San Jorge sobre Sistemas Inteligentes y tratamiento de la información. Durante los últimos 5 años es el responsable de ITA dentro de la asociación BDVA (*Big Data Value Association*) y ha participado dentro de la asociación en diferentes eventos, como organizador y experto. En la actualidad es el responsable del grupo de Inteligencia Artificial en el Digital Innovation Hub, creando y dinamizando el desarrollo de servicios entorno a las PYMES y la inteligencia artificial. Ha participado en múltiples proyectos internacionales relacionados con la Gestión de la Información y la Inteligencia Artificial, financiados por la Unión Europea en el 5º, 6º y 7º PM, H200 y el programa Eureka (Celta) y programas nacionales como el Plan Avanza o programas de Colaboración Público Privada, trabajando como IP y en sus inicios como técnico de referencia. Ha participado en múltiples proyectos con empresas privadas, tanto multinacionales como Telefónica, MasMovil, Bayer, DKV y con empresas pequeñas como MindKind, Origen o Imascono. En los últimos 5 años está involucrado como IP en más de 4 proyectos



Europeos, IP en una misión de Inteligencia Artificial, 2 CDTIs y diversos proyectos regionales para las PYMES. Aunque su primer foco es el trabajo con empresas para la transferencia de conocimiento, posee un gran número de publicaciones en revistas con gran impacto. Sus intereses de investigación incluyen la minería de datos, la inteligencia artificial, los sistemas cognitivos, la semántica y el Big Data en general.

- **Paula Peña** es Ingeniera informática y Máster en Ingeniería de Sistemas e Informática por la Universidad de Zaragoza (España). En la actualidad, trabaja como responsable de proyectos y técnica I+D en el grupo de Inteligencia Artificial, Big Data y Sistemas Cognitivos del Instituto Tecnológico de Aragón (ITA). Durante los últimos años, ha dirigido y participado en proyectos relacionados con la Inteligencia Artificial y la captura, almacenamiento, procesamiento y análisis de información con tecnologías de Big Data, tecnologías semánticas para representar la información (ontologías o grafos generales) y técnicas de minería de contenidos (perfilado de usuarios, análisis de tendencias, etc.). Además, ha contribuido en proyectos para mejorar la estructuración de la información y la estandarización y armonización de los datos, y para proporcionar soluciones y servicios destinados a permitir la explotación, el análisis y la reutilización de la información. Ha participado en congresos nacionales e internacionales y ha realizado diversas publicaciones científico-tecnológicas en congresos y revistas del ámbito de los Sistemas de Información y la Inteligencia Artificial. Desde 2004 ha colaborado y participado activamente en varios proyectos nacionales, así como en proyectos internacionales de I+D financiados por la Unión Europea. Sus principales intereses se centran en Inteligencia Artificial, Semántica, Minería de Datos y Big Data.
- **José Ignacio Calvo** es Ingeniero en Informática por la Universidad de Zaragoza, trabaja como ingeniero en investigación y desarrollo en el Instituto Tecnológico de Aragón (ITA) en el área de Big Data y Sistemas Cognitivos. Ha participado en distintos proyectos dentro del Instituto que van desde los sistemas de video asistencia basados en Software Libre hasta plataformas para logística colaborativa. Amplia experiencia en el desarrollo de sistemas distribuidos gracias a la participación en el desarrollo y mantenimiento de la plataforma Cervantes y de su mayor caso de éxito, el sistema de Gestión de Flotas de taxis de NITAX. Durante varios años he centrado el trabajo en sistemas de información geográfica basados en cartografías *opendata* y en el transporte multimodal. Ha participado en el desarrollo y evolución de la herramienta Moriarty Instituto para prototipado rápido de soluciones para Big Data. Actualmente forma parte del equipo DevOps que se encarga de despliegue de los proyectos desarrollados por el equipo de BigData.



## VALORACIÓN ECONÓMICA

En la valoración económica (impuestos incluidos) se incluye tanto el coste del desarrollo de la herramienta como la infraestructura que la soporta.

### DESARROLLO

	Periodo	Presupuesto
<i>Etapa 1: Primeros pasos</i>	2020-2021	85.000,00 €
<i>Etapa 2: Evolución. Deep Learning</i>	2022-2023	30.000,00 €
<i>Etapa 3: Perfeccionamiento IA Generativa y Copilot</i>	2024	30.000,00 €
<i>Total:</i>		145.000,00 €

### INFRAESTRUCTURA

	Periodo	Presupuesto
<i>Infraestructura entornos pre y productivo:</i>	Anual	4.926,85 €
<i>Total:</i>	2020-2024	24.634,25 €

Presupuesto total actual: **169.634,25 €**



## PLAZOS DE CUMPLIMIENTO

El plazo de ejecución de los trabajos se ha llevado a cabo desde julio de 2020 hasta su evolución tecnológica actual en 2025.

Las tareas planificadas en las diversas etapas se describen respectivamente en los siguientes cronogramas.

### ETAPA 1 (2020): PRIMEROS PASOS

Acro.	Tarea	2020 - 2021																																							
		Mes 1				Mes 2				Mes 3				Mes 4				Mes 5				Mes 6				Mes 7				Mes 8				Mes 9							
		S1	S2	S3	S4	S1	S2	S3	S4	S1	S2	S3	S4	S1	S2	S3	S4	S1	S2	S3	S4	S1	S2	S3	S4	S1	S2	S3	S4	S1	S2	S3	S4	S1	S2	S3	S4	S1	S2	S3	S4
<b>FASE 0</b>																																									
P0	Gestión del proyecto																																								
T0.1	Lanzamiento del Proyecto																																								
T0.2	Seguimiento del proyecto																																								
T0.3	Cierre del Proyecto																																								
<b>FASE 1</b>	<b>Iteraciones</b>	Iteración 1				Iteración 2				Iteración 3				Iteración 4				Iteración 5				Iteración 6																			
P1	Arquitectura e infraestructura necesaria para el proyecto																																								
T1.1	Definición de la arquitectura tecnológica																																								
T1.2	Análisis de las necesidades para el despliegue y configuración de la solución																																								
P2	Procesamiento de datos																																								
T2.1	Captura y procesamiento de fuentes de datos																																								
T2.2	Estructura de taxonomía para la clasificación																																								
T2.3	Generación del motor semántico																																								
T2.4	Generación de consultas específicas sobre la base de datos																																								
T2.5	Desarrollo de un pipeline de procesado de lenguaje natural																																								
P3	Visualización de resultados																																								
T3.1	Desarrollo de la búsqueda de patrones																																								
T3.2	Visualización de resultados																																								
P4	Integración, Despliegue y Validación del sistema																																								
T3.1	Realización de pruebas																																								
T3.2	Despliegue de la solución																																								
T3.3	Validación con personal de la Dirección General de Tributos																																								

### Entregables:

A la finalización de la actividad, ITA entregó los siguientes documentos:

- Presentaciones y actas de las reuniones
- Modelo
- Código fuente
- Documento de cierre de proyecto, con los resultados de investigación obtenidos.
- Documento de instalación/despliegue del sistema entregado
- Manual de usuario del sistema



## ETAPA 2 (2022): EVOLUCIÓN DEEP LEARNING

Acro.	Tarea	2022																																																			
		Mes 1				Mes 2				Mes 3				Mes 4				Mes 5				Mes 6				Mes 7				Mes 8				Mes 9				Mes 10				Mes 11				Mes 12							
		S1	S2	S3	S4	S1	S2	S3	S4	S1	S2	S3	S4	S1	S2	S3	S4	S1	S2	S3	S4	S1	S2	S3	S4	S1	S2	S3	S4	S1	S2	S3	S4	S1	S2	S3	S4	S1	S2	S3	S4	S1	S2	S3	S4	S1	S2	S3	S4	S1	S2	S3	S4
<b>FASE 0</b>																																																					
PT0	Gestión del proyecto																																																				
T0.1	Lanzamiento del Proyecto																																																				
T0.2	Seguimiento del proyecto																																																				
T0.3	Cierre del Proyecto																																																				
<b>FASE 1</b>																																																					
<b>Iteraciones</b>																																																					
PT1	Soporte del sistema actual																																																				
T1.1	Soporte del sistema																																																				
PT2	Prototipo funcional																																																				
T2.1	Desarrollo de un prototipo funcional																																																				
PT3	Arquitectura e infraestructura necesaria para el proyecto																																																				
T3.1	Definición de la arquitectura tecnológica																																																				
T3.2	Análisis de las necesidades para el despliegue y configuración de la solución																																																				
PT3	Procesamiento de datos																																																				
T3.1	Captura y procesamiento de fuentes de datos																																																				
T3.2	Estructura de taxonomía para la clasificación																																																				
T3.3	Generación del motor semántico																																																				
T3.4	Generación de consultas específicas sobre la base de datos																																																				
T3.5	Desarrollo de un pipeline de procesamiento de lenguaje natural																																																				
PT4	Visualización de resultados																																																				
T4.1	Desarrollo de la búsqueda de patrones																																																				
T4.2	Visualización de resultados																																																				
PT5	Integración, despliegue y Validación del sistema																																																				
T5.1	Realización de pruebas																																																				
T5.2	Despliegue de la solución																																																				
T5.3	Validación con personal de la Dirección General de Tributos																																																				

### Entregables:

A la finalización de la actividad, ITA entregó los siguientes documentos:

- Presentaciones y actas de las reuniones
- Modelo
- Código fuente
- Documento de cierre de proyecto, con los resultados de investigación obtenidos.
- Documento de instalación/despliegue del sistema entregado
- Manual de usuario del sistema





ETAPA 3 (2024-2025): PERFECCIONAMIENTO CON IA GENERATIVA Y PRUEBA CONCEPTO COPILOT

Acro.	Paquete de Trabajo	Actividad	2024											
			M3	M4	M5	M6	M7	M8	M9	M10	M11	M12		
PT0	Gestión del proyecto													
T0.1		Reunión de lanzamiento (kick-off)	◆											
T0.2		Seguimiento y reporte de actividades	◆	◆	◆	◆	◆	◆	◆	◆	◆	◆	◆	◆
T0.3		Cierre proyecto												◆
PT1	Evolución del Modelo inteligente de análisis automático de texto en escrituras													
T1.1		Adopción modelo basado en <i>Transformers</i>												
T1.2		Entrenamiento del modelo con escrituras												
T1.3		Generación de un modelo basado en <i>Transformers</i>												
T1.4		Optimización del modelo												
PT2	Dearrollar prueba de concepto de Co-Piloto													
T2.1		Desarrollo de un asistente inteligente Co-Piloto												
PT3	Análisis de resultados de uso del prototipo actual													
T3.1		Análisis de resultado de uso												
T3.2		Correcciones posibles errores												
PT4	Visualización de resultados													
T4.1		Generación interfaz de visualización de resultados												
PT5	Integración, Despliegue y Validación del sistema													
T5.1		Integración y realización de pruebas												
T5.2		Validación del sistema												
T5.3		Despliegue en infraestructura AST												

**Entregables:**

A la finalización de la actividad, ITA entregará los siguientes documentos:

- Presentaciones y actas de las reuniones
- Modelo
- Código fuente
- Documento de cierre de proyecto, con los resultados de investigación obtenidos.
- Documento de instalación/despliegue del sistema entregado
- Manual de usuario del sistema