

# Especificaciones técnicas

---

## **GENERACIÓN AUTOMÁTICA DE TEXTOS APLICADA A COBERTURA DE ELECCIONES GENERALES**

---



## **1 Introducción**

La irrupción de la Inteligencia Artificial (IA), y de las tecnologías basadas en ella, está suponiendo un hecho más que relevante en el sector audiovisual. Muchas compañías están aplicando su utilización en distintos contextos para entender su uso, comprender sus limitaciones y determinar la viabilidad de su utilización en sus distintos procesos de operación.

Desde hace algún tiempo y como parte de las actuaciones de Innovación en RTVE se está estudiando la aplicación de IA a las distintas fases de la generación de Informativos, desde las fases tempranas de detección de alertas informativas, hasta las fases últimas de presentación de las mismas a los consumidores finales. Éstas son especialmente importantes ya que se mejoraría la productividad del equipo editorial y la experiencia de los usuarios, pudiendo llegar a más personas y con más información, un factor clave en la prestación de servicio público.

La aplicación de estas tecnologías a procesos que de otra manera no podrían llevarse a cabo, como es la producción personalizada de contenidos para las poblaciones de menos de mil habitantes, abre la posibilidad de ampliar y mejorar el servicio público que ofrece RTVE, en este caso para la España vaciada.

## **2 Objeto del contrato**

Desde el año 2020 se está llevando a cabo un proyecto de investigación aplicada, S05158-20200925 *“Generación automática de textos aplicada a cobertura de elecciones municipales”* que, utilizando un sistema basado en tecnologías de IA y a partir de los datos suministrados por las fuentes de información oficiales, interprete éstos y genere noticias creadas automáticamente con texto en lenguaje natural y relativas a información electoral. El proyecto culminará en una prueba de concepto en la jornada electoral de las próximas elecciones municipales de mayo de 2023, en la que se obtendrán hasta un total de tres noticias para cada una de las poblaciones de menos de 1.000 habitantes.

Como consecuencia de la evolución del proyecto, se han podido incorporar voces a la noticia y profundizar en otros aspectos de accesibilidad en colaboración con la ONCE, de forma que las noticias podrán ser también entendidas por las herramientas de transcripción de texto a voz, para móviles y PC’s utilizadas habitualmente por personas con visibilidad reducida.

Dados los buenos resultados de los trabajos llevados a cabo hasta la fecha, en el presente expediente se plantea la realización de una nueva prueba de concepto cuyo fin será la generación automática de noticias sobre los resultados electorales de las próximas elecciones generales, a celebrar previsiblemente en diciembre de 2023, para poblaciones de menos de 1.000 habitantes como objetivo inicial, aprovechando los conocimientos adquiridos en todo el trabajo realizado hasta la fecha. Por lo tanto, el oferente a este expediente deberá disponer, en el momento de realizar la oferta, de un sistema capaz de generar noticias electorales según lo descrito en el apartado 3.5 de este Pliego.

El alcance del presente expediente incluye tanto la infraestructura tecnológica, Hardware y Software, como las tareas de Científicos de Datos, Ingenieros de IA, Editor de Automatización y Estilo Media (para el diseño de infografías/visualizaciones).

Además, este proceso se hará sin intervención personal alguna, aunque siempre estará supervisado por profesionales para certificar y verificar su calidad informativa, editorial y estética.

## **3 Descripción del servicio**

El adjudicatario llevará a cabo las siguientes tareas:

- Analizar las noticias generadas en las elecciones municipales, detectando posibles fallos o errores, deficiencias en estructura de la noticia, en los gráficos generados, textos de las voces y problemas de accesibilidad. Esta labor deberá ser realizada en el plazo de 20 días desde el inicio del contrato.
- Planificar y ejecutar herramientas y desarrollos que permitan incorporar mejoras en aquellos puntos que el análisis anterior detecte como puntos débiles. Los puntos débiles a tratar y su prioridad serán definidos por RTVE.
- Revisar el grado de accesibilidad alcanzado, desarrollando aquellas funcionalidades en las que no se haya logrado el 100% de accesibilidad.
- Realizar los cambios necesarios para adecuar el formato de las noticias a unas elecciones generales.
- Realizar una prueba de concepto (PoC) durante la jornada de las elecciones generales, consistente en la generación de al menos tres noticias para cada una de las poblaciones de menos de 1.000 habitantes. A priori, la primera de las noticias se realizará con los datos de participación cuando esta información se haga pública y el resto una vez cerrados los colegios electorales, la segunda con información parcial de resultados y la tercera con el escrutinio finalizado. No obstante, CRTVE podrá designar otros momentos de generación de las noticias si así lo considera oportuno.
- Crear una web que albergue las noticias generadas durante la prueba de concepto.
- Análisis posterior de los resultados obtenidos en las noticias generadas sobre las elecciones generales, que se plasmará en un entregable, de forma que permita un traspaso de conocimiento para proyectos posteriores similares al actual.

### **3.1 Requerimiento del servicio**

El sistema ofertado dispondrá de las siguientes prestaciones:

- Estará basado, de forma demostrable, en las siguientes tecnologías: GPT-3, GPT-J/GPT-Neox y GALACTICA.
- Se suministrarán como mínimo tres noticias por cada una de las poblaciones seleccionadas (inicialmente, para aquellos municipios de menos de 1.000 habitantes).
- Proporcionará por cada una de las noticias un texto en lenguaje natural en español y relativo a los datos adquiridos.
- Los textos en lenguaje natural deberán ser únicos en estilo y redacción a los generados para usuarios distintos a la CRTVE, la herramienta será de uso propio y personalizada a las necesidades y objetivos de la CRTVE.
- El texto resultante se conformará en forma de noticia, mediante titular diferenciado, imagen y texto, además de otros recursos gráficos, voz y vínculos y recomendaciones a otros contenidos de la CRTVE. La estructura, en estilo e imagen, de las noticias deberá ser validada por la CRTVE. Estas noticias deberán configurarse en formato HTML para su publicación en Web.
- En la noticia deberá aparecer un titular principal y los titulares secundarios que se estimen oportunos, pero nunca menos de uno.
- Las noticias deberán generarse y aparecer publicadas inmediatamente después de estar disponibles los datos electorales.
- Será responsabilidad del adjudicatario del servicio el acceso a la fuente o fuentes de datos, que serán siempre las oficiales, para asegurar la precisión y calidad de los mismos. En caso de fallo de la fuente de datos será responsabilidad del adjudicatario proveer una fuente alternativa asegurando en todo momento la continuidad, calidad y precisión del servicio.
- El adjudicatario deberá proporcionar un dominio/sitio web gestionado por él mismo para la presentación de las noticias. Dicho sitio web deberá ser capaz de:
  - Ser accesible a personas con discapacidad visual integrando en el código web las pautas proporcionadas por la ONCE.

- Almacenar en tiempo real todas las noticias generadas automáticamente por IA durante la jornada electoral.
- Incluir información visible (imágenes, video, textos) proveniente de fuentes externas al servicio o proporcionadas por la CRTVE, con el fin de enriquecer las noticias y a petición de la CRTVE.
- Almacenar gráficos generados automáticamente (en integración con un proveedor de gráficos externo).
- Almacenar audios sintéticos (generados por otro proveedor) e imágenes estándar por cada municipio.
- Procesar y soportar todas las peticiones de usuarios de cualquier lugar sin cortes de servicio.
- Ofrecer un servicio que muestre resultados y sea de carga inmediata para un rápido acceso.
- Generar un único XML en cada uno de los 3 momentos en los que se generen noticias. Este XML contendrá un texto para la transcripción de texto a voz, de menor duración que el texto de la noticia web y con el contenido adaptado para ser escuchado, para cada una de las poblaciones objeto de este expediente.
- Adicionalmente, desde el inicio del proyecto y de forma transitoria, el adjudicatario facilitará un dominio Web de acceso restringido por usuario y contraseña, para la publicación de pruebas y ensayos relacionados con el proyecto.
- El adjudicatario proporcionará todos los elementos necesarios para la publicación de los resultados en las principales redes sociales, al menos en tres de ellas, y cuentas de CRTVE que ésta considera oportuno.
- El servicio deberá proporcionar acceso a su API para posibles integraciones con terceros, siempre y cuando estos terceros hayan sido autorizados por la CRTVE y tengan su consentimiento.
- Toda la tecnología desarrollada, así como las bases de datos generadas, quedarán a disposición de la CRTVE para su utilización en coberturas electorales posteriores y la elaboración de otras noticias de ámbito diferente al electoral que se puedan enriquecer con cualesquiera de los logros de este proyecto.
- A la finalización del proyecto el adjudicatario aportará los recursos necesarios para la transferencia de todo lo almacenado durante el proyecto (noticias generadas, gráficos, ficheros de audio, fotografías...) a un nuevo almacenamiento que RTVE determine.

El texto generado a partir de los datos comunes de la fuente, que están disponibles y abiertos a medios de comunicación y a la sociedad, deberá diferenciarse de los contenidos ofrecidos por otros medios de comunicación, en la medida que no penalice en modo alguno en el posicionamiento de SEO y que, en todo caso, lo mejore al ser un contenido más original.

Dado que se trata de un proyecto de investigación aplicada y la vertiginosa velocidad con la que evolucionan las tecnologías de Inteligencia Artificial, se contemplará durante el tiempo de vida del mismo, la posibilidad de incorporar aquellas otras cuestiones de IA que permitan complementar y mejorar el alcance del proyecto y que en estos momentos no es posible definir.

Si por cualquier causa la celebración de las elecciones se retrasara, el adjudicatario se comprometerá a la prestación del servicio la jornada electoral, ajustando los plazos del expediente si fuera necesario y sin coste adicional para la CRTVE. Por el contrario, si las elecciones se adelantaran, se replanificarán los trabajos para poder acometer el objetivo en la jornada electoral.

Se formará un equipo de trabajo que estará compuesto por recursos de la empresa adjudicataria, cuyos componentes responderán a los perfiles de Científicos de Datos, Ingenieros de IA, además de diseñadores de procesos y de automatización, y personal de la CRTVE, que aportará periodistas y editores, así como personal de Innovación Tecnológica, que serán los responsables de la guía y coordinación del proyecto. Se requiere que los recursos de la empresa adjudicataria adscritos a este proyecto tengan una experiencia demostrable de dos o más años en tecnologías de Generación de Lenguaje Natural usando Machine Learning e Inteligencia Artificial.

Tal y como se indica en el Pliego de Condiciones Generales (PCG), se valorará positivamente que se hayan generado noticias en medios internacionales además de nacionales y que se haya participado en proyectos de investigación con universidades o medios públicos de comunicación.

### **3.2 Servicio**

---

El servicio se desarrollará en tres fases:

#### **Fase 1: Puesta a punto y mejora de la herramienta**

- Se trabajarán aspectos relacionados con las fuentes de datos que completarían las noticias, sobre datos demográficos de la población, históricos de elecciones, corpus de ejemplos, fotografías, etc. creando un repositorio que albergue los datos seleccionados.
- Se trabajará en el desarrollo de la herramienta propiamente dicha: estructura, aspectos de dominio, estilo-media (infografías y visualizaciones), etc.
- Se trabajarán aspectos de ajuste y puesta en marcha.
- Se realizarán un máximo de tres pruebas completas, end to end, que simulen y validen los modelos propuestos. Dada la probabilidad de que en este periodo no se celebren elecciones en nuestro país, se podrá contemplar la posibilidad de utilizar datos de elecciones anteriores o de otros países para la realización de estas pruebas.

#### **Fase 2: Jornada electoral**

- Incluye la publicación efectiva de las noticias generadas en base a los resultados de las elecciones generales previstas, a día de hoy, para diciembre de 2023.

#### **Fase 3: Análisis de resultados**

- Análisis posterior de los resultados obtenidos en las elecciones generales, de forma que permita un traspaso de conocimiento para proyectos posteriores similares al actual. Tendrá una duración de un mes a contar desde la fecha de celebración de las elecciones.

### **3.3 Tipo de servicio**

---

El servicio a contratar será en el modelo SaaS. No se permitirá la instalación de equipamiento alguno en las dependencias de la CRTVE.

### **3.4 Condiciones de aceptación del servicio**

---

En la primera fase, la aceptación del servicio estará relacionada con la buena marcha de los trabajos. En este sentido, al final de cada mes, CRTVE podrá solicitar al adjudicatario un documento, en formato Excel, que recoja los trabajos realizados en ese mes y el tiempo y recursos dedicados a cada uno de ellos.

En la segunda fase, la aceptación estará ligada al umbral de éxito en la generación de las noticias de la jornada electoral que queda definido por los siguientes aspectos:

- El nivel de publicación de noticias válidas tendrá que ser igual o mayor al 95%.
- La personalización, tanto de las noticias como del HTML, se tendrá que ajustar al modelo que en su momento la CRTVE aporte al respecto.
- La publicación de las noticias se llevará a cabo sin incidencia alguna.
- El acceso a la web donde están publicadas las noticias no se colapsará, ni se verá interrumpido en ningún momento durante el día y la noche electoral.
- No se producirá penalización en SEO.

### **3.5 Prueba para una mejor evaluación**

Tal y como se especifica en el PCG, a la vista de las ofertas recibidas y para su mejor evaluación, la CRTVE podrá requerir a todos los licitantes, una demostración del sistema ofertado, en las instalaciones de CRTVE o en otro lugar que previamente se determine, sin ningún coste adicional, donde deberá quedar demostrada la idoneidad de la solución presentada.

Para dicha demostración y el mismo día de la celebración de la misma, CRTVE proporcionará a los oferentes una BBDD con los datos de algunas de las últimas elecciones municipales celebradas y el oferente deberá:

- Generar y presentar en una página Web propiedad del oferente, en menos de 30 minutos, una noticia de todas las poblaciones de menos de 1000 habitantes de España (aproximadamente 5000 poblaciones).
- Las noticias deberán:
  - Tener un aspecto y formato similar a las noticias publicadas en la Web de RTVE.
  - Incluir fotografías diferentes que guarden relación con el municipio o con la jornada electoral.
  - No se admitirá la misma redacción para todas las distintas poblaciones, es decir, no podrán tener la misma estructura sintáctica y semántica.
  - Generar un XML único, para todas las poblaciones, con los textos de menor duración para la transcripción de texto a voz, aunque no se exigirá la inclusión de las voces en la noticia.

Los resultados deberán ser entregados al finalizar la prueba en un formato PDF o Word para al menos 50 poblaciones. RTVE determinará el nombre de las poblaciones seleccionadas.

Si el oferente no presenta los resultados de esta prueba en el plazo indicado, la oferta será desestimada.

## **4 Gestión del proyecto**

Durante la duración del contrato se establecerá un equipo de trabajo permanente en el que estarán representados, por parte de la CRTVE, periodistas especializados en Información, Edición y nuevas tecnologías, así como personal de Innovación y en el que la empresa adjudicataria estará representada por cuantos recursos considere oportunos.

El adjudicatario nombrará un responsable del servicio dentro de su equipo, que será el interlocutor único con el personal de la CRTVE para la coordinación y priorización de tareas. Así mismo, desde la CRTVE también habrá otro coordinador, que será su único contacto para cualquier duda o problema que pueda surgir.

Algunas de las actividades que realizará el equipo de trabajo serán:

- Planificación y asignación de tareas.
- Supervisión de actividades.
- Supervisión del cumplimiento de los trabajos.
- Gestión de prioridades.
- Optimización de procesos.
- Asistencia a las reuniones de coordinación.
- Elaboración de informes que permitan medir el rendimiento, la efectividad y la calidad del servicio.

Por su parte, la Dirección de Proyecto nombrada por la CRTVE tendrá las atribuciones necesarias para definir, en cada momento, las prioridades dentro del mismo, así como para convocar cuantas veces considere oportunas reuniones de seguimiento con el grupo de trabajo.

No obstante, y sin perjuicio de las reuniones que la Dirección del Proyecto pueda convocar fuera de agenda, el grupo de trabajo se citará en reuniones de seguimiento con periodicidad semanal durante la duración del

proyecto. En todos los casos la empresa adjudicataria estará encomendada de levantar acta de reunión y hacérsela llegar a todo el grupo de trabajo.

En general y salvo causas de fuerza mayor, el cauce de comunicación en el seno del proyecto será entre los coordinadores de Proyecto de ambas partes.

## **5 Otros**

Ante la posibilidad de que sean necesarios durante el periodo de vigencia del contrato desarrollos o actividades imposibles de prever en esta fase del proyecto, se contemplan unas horas de servicios profesionales variables.

Los desarrollos o actividades, de ser necesarios, se definirán en su momento y el adjudicatario presentará una propuesta específica con una planificación de los trabajos, plazos, entregables y horas que el responsable de RTVE deberá evaluar y aceptar antes de su inicio. Las horas indicadas en la propuesta serán las máximas que se certificarán una vez terminado el desarrollo o actividad.

## **6 Formato de la oferta**

Es imprescindible que la documentación presentada en la oferta técnica tenga la información que se indica a continuación:

- Relación de trabajos similares en los últimos tres años.
- Relación de instalaciones, medios técnicos.
- Relación de recursos humanos, aportando documentación que acredite la experiencia solicitada en este Pliego de condiciones técnicas.
- Relación de las tecnologías utilizadas
- Compromiso de calidad del servicio.
- Información acerca de la forma en que se podrá acceder a su API: FTP, conexión abierta por Internet, IP, etc.
- Organización del trabajo y sistema de seguridad de la calidad del servicio.
- Cumplimiento y mejoras de las especificaciones técnicas requeridas. En este sentido, se incluirá en la Oferta Técnica todos los documentos que acrediten el cumplimiento de los criterios técnicos de valoración de las ofertas descritos en el Pliego de Condiciones Generales
- Planificación de tiempos, lo más detallada posible, de las distintas fases del proyecto, incluyendo información de los recursos asignados en cada fase.