

Especificaciones técnicas

**GENERACIÓN AUTOMÁTICA DE TEXTOS APLICADA A
COBERTURA DE ELECCIONES AUTONÓMICAS DE
CATALUÑA**



1. Introducción

En el dinámico mundo del periodismo moderno, la cobertura de elecciones autonómicas representa un desafío significativo, especialmente en regiones menos pobladas o alejadas de los centros urbanos, comúnmente conocidas como la "España vaciada". En el equipo multidisciplinar formado por informadores y personal técnico liderados por la Dirección de Estrategia Tecnológica seguimos explorando la revolucionaria aplicación de la inteligencia artificial, específicamente en la generación automática de textos, como herramienta para superar dichos desafíos. La premisa central no es la sustitución del periodista, sino su empoderamiento y asistencia en tareas que pueden resultar monótonas o extenuantes, permitiendo una cobertura más amplia y equitativa del territorio a cubrir.

La inteligencia artificial, aplicada a la generación de textos, se presenta como una solución innovadora y eficiente, capaz de procesar grandes volúmenes de datos y convertirlos en narrativas coherentes y útiles. Esta tecnología no solo optimiza el tiempo y los recursos, sino que también garantiza una representación más diversa y completa de todos los municipios, incluyendo aquellos en la España vaciada. De este modo, se democratiza la información, asegurando que ningún rincón del país quede sin voz durante periodos electorales tan cruciales.

2. Objeto del contrato

El presente expediente contempla la realización de una prueba de concepto mediante un proyecto de investigación aplicada, cuyo fin será la generación automática de noticias, tanto en español como en catalán, que muestren información relacionada con la participación y con los resultados electorales de las próximas elecciones autonómicas de Cataluña, a celebrar en 2025, para poblaciones de menos de 10.000 habitantes como objetivo inicial, así como la incorporación de voces sintéticas que informen de los resultados obtenidos tanto en español como en catalán. No es objeto de este expediente la generación de las voces ni de los audios.

El proyecto se plantea tras el éxito obtenido con los expedientes, S-05158-20200925 "Generación automática de textos aplicada a cobertura de elecciones municipales" y S-01392-2023 "Generación automática texto elecciones generales" que, basándose en tecnologías de Inteligencia Artificial interpretan los datos suministrados por fuentes de información oficiales y los presentan en una noticia relativa a información electoral creada automáticamente en lenguaje natural. El proyecto se hará sin intervención humana alguna, aunque siempre estará supervisado por profesionales para certificar y verificar su calidad informativa, editorial y estética.

Dicho texto, deberá ir acompañado de otros elementos como gráficos, videos, imágenes o recomendaciones a otras publicaciones informativas o de archivo de la CRTVE, por ejemplo. Se configurará en forma de noticia, en formato HTML para su publicación en la Web de dominio www.rtveia.es, si bien en las primeras fases del proyecto, el proveedor deberá proporcionar un entorno Web de pruebas que será accesible únicamente a los usuarios de la CRTVE vinculados al proyecto a través de usuario y contraseña. Se estudiará también la posible difusión de estas noticias en las redes sociales y cuentas de CRTVE que se considere oportuno.

Además, para las poblaciones de más de 10.000 habitantes se generarán textos cuyo formato deberá ser el adecuado para que sirva de ayuda a los periodistas de RTVE que estén trabajando en noticias relacionadas con estas poblaciones. El formato de estos textos de ayuda será definido por el equipo de trabajo de RTVE junto con la Dirección de Centros Territoriales y el CPP de RTVE en Cataluña y consensuada con el adjudicatario una vez adjudicado el expediente.

No obstante, una vez adjudicado el proyecto, RTVE podrá cambiar si así lo considera poblaciones que estarán incluidas en cada bloque, siempre en función del número de habitantes.

El alcance del presente expediente incluye tanto la infraestructura tecnológica, Hardware y Software, como las

tareas de Data Scientist, AI Engineer, Editor de Automatización y Estilo Media (para el diseño de infografías/visualizaciones).

3. Generación automática de textos

El adjudicatario llevará a cabo las siguientes tareas:

- Analizar, junto con el equipo de RTVE las noticias generadas en las elecciones generales y municipales de 2023, detectando posibles fallos o errores, deficiencias en la estructura de la noticia, en los gráficos generados, en los textos de las voces y los 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 en las experiencias previas, 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 autonómicas y la incorporación de textos también en catalán. La forma en la que los usuarios accedan a la información en español o catalán y la estructura de la noticia en ambos idiomas será definida por el equipo de trabajo de RTVE una vez adjudicado el proyecto.
- Una vez desarrollada e implementada en la herramienta de generación de textos el modelo de lenguaje en catalán, hasta el día de la celebración de elecciones, y con el fin de evaluar el resultado del procesamiento del lenguaje natural en dicha lengua, se realizarán pruebas del funcionamiento del sistema en catalán consistentes en la generación de noticias de otras temáticas a determinar por RTVE y/o noticias de carácter electoral con los resultados de elecciones pasadas, aunque éstas no sean elecciones autonómicas en Cataluña.
- Realizar la prueba de concepto (PoC) durante la jornada de las elecciones autonómicas de Cataluña, consistente en la generación de al menos tres noticias en cada idioma para cada una de las poblaciones de menos de 10.000 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.
- Planificar y ejecutar herramientas y desarrollos que permitan la generación de textos de ayuda al periodista para poblaciones de más de 10.000 habitantes según el formato definido por RTVE en los mismos momentos del día de las elecciones, enumerados en el punto anterior.
- En la actualidad, las noticias sobre elecciones municipales y generales generadas con IA están publicadas en la web www.rtveia.es, cuyo soporte y mantenimiento están a cargo de la empresa que proporcionó estos servicios. Será responsabilidad del adjudicatario de este Lote todas las tareas necesarias para el traslado de la totalidad de esta web y la información que contiene a un nuevo almacenamiento gestionado por dicho adjudicatario.
- Las noticias generadas el día de las elecciones catalanas se publicarán en la web de www.rtveia.es. Será responsabilidad del adjudicatario el soporte y mantenimiento de esta Web hasta diciembre de 2025.
- Análisis posterior de los resultados obtenidos en las noticias generadas sobre las elecciones autonómicas de Cataluña, que se plasmará en un entregable, de forma que permita un traspaso de conocimiento para proyectos posteriores similares al actual. Este análisis incluirá información sobre la utilidad de los textos generados para ayuda a periodistas, márgenes de mejora, y cualquier otra valoración obtenida de los usuarios de estos textos.

3.1 Requerimiento del servicio

El sistema ofertado estará basado, de forma demostrable, en las siguientes tecnologías: GPT-4, GPT-3.5, Llama 2 y

otras tecnologías de generación de textos, y dispondrá de las siguientes prestaciones:

Para la **generación de noticias para usuarios finales:**

- Se suministrarán como mínimo tres noticias por cada una de las poblaciones seleccionadas (inicialmente, para los municipios de menos de 10.000 habitantes).
- Proporcionará por cada una de las noticias un texto en lenguaje natural en español y otro en catalán.
- El texto resultante se conformará en forma de noticia, mediante titular diferenciado, imagen y texto, además de otros recursos como fotografías y gráficos 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.
- En cada noticia deberá aparecer un titular principal y los titulares secundarios que se estimen oportunos, pero nunca menos de uno.
- Cada noticia incluirá al menos dos fotografías. A tal efecto, el adjudicatario gestionará una base de datos de imágenes proporcionada por RTVE en la que seleccionará las imágenes a incluir. En el caso de que la base de datos contenga fotografías de los distintos municipios catalanes, las noticias incluirán la foto del municipio de que se trate. En caso contrario se utilizarán fotografías neutras. El pie de foto deberá incluir la autoría de la misma.
- Los textos en lenguaje natural deberán ser únicos en estilo y redacción a los generados para usuarios distintos a la CRTVE al ser la herramienta de uso propio y personalizada a las necesidades y objetivos de la CRTVE.
- La noticia generada deberá aparecer publicada inmediatamente después de estar disponibles los datos electorales. Podría plantearse la publicación de distintas noticias a lo largo de la jornada electoral en momentos como publicación de datos sobre participación, y durante el proceso de recuento, validación y cifras definitivas de los votos.
- El tiempo desde que estén publicados los datos hasta que las noticias sean visibles en la web será inferior a 5 minutos.
- Para cada noticia se generará además una versión más corta que servirá para la generación del audio que acompaña la noticia. La estructura gramatical y el lenguaje empleado deberá ser el adecuado para que el contenido sea escuchado y no leído. La generación de este audio no será responsabilidad del adjudicatario de este expediente, si lo será el envío de los ficheros de texto hasta el proveedor responsable del audio.
- 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 del mismo. 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 proporcionará un HTML final, personalizado de acuerdo con las especificaciones y modelo definidos en su momento por la CRTVE. El HTML generado deberá publicarse en el/los dominios que la CRTVE asigne al efecto.
- En la página de inicio de la web se implementará una ventana que permita decidir al lector si quiere la información en español o en catalán. En este sentido será necesario disponer del nombre de los pueblos en ambos idiomas.
- Se realizarán todos los desarrollos necesarios para que la web sea accesible para personas con visibilidad reducida en lo referente a back-end, etiquetas de lectura, encabezados y navegación estructurados, elementos front para elección de población, etc.
- Adicionalmente, 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, que serán definidas por RTVE.
- Con el fin de enriquecer las noticias y a petición de la CRTVE, el servicio deberá incluir en el HTML información visible (imágenes, video, audio, textos) proveniente de fuentes externas al servicio o

proporcionadas por la CRTVE, con el fin de enriquecer la noticia.

- 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.
- 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 en buscadores (SEO en inglés) y que, en todo caso, lo mejore al ser un contenido más original. Además, se seguirán todas las pautas dadas por RTVE en cuanto a arquitectura de la información, redireccionamiento de contenido, etc.

Para la **generación de textos de ayuda a periodistas:**

- Se suministrarán como mínimo tres noticias por cada una de las poblaciones seleccionadas (inicialmente, para los municipios de más de 10.000 habitantes).
- Proporcionará por cada una de las noticias un texto en lenguaje natural en español y en catalán.
- La forma en la que se conformará el texto resultante y los elementos que deberá incluir como fotografías o gráficos será definida y validada por la CRTVE durante la fase 1.
- Los textos en lenguaje natural deberán ser únicos en estilo y redacción a los generados para usuarios distintos a la CRTVE al ser la herramienta de uso propio y personalizada a las necesidades y objetivos de la CRTVE.
- Los textos generados deberán aparecer publicados inmediatamente después de estar disponibles los datos electorales. Podría plantearse la publicación de distintas noticias a lo largo de la jornada electoral en momentos como publicación de datos sobre participación, y durante el proceso de recuento, validación y cifras definitivas de los votos.
- El tiempo desde que estén publicados los datos hasta que las noticias sean visibles en la web será inferior a 5 minutos.
- 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 del mismo. 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 proporcionará un HTML final, personalizado de acuerdo con las especificaciones y modelo definidos en su momento por la CRTVE. El HTML generado deberá publicarse en el/los dominios que la CRTVE asigne al efecto.
- En la página de inicio de la web se implementará una ventana que permita decidir al lector si quiere la información en español o en catalán.
- Se realizarán todos los desarrollos necesarios para que la web sea accesible para personas con visibilidad reducida en lo referente a back-end, etiquetas de lectura, encabezados y navegación estructurados, elementos front, para elección de población, etc.
- El adjudicatario facilitará un dominio Web de acceso restringido por usuario y contraseña, para la publicación de estos textos, tanto durante las fases de pruebas y ensayos como el día de las elecciones.
- 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.

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 otras funcionalidades 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 noche electoral, ajustando los plazos del expediente si fuera necesario y sin coste adicional para la CRTVE.

Tal y como se indica en el Pliego de Condiciones Generales, 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. Ambas experiencias se demostrarán mediante la presentación de un certificado emitido por el medio internacional con el que se haya colaborado en el primer caso y en el segundo mediante un documento con información acerca del proyecto, y su nivel de participación en el mismo.

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 autonómicas previstas, a día de hoy, para el mes de febrero de 2025. No obstante, el adjudicatario deberá planificar los trabajos de la fase 1 de forma que si se produce un adelanto electoral pueda reajustar las tareas para que las noticias generadas en la jornada electoral no se vean afectadas ni en cantidad, ni en calidad.

Fase 3: Análisis de resultados

- Análisis posterior de los resultados obtenidos en las elecciones autonómicas, 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 cuyo seguimiento se realizará mediante las reuniones con el equipo de RTVE.

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

- El nivel de publicación de noticias válidas tendrá que ser igual o mayor al 85%.
- 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.

Además, a la finalización de cada una de las fases el adjudicatario elaborará un entregable que recoja no solo la información sobre los trabajos realizados que contemplan los informes mensuales, sino también las tecnologías aplicadas, acuerdos alcanzados y cuanta información se genere en el transcurso del proyecto. Este material podría servir de base para generar los textos, académicos o periodísticos, que se crean oportunos para difundir esta experiencia y sus logros.

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 o generales celebradas y el oferente deberá:

- Generar y presentar en una página Web propiedad del oferente, en menos de 30 minutos, una noticia para cada una de las poblaciones de menos de 10.000 habitantes de Cataluña en español.
- 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 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 y con el lenguaje y estructura gramatical adecuados para ser leído, 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.

Si el oferente no presenta los resultados de esta prueba en el plazo indicado o los mismos no cumplen alguno de los aspectos descritos, 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 Estrategia Tecnológica.

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.

5. Formato de la oferta

Es imprescindible que la oferta técnica presentada tenga la estructura que se indica a continuación.

Propuesta técnica:

- Relación de instalaciones, medios técnicos.
- Relación de recursos humanos, aportando documentación que acredite la experiencia solicitada en el Pliego de Condiciones Generales.
- Relación de las tecnologías utilizadas.
- 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.