

Especificaciones técnicas

GENERACIÓN DE TEXTOS Y VOCES EN CATALÁN MEDIANTE IA



1 Introducción

Como parte de las actuaciones de la dirección de Estrategia Tecnológica en CRTVE, desde el año 2021 se han llevado a cabo pruebas de concepto (PoC) relacionadas con información electoral en poblaciones pequeñas, de menos de 1.000 habitantes, que, basándose en tecnologías de Inteligencia Artificial (IA) y a partir de los datos suministrados por las fuentes de información oficiales, los interpretaba y presentaba una noticia creada automáticamente.

Para mejorar el alcance en la difusión y acceso a la información, las PoC incluían en cada noticia la generación de un texto para leer que podía reproducirse con voces sintéticas únicas desarrolladas partiendo de voces humanas correspondientes a profesionales de la comunicación.

Como parte de estas actuaciones, en estos momentos se está llevando a cabo una PoC más, que finalizará próximamente, mediante la que se están generando noticias de carácter meteorológico para 5 poblaciones de menos de 1.000 habitantes de la provincia de Lleida en catalán y español, testeando así el funcionamiento de las nuevas tecnologías aplicadas a otros idiomas cooficiales.

Sin embargo, CRTVE considera muy interesante poder continuar con el servicio de generación de contenidos audiovisuales relacionados con aspectos meteorológicos en las mismas condiciones que el que se tiene actualmente para poder seguir experimentando nuevas posibilidades en las noticias textuales y perfeccionando las voces en catalán

Por lo tanto, el presente expediente propone la contratación de un servicio de generación de noticias de carácter meteorológico para cinco poblaciones de Cataluña de menos de 1000 habitantes incluyendo además de la noticia escrita, los audios correspondientes, y hacerlo por un periodo de seis meses.

2 Objeto del contrato

Mediante este expediente se contratará un servicio que:

- genere noticias sobre información meteorológica de 5 municipios de la provincia de Lérida de menos de 1.000 habitantes mediante IA.
- utilice **voces sintéticas únicas** (masculina y femenina) creadas partiendo de voces humanas, para generar las piezas de audio que acompañen a las noticias anteriores.

3 Lote 1. Generación de textos automáticos

3.1 Requerimiento del servicio

El sistema ofertado estará basado, de forma demostrable, en las siguientes tecnologías: GPT-4o, GPT-4, Llama 3.1 y otras tecnologías de generación de textos, y dispondrá de las siguientes prestaciones:

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

- Se suministrará una noticia por cada una de las poblaciones seleccionadas.

- 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.
- Podrá incluir vínculos y recomendaciones a otros contenidos de la CRTVE si así se solicita.
- La estructura, en estilo e imagen, de las noticias será de características iguales a la que se dispone en la actualidad y deberá ser validada por la CRTVE.
- En cada noticia deberá aparecer un titular principal y, si CRTVE así lo considerara, los titulares secundarios que se estimen oportunos.
- Cada noticia incluirá al menos una fotografía y un gráfico con iconos animados relacionados con fenómenos atmosféricos.
- Los textos en lenguaje natural deberán ser únicos en estilo y redacción, diferentes a los generados para clientes 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 de la AEMET.
- 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 lote, sí lo será el envío de los ficheros de texto hasta el adjudicatario del lote 2 al igual que la recepción de los ficheros de audio con formato mp3 generados por dicho adjudicatario y su incorporación en la noticia.
- 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.
- 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. Será responsabilidad del adjudicatario la coordinación con la empresa responsable del mantenimiento de la web donde se aloje la información generada. Se deberá tener en cuenta que 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 y 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.
- 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.
- 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 la elaboración de otras noticias de ámbito diferente 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 CRTVE en cuanto a arquitectura de la información, redireccionamiento de contenido, etc.

Tal y como se indica en el Pliego de Cláusulas Administrativas Particulares, 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 Prueba para una mejor evaluación

Tal y como se especifica en el PCAP, 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 el nombre de 5 poblaciones de Lérida y el oferente deberá:

- Generar y presentar en una página Web propiedad del oferente, en menos de 5 minutos, una noticia para cada una de las poblaciones en español y en catalán.
- 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.
 - No se considerará válida 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.

El equipo de CRTVE comprobará los resultados de la prueba en la web del oferente. No obstante, 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 Lote 2. Generación de voces sintéticas y audios

4.1 Requerimiento del servicio

Se contratará un servicio para la generación de piezas de audio que complementen las noticias meteorológicas creadas en el lote 1 del presente expediente, mediante voces sintéticas únicas en castellano y catalán (masculina y femenina), desarrolladas a partir de voces humanas.

En el caso de las voces en español el acento será del centro peninsular, y en el caso del catalán, la voz femenina tendrá acento central y la masculina, acento del noroeste.

El adjudicatario de este lote deberá garantizar la conectividad con la empresa adjudicataria del lote 1, asegurando que el sistema ofertado tenga las prestaciones suficientes para que no se colapse el flujo de datos ni se produzcan retrasos en la entrega de la noticia.

Para cada población se creará una versión en castellano y otra en catalán. Además de las voces, los audios incorporarán efectos sonoros relacionados con los fenómenos meteorológicos (truenos, lluvia, viento) así como una sintonía que serán proporcionados por CRTVE.

La empresa encargada de la generación de textos creará uno específico, con las características adecuadas de duración y formato del contenido, para su transcripción a voz, para cada una de las noticias, y en cada uno de los idiomas. Con estos textos, el adjudicatario de este expediente generará los audios correspondientes, así como clips en los que las voces aparezcan mezcladas con las sintonías y efectos sonoros. Estos clips deberán estar disponibles y publicados en la Web en el mismo momento en el que las noticias sean accesibles para el usuario final, en un tiempo no superior a 15 minutos desde la generación de los textos.

A continuación, se describen las características técnicas y funcionales con las que deberá contar el servicio ofertado.

- Mejora de los modelos de lenguaje en catalán, en especial de la fonética, en caso de ser necesarios según la percepción del grupo de trabajo multidisciplinar de la PoC que está compuesto por personal de CRTVE, así como de las Universidades de Castilla la Mancha y de Lleida.
- Personalización del servicio a partir de la creación de herramientas NLP y diccionarios fonéticos en castellano y catalán para adecuar la síntesis al vocabulario de información meteorológica.
- Disponibilidad del servicio en tiempo real y 24/7.
- Los audios generados se incorporarán a la página Web seleccionada por CRTVE, donde se podrán escuchar en una sola voz, con una probabilidad del 50% para voz masculina y 50% para la femenina. Para el caso del castellano, las dos voces tendrán el mismo acento, para el catalán, los acentos serán los descritos anteriormente.
- Las noticias generadas, serán de duración corta (20 segundos aproximadamente), e incluirán sintonías y efectos sonoros proporcionados por CRTVE.
- Las noticias con información meteorológica se publicarán en la página Web, con una periodicidad diaria, para ello, el adjudicatario convertirá a voz los textos recibidos de la empresa que genera los mismos, creará los clips correspondientes a cada municipio, en español y en catalán y enviará a dicha empresa una URL con el audio en mp3 por cada clip generado.
- Se realizará al menos un reentrenamiento del modelo en catalán lo que implicará el corte, limpieza y preparación de nuevos audios.
- Los audios en castellano deberán mantener la pronunciación de los municipios en catalán.
- Las voces generadas para este proyecto serán propiedad de CRTVE y no podrán ser utilizadas para ninguna otra prueba, proyecto o actividad ajena a CRTVE.

A la finalización del servicio, los archivos de audio del TTS de las voces deberán almacenarse hasta el 31 de diciembre del 2025, entregándose los mismos a CRTVE a la finalización del contrato.

4.2 Prueba para una mejor evaluación

Será condición imprescindible para aceptar la oferta que **los oferentes dispongan en el momento de presentar la oferta de lo siguiente:**

- ❖ Al menos dos voces sintéticas únicas, una masculina y otra femenina, que conviertan el texto de noticias de corta duración a voz (Text to Speech – TTS) en español con acento de España del centro peninsular y en catalán con acento central y del noroeste. Los modelos de voz deberán cumplir las siguientes características:

- Se habrán generado partiendo de **voces humanas** clonadas **de periodistas**, sin embargo, no deben identificar a ninguno concreto.
- Clonadas en el tono de voz y en lo referente a la manera de leer los textos, entonación, cadencia, énfasis, manera de leer los números, etc. Se trata de evitar voces fuente robóticas, poco naturales o diseñadas para otros menesteres.
- Incluirán y podrán controlar atributos relacionados con las propiedades características de la voz en términos de prosodia tales como: ritmo, cadencia, pausas de respiración, velocidad, tono y timbre y estar generados con las técnicas más actuales del estado del arte de la síntesis de voz.
- La naturalidad de la voz debe tener un MOS (Mean Opinion Score) superior a un 4.
- ❖ Capacidad de generar con estas voces, las piezas de audio que incorporen las sintonías asociadas y los efectos sonoros en los planos de tiempo correctos.
- ❖ Generar los audios en tiempo real.

CRTVE podrá solicitar a todos los oferentes, si lo considera oportuno, una demostración de que cumplen las condiciones anteriormente mencionadas.

5 Requerimientos comunes a ambos lotes

5.1 Características del Servicio

El servicio comenzará el 1 de enero o a la firma de contrato si fuera posterior y se prolongará por seis meses durante los cuales se generarán diariamente las noticias textuales sobre información meteorológica para las poblaciones de Lérida elegidas.

5.2 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.

5.3 Condiciones de aceptación del servicio

La aceptación del servicio estará relacionada con la buena marcha de los trabajos cuyo seguimiento se realizará mediante las reuniones con el equipo de CRTVE.

5.4 Coordinación con CRTVE

La empresa adjudicataria nombrará un responsable del servicio dentro de su equipo, que será el coordinador. Este coordinador será el interlocutor único con el personal de CRTVE para la coordinación y priorización de tareas. Así mismo, desde CRTVE también habrá otro coordinador que será su único contacto para cualquier duda o problema que pueda surgir. Se establecerán reuniones ordinarias semanales para el seguimiento de las tareas. Será labor de los coordinadores la convocatoria de reuniones extraordinarias si se considerara necesario.

5.5 Formato de la oferta

La oferta técnica presentada tendrá la estructura que se indica a continuación:

- Organización del trabajo y sistema de aseguramiento de la calidad del servicio.
- Relación de recursos humanos, aportando documentación que acredite la experiencia solicitada en el Pliego de Cláusulas Administrativas Particulares.
- 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 Cláusulas Administrativas Particulares.
- Relación de todos los trabajos similares realizados hasta el momento de entregar las ofertas.