



Expediente: KM/2024/033

Asistencia técnica para realizar transcripciones de audio

[Proyecto «ITZULI»]



Pliego de Prescripciones Técnicas

Índice

Antecedentes	3
Objeto	4
Tareas a realizar	5
Entregables.....	7
Calidad, Control y Seguimiento.....	8
Equipo de Trabajo	10
Estructura de las ofertas.....	11
Propiedad Intelectual.....	12
Seguridad y Confidencialidad.....	13
Criterios de valoración	14

Antecedentes

La «Inteligencia Artificial» (IA) [*Artificial Intelligence*], en inglés] no es una tecnología nueva, puesto que se creó en los años 50. Sin embargo, hoy le podemos sacar el máximo rendimiento gracias, sobre todo, a los avances tecnológicos que se han producido en los últimos años.

El crecimiento de la capacidad de computación, la disponibilidad de gran cantidad de datos que existe actualmente sobre cualquier tema y el avance de los algoritmos han convertido a la Inteligencia Artificial en una de las tecnologías estratégicas del Siglo XXI.

Por todo ello, los sistemas basados en inteligencia artificial se están convirtiendo en un elemento clave para muchas organizaciones, incluidas las Administraciones Públicas.

Sin embargo, hay que tener en cuenta dos aspectos importantes a la hora de hablar de esta tecnología:

1. La inteligencia artificial «aprende» de los datos que le damos. Esto significa que hay que depurar los datos para que el sistema nos pueda dar resultados correctos y útiles.
2. Los sistemas basados en IA se entrenan específicamente para realizar una tarea concreta. Esto significa que el mismo sistema que utilizamos para conducir un vehículo no podríamos utilizarlo para detectar fraudes en operaciones bancarias, por ejemplo.

Uno de los últimos proyectos puestos en marcha por el Gobierno Vasco es el **traductor neuronal** [<https://www.euskadi.eus/traductor>], basado en inteligencia artificial. De hecho, «Itzuli» es el primer proyecto práctico para uso público promovido por el Gobierno Vasco.



Es importante destacar que los sistemas de IA no son sistemas aislados, sino que deben basarse en otras tecnologías para desplegar todo su potencial. Por ejemplo, necesitamos Unidades de Procesamiento Gráfico (GPU), técnicas de *BigData*, algoritmos avanzados, etc. para sacar el máximo provecho.

Como ocurre con las personas, para que un sistema de Inteligencia Artificial pueda «aprender» (o, técnicamente, ser capaz de «generalizar conceptos»), debe pasar por varias fases: la primera se denomina «aprendizaje» y, a continuación, pasaríamos a la fase de «entrenamiento», en la que el sistema IA debería poner en práctica la teoría. Para ello es imprescindible disponer de información y datos suficientes.

Por lo tanto, el objeto del presente expediente es la contratación del servicio de asistencia técnica necesaria para crear el material necesario (en cuanto a calidad y cantidad) para el Gobierno Vasco, y que formará parte del Banco de Recursos Lingüísticos del Gobierno Vasco.

Objeto

Este expediente tiene como objeto contratar los servicios de asistencia técnica necesaria para realizar la transcripción de una serie de vídeos/audios. Estos nuevos recursos, que serán incorporados al Banco de Recursos Lingüísticos del Gobierno Vasco, servirán de base para el «entrenamiento» de los sistemas de Inteligencia Artificial desarrollados por el Gobierno Vasco, como puede ser el traductor neuronal «Itzuli».

El presente Pliego de Bases Técnicas (PBT) describe, por tanto, el alcance, los objetivos y/o requisitos a cumplir por la/s empresa/s adjudicataria/s.

Tareas a realizar

A continuación, se definen las tareas y/o funciones que deberá asumir la empresa que resulte adjudicataria del expediente, así como los recursos que serían necesarios, la dedicación estimada y los productos finales a entregar. La empresa que lleve a cabo dichas tareas actuará en todo momento bajo la supervisión directa de la Dirección de Tecnologías de la Información y la Comunicación [DTIC], o de aquellas personas que ésta designe.

Tal y como se ha indicado, los trabajos a realizar consistirán en:

- Realizar **transcripción** del contenido de distintas grabaciones de audio y/o vídeo

Hay que tener en cuenta que:

- ✓ Para entrenar un modelo de Inteligencia Artificial para detectar entidades, hacen falta ejemplos que se asemejen lo más posible en estilo y contenido a los documentos que posteriormente se tiene intención de procesar.
- ✓ Hay que saber qué se quiere detectar (o qué se puede querer en el futuro). Qué inventario de entidades, y qué granularidad. Esto está relacionado con el caso o casos de uso que se prevé atender.
- ✓ Lo que no figure en los datos de entrenamiento, el modelo de Inteligencia Artificial no podrá detectarlo.

Transcripción del contenido de distintas grabaciones de audio y/o vídeo

Las tareas a realizar consistirán en la transcripción de audios de las grabaciones de audios de distintos **programas de televisión y/o radio**. Los audios y/o contenidos se entregarán a la empresa adjudicataria al inicio de la contratación.

El objetivo final es generar un mínimo de **horas** de contenido que se detalla más adelante.

En cuanto a la **metodología** a seguir por parte de la empresa adjudicataria, se deberán tener en cuenta los siguientes criterios:

- Se anotarán datos acústicos en formato audio (PCM WAV) del dominio: «genérico».
- Los contenidos del ámbito «genérico» incluirán principalmente programas de televisión que incorporen riqueza acústica, de la persona que actúe como locutora y fenómenos relacionados con el habla espontánea. Estos programas serán del género «Entrevistas», «Debates» y «Realities» (por ejemplo, programas de EITB y/o Euskadi Irratia/Radio Euskadi como «Eztabaidan», «Baipasa» o similares), pudiendo ser el contenido tanto en castellano como en euskera.
- Las anotaciones de estos audios se realizarán a diferentes niveles, incluyendo la anotación de los siguientes elementos:
 - Transcripción literal del audio

- Puntuación y capitalización
- Turnos e información del locutor/a (identificador común entre diferentes contenidos para el mismo locutor/a)
- Información acústica de ambiente y audiodescripción (música, ruidos, aplausos, etc.)
- Solapamientos de locutores/as
- Entidades nombradas
- Identificación de temática
- Muletillas y fenómenos del habla espontánea (contracciones, segmentos ininteligibles, repeticiones, correcciones, segmentos eliminados, etc.)
- Definición de convenciones lingüísticas homogéneas (tildes en EU, expansiones de números, abreviaciones, etc.)

A modo orientativo, indicar que el desglose de tipos de programas y horas de grabación a procesar será el siguiente:

Programas	Castellano (espontáneo)	Euskara (espontáneo)	Total de horas
Entrevistas + Debates + Realities	67 horas	67 horas	134 horas

Nota: Esta distribución de horas podrá verse modificada en función de las necesidades del servicio y/o la disponibilidad de las grabaciones correspondientes, lo cual se notificará a la empresa adjudicataria tan pronto como se tenga conocimiento de ello; manteniéndose invariable en todo momento el **total** de horas.

Al inicio de la contratación se intentará facilitar a la empresa adjudicataria una transcripción generada automáticamente de los contenidos en **formato XML** (junto con los contenidos [vídeos]). Y como resultante de la contratación la empresa adjudicataria deberá suministrar como entregable la transcripción rectificada que será validada por parte de la DTIC.

Dada la duración que se estima necesaria para llevar a cabo el proyecto, la empresa adjudicataria deberá reportar a la DTIC, al menos **mensualmente**, un resumen de las tareas realizadas durante el periodo anterior, así como de aquellos aspectos que se consideren más significativos o que puedan afectar al buen desarrollo de las tareas planificadas.

Entregables

El trabajo a realizar consistirá en entregar las **transcripciones** de las distintas grabaciones de audio que se faciliten, cuyas características se detallan en el apartado «*Tareas a realizar*».

Asimismo, y en caso de que así se solicite por parte de la DTIC, la empresa adjudicataria deberá también redactar y entregar las **actas** de aquellas reuniones que se indiquen.

En este caso, los documentos que se elaboren deberán ser entregados a la DTIC en soporte ofimático, estando toda esa documentación en formato compatible con las propias herramientas y estándares establecidos en el Gobierno Vasco.

Asimismo, en caso de que así lo establezca la DTIC, todos estos documentos (salvo las ofertas) se deberán ajustar a las señas de identidad o imagen corporativa del Gobierno Vasco.

Calidad, Control y Seguimiento

Las empresas licitadoras deberán indicar en la oferta si utilizan algún método o plan para garantizar la calidad de los trabajos y su descripción, cuánto tiempo llevan utilizándolo y la experiencia en proyectos cuya calidad haya sido controlada con dicho método durante los últimos dos años (número de proyectos e importe total de los mismos).

En todo caso, la Administración se reserva el derecho de rescindir el contrato si se considera que la empresa adjudicataria no ofrece un servicio de la calidad precisa.

La empresa licitadora deberá establecer todos aquellos procedimientos de control (y describirlos en la oferta que presente) que permitan verificar de forma continuada el cumplimiento de los objetivos especificados, tanto en los plazos como en la calidad final.

Con objeto de realizar el seguimiento más adecuado de los trabajos que deben llevarse a cabo (y que se han detallado en los apartados anteriores), la DTIC realizará una serie de **reuniones** que se detallan a continuación:

- **Lanzamiento:**

Se trata de la **primera reunión** que se celebrará al inicio del proyecto. En ella intervendrán tanto las personas interlocutoras que designe la DTIC como las personas asignadas al proyecto por parte de la empresa adjudicataria. Y el objeto de la misma será establecer las principales líneas del proyecto.

- **Seguimiento:**

Mientras esté vigente el proyecto se realizará una reunión de seguimiento una vez al mes (preferiblemente la primera semana del mes).

Durante la misma, la empresa adjudicataria, en primer lugar, hará un resumen de los trabajos realizados desde la última reunión de seguimiento, es decir, durante el periodo correspondiente al mes anterior (si ha lugar). De esta manera, la DTIC revisará los trabajos ya realizados y comprobará si se corresponden con las propuestas y tareas encargadas/planificadas en su momento. En caso de que todo sea conforme a lo planificado, la DTIC dará el visto bueno a los mismos. En caso contrario, las tareas que queden pendientes deberán ser replanificadas para llevarlas a cabo en el plazo más breve posible, ya que el pago de las facturas quedará supeditada al visto bueno de todas las tareas definidas.

En esta reunión la empresa entregará a la DTIC **un resumen** (incluyendo como mínimo, el grado de avance de las tareas a realizar) de los trabajos realizados y de los trabajos pendientes.

En segundo lugar, la empresa propondrá a la DTIC todas aquellas tareas que estime interesantes y que pretenda llevar a cabo durante ese mes, las cuales podrán ser aceptadas o rechazadas por la DTIC.

- **Extraordinaria:**

Reunión que se llevará a cabo a petición de cualquiera de las partes (DTIC o empresa adjudicataria) en cualquier momento durante el proyecto, siendo su objetivo solventar cualquier **incidencia grave** que requiera una respuesta inmediata y que no pueda esperar a la reunión mensual de seguimiento.

- **Cierre:**

Se trata de la **última reunión** del proyecto.

Durante la misma se hará un repaso general de los objetivos planteados al comienzo del proyecto y cuáles han sido los trabajos realizados para alcanzar esos objetivos.

La empresa adjudicataria entregará al comienzo de la misma un informe final con las tareas llevadas a cabo durante el proyecto.

Equipo de Trabajo

El programa de trabajo se circunscribe a la realización de las tareas descritas en el apartado «*Tareas a realizar*» de este Pliego. La empresa adjudicataria coordinará las relaciones, supervisando la calidad de los trabajos. Para ello deberá aportar su conocimiento y experiencia en trabajos similares. La DTIC, por su parte, se encargará de:

- Ejercer la dirección de todos los trabajos. (La empresa que resulte adjudicataria realizará todas las tareas bajo la supervisión de la DTIC).
- Decidir sobre cualquier asunto que tenga influencia sobre el desarrollo del proyecto
- Aceptar (o rechazar en su caso) los productos intermedios y/o finales

El **equipo de trabajo** estará dirigido por una persona que ocupará el puesto de Responsable del Proyecto con dedicación estable, quien coordinará la realización de todas las tareas implicadas en el proyecto y actuará como interlocutora principal con la persona designada en la DTIC.

El equipo de trabajo estará formado por personal con la categoría profesional y nivel de especialización adecuados a las necesidades planteadas, de acuerdo con las actividades que se vayan desarrollando.

Si por cualquier circunstancia fuese necesaria la sustitución de alguno de los miembros del equipo de trabajo realizada inicialmente por la empresa adjudicataria, la empresa quedará obligada a proponer otras personas con categoría y circunstancias personales y profesionales idénticas o similares a las inicialmente propuestas, sin que dicha circunstancia altere, en ningún caso, el precio ofertado ni cualquier otro concepto.

En las ofertas que se presenten, se deberá incluir en el capítulo «*Equipo de trabajo*» [ver apartado «*Estructura de las ofertas*»] la siguiente información:

El **organigrama** del equipo de trabajo (especificando las funciones y responsabilidades de cada una de las personas que lo integran por categoría profesional y una breve referencia a sus *Curricula Vitae*). Detallando para cada participante del equipo de trabajo, como mínimo, la siguiente información:

- Titulación
- Categoría profesional
- Experiencia en proyectos similares en los últimos años



Estructura de las ofertas

Toda la documentación técnica incluida en la oferta presentada seguirá el siguiente esquema:

1. **Introducción.** Breve descripción de la estructura de la documentación presentada.
2. **Equipo de trabajo.** Descripción (organización y funciones) de las tareas del personal que forma parte del equipo de trabajo de la empresa/entidad para la realización de las tareas.
3. **Planificación y Metodología.** Descripción de las fases y acciones a realizar. Se especificarán las fases en las que se divide el proyecto para facilitar su desarrollo. Para cada fase se podrán especificar objetivos, acciones a realizar, herramientas y/o técnicas a utilizar, productos a obtener, etc.
4. **Control y Seguimiento.** Descripción de los procedimientos que faciliten el control y seguimiento de los trabajos intermedios, detallando la relación con la dirección responsable del proyecto en el Gobierno Vasco.

La empresa podrá incluir en los apartados anteriores cualquier otra cuestión que considere relevante.

En cualquier momento, anterior o posterior a la adjudicación y durante la ejecución de los trabajos, la Administración se reserva la facultad de solicitar cualquier otra documentación complementaria que permita comprobar los datos ofrecidos por la empresa licitadora, tanto de la propia empresa como del personal propuesto.

Propiedad Intelectual

La empresa adjudicataria se compromete a aceptar las siguientes cláusulas:

- ✓ Los derechos de propiedad intelectual relacionados con el trabajo realizado pertenecerán a la Administración.

Cualquier producto o subproducto derivado del mismo que cree o haya creado la empresa, no podrá ser utilizado por ésta para otros fines fuera del ámbito que le corresponda sin el permiso expreso por escrito de la Administración (en este caso el único interlocutor válido será la DTIC).

- ✓ La documentación generada durante la ejecución del trabajo es propiedad exclusiva de la Administración, sin que la empresa adjudicataria para llevar a cabo los trabajos correspondientes pueda conservarla, ni obtener copia de la misma o facilitarla a terceros sin la expresa autorización escrita de la Administración (en este caso, el único interlocutor válido será la DTIC), que la daría, en su caso, previa petición formal de la empresa con expresión del fin que pretende. Lo dispuesto en este párrafo se entiende salvo conservación a los efectos de prueba de realización del trabajo contratado. La Administración autoriza expresamente la utilización de los resultados del presente trabajo como referencia y como base de futuros trabajos a realizar.



Seguridad y Confidencialidad

Tanto la empresa adjudicataria como su equipo de trabajo quedarán expresamente obligados a mantener absoluta **confidencialidad** y reserva sobre cualquier dato que pudieran conocer con ocasión del cumplimiento del contrato, especialmente los de carácter personal, que no podrá copiar o utilizar con fin distinto al que figura en este Pliego de Bases Técnicas, ni tampoco ceder a otros ni siquiera a efectos de conservación.

La empresa deberá garantizar la confidencialidad de la información facilitada por la Administración quedando especialmente sujeta, en cuanto a datos de carácter personal se refiere, a la legislación vigente en la materia —es decir, cumplimiento de lo dispuesto en el Reglamento General de Protección de Datos [RGPD] y en la Ley Orgánica 3/2018, de 5 de diciembre, sobre Protección de Datos Personales y Garantía de los Derechos Digitales [LOPDP-GDD]].

Todas las personas trabajadoras de la empresa adjudicataria se comprometen a mantener el más absoluto secreto profesional sobre aquellos datos, informaciones y documentos que pudieran conocer con motivo de la realización del trabajo, cualquiera que sea la forma bajo la que se representen (oral o escrita). Exclusivamente tendrán acceso a dichas informaciones los colaboradores que sean imprescindibles para el desarrollo de las tareas que les hayan sido asignadas dentro del ámbito al que nos estamos refiriendo. Todos ellos serán advertidos sobre el carácter estrictamente confidencial y reservado de la información a la que tendrán acceso.

A estos efectos, deberán establecerse las máximas cautelas en el acceso a los datos. Cualquier infracción en este sentido será calificada como grave y será causa de resolución del contrato, sin perjuicio de las responsabilidades penales, o de otro tipo, en que se puedan incurrir.

La empresa adjudicataria se compromete a no suministrar a terceros ni utilizar para sí ni para otros los datos facilitados por la Administración para fines distintos a los que contempla este trabajo.

Asimismo, la empresa queda obligada a informar a la Administración, de modo inmediato, sobre cualquier sospecha relacionada con fallos o fugas del sistema de seguridad y protección de la información, que pudiera ser detectada durante el desarrollo del trabajo correspondiente.



Criterios de valoración

En referencia a los criterios (y la ponderación) en que se basarán los criterios de valoración de las ofertas presentadas, consultar el siguiente documento:

«PLIEGO DE CLÁUSULAS ADMINISTRATIVAS PARTICULARES»

En Vitoria-Gasteiz, a la fecha de la firma electrónica

PABLO PEREZ LAFUENTE -
22753603W
Digitalki sinatuta / Firmado
digitalmente
2024.05.08 19:45:59 +02'00'

Pablo Pérez Lafuente

Responsable Sistemas Informáticos

Dirección de Tecnologías de la Información y la Comunicación