

## **PLIEGO DE CLÁUSULAS TÉCNICAS PARTICULARES DE LA FUNDACIÓN CENTRO TECNOLÓGICO DE TELECOMUNICACIONES DE GALICIA (GRADIANT) PARA LA CONTRATACIÓN DE SERVICIOS POR PROCEDIMIENTO NEGOCIADO CON PUBLICIDAD**

En el marco de la cuarta convocatoria de la Compra Pública Precomercial lanzada por el Instituto Nacional de Ciberseguridad (INCIBE), Gradiant ha sido adjudicataria del proyecto fAIr: "Fight Fire with "fAIr": Suite de herramientas basadas en IA multimodal para combatir las amenazas derivadas del uso malicioso de la IA.

El objetivo final de este proyecto es el desarrollo de una plataforma de IA multimodal confiable, robusta y centrada en los datos que permita desarrollar modelos específicos para enfrentar los peligros asociados a un uso malintencionado de la IA. El proyecto incluye la creación de varias herramientas o módulos de detección de Deepfakes de audio, video, imagen, texto, así como demostradores y versiones de pruebas, detección de contenidos sintéticos, protección de contenidos y el análisis de trazabilidad de estos contenidos.

El objeto del presente pliego de contratación es la realización de un servicio de anotado de datos con el fin de etiquetar información para su uso en el proyecto fAIr. El uso de estos datos se realizará tanto en procesos de entrenamiento como de prueba de los modelos desarrollados a lo largo del desarrollo del proyecto, y será de carácter multimedia (imágenes, audios y videos).

El objetivo del pliego es la contratación de una empresa externa para colaborar en el anotado de datos, cubriendo la falta de perfiles adecuados por parte de Gradiant para ejecutar esta tarea.

### **Especificaciones técnicas**

En base a las necesidades y retos planteados, el alcance de los servicios consistirá en el anotado de un conjunto de datos con las siguientes características:

#### **Etiquetado de Imágenes**

- Sexo (masculino, femenino)
- Raza (afroamericano, caucásico, asiático, árabe)
- Edad (joven, adulto, anciano)
- Calidad de la manipulación
  - Baja: Los defectos de la manipulación son evidentes
  - Media: Existen defectos en la imagen, pero no son evidentes
  - Alta: No existen defectos visualmente detectables
- Detección de elementos disruptivos en la imagen (gafas, barba, bigote, sombrero, oclusiones capilares)

## **Etiquetado de Videos**

- Sexo (masculino, femenino)
- Raza (afroamericano, caucásico, asiático, árabe)
- Edad (joven, adulto, anciano)
- Calidad del Deepfake:
  - Baja: Los defectos de la manipulación son evidentes
  - Media: Existen defectos en el vídeo, pero no son evidentes
  - Alta: No existen defectos visualmente detectables
- Detección de elementos disruptivos en la imagen (gafas, barba, bigote, sombrero, oclusiones capilares)

## **Etiquetado de Audios**

- Calidad del deepfake:
  - Baja: Voz muy poco natural y artefactos claramente audibles
  - Media: Naturalidad de la voz media y artefactos poco evidentes
  - Alta: Voz natural y sin artefactos
- Ruido (si, no)
- Sexo (masculino, femenino)
- Idioma (Inglés, castellano, alemán, francés, italiano, otro)

## **Especificaciones técnicas de los datos:**

- Las imágenes tendrán los siguientes formatos: JPEG, PNG, TIFF
- Los videos tendrán los siguientes formatos: H.264, H.265, MPEG4 y AVI
- Los audios tendrán los siguientes formatos: MP3 y WAV

Calidad del servicio: Gradiant verificará que las anotaciones de datos realizadas en el marco del presente documento se ajustan a los requisitos de calidad establecidos por Gradiant para poder hacer uso de los mismos. Con el objetivo de garantizar la calidad del servicio de etiquetado se realizará una revisión aleatoria de los datos etiquetados en cada uno de los entregables. Los criterios específicos para realizar esta evaluación se acordarán conjuntamente con el proveedor para garantizar la calidad del servicio