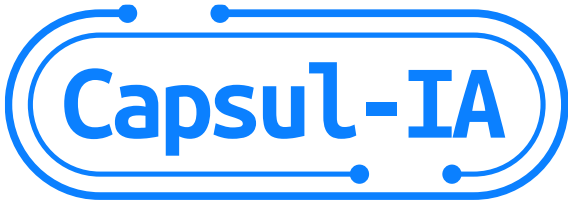


E2.1 REPORT DE TRATAMIENTO DE DATOS

Versión 1.0

Información de la Documentación

Web del Proyecto	https://capsul-ia.es/
Fecha	31/12/2025
Nivel Diseminación	Público
Autor	Joel Frias, Andrés Pacheco
Colaboradores	Jaume Abella, Diego Gil, Alfonso González
Revisor	Lorea Belategi, Alfonso González
Palabras clave	Datos, Privacidad, Ética, IA, Seguridad

Project MIG-20232044, subsidized by the CDTI – Center for Industrial Technological Development, supported by the Ministry of Science and Innovation, and approved in the call for the year 2023 of the procedure for granting aid for “Science and Innovation Missions” within the framework of the TransMisiones 2023 initiative, within the scope of the State Program to Catalyze Innovation and Business Leadership of the State Plan for Scientific and Technical Research and Innovation 2021-2023. Project PLEC2023-010240 funded by MICIU/AEI/10.13039/501100011033”

Registro de Modificaciones

Versión	Descripción del cambio
V1.0	Primera versión del documento

Tabla de Contenidos

1. Introducción

El presente documento constituye el informe de tratamiento de datos del proyecto CAPSUL-IA, una iniciativa de investigación e innovación enmarcada en el programa "Ciencia y Misiones de Innovación" del Centro para el Desarrollo Tecnológico Industrial (CDTI), con el apoyo del Ministerio de Ciencia e Innovación. Este proyecto forma parte de la iniciativa TransMisiones 2023, dentro del Plan Estatal de Investigación Científica y Técnica e Innovación 2021-2023.

El proyecto busca aportar soluciones innovadoras que contribuyan a la transformación digital de sectores clave, potenciando la competitividad y el liderazgo tecnológico español en el campo de la IA.

En este contexto, el tratamiento de datos representa un componente crítico del proyecto, tanto desde la perspectiva técnica como desde el cumplimiento normativo. El presente informe detalla los procedimientos, técnicas y salvaguardas implementados para garantizar que el procesamiento de datos se realiza con el máximo respeto a la privacidad, confidencialidad y seguridad, cumpliendo estrictamente con el marco regulatorio vigente, en particular con el Reglamento General de Protección de Datos (RGPD) y la Ley Orgánica de Protección de Datos Personales y garantía de los derechos digitales (LOPDGDD).

Además de cumplir aspectos claramente regulados, como aquellos relacionados con la privacidad, confidencialidad y seguridad de los datos, debe darse cumplimiento también con criterios éticos para el uso fiable de la IA. Aunque los principios éticos no están regulados mediante leyes que los describen con precisión, existen claras referencias en las que basarse, como son los documentos “AI Act” y “Ethics Guidelines for Trustworthy Artificial Intelligence”, cuyo desarrollo ha sido promovido y apoyado por la Comisión Europea.

2. Base legal

El tratamiento de datos personales y sensibles en el marco del proyecto CAPSUL-IA se lleva a cabo en cumplimiento con la normativa legal vigente:

- Reglamento (UE) 2024/1689 por el que se establecen normas armonizadas sobre inteligencia artificial (Ley de IA).
- Reglamento (UE) 2016/679 del Parlamento Europeo y del Consejo, relativo a la protección de las personas físicas en lo que respecta al tratamiento de datos personales y a la libre circulación de estos datos (RGPD).
- Ley Orgánica 3/2018 de 5 de diciembre, de Protección de Datos Personales y garantía de los derechos digitales (LOPDGDD).
- Proyecto de Ley Orgánica para la protección de las personas menores de edad en los entornos digitales.

La Ley de IA establece un conjunto de normas basadas en el riesgo para los desarrolladores e implementadores de IA en relación con usos específicos. Esta Ley garantiza que los europeos puedan confiar en lo que la IA tiene para ofrecer. Aunque la mayor parte de sistemas de Inteligencia Artificial exhiben un riesgo limitado a cero y pueden ser utilizados para tratar de resolver una amplia variedad de desafíos, otros crean riesgos que se deben abordar para evitar resultados indeseables.

Esta Ley tiene un enfoque basado en el riesgo y define cuatro niveles para los sistemas de Inteligencia Artificial.

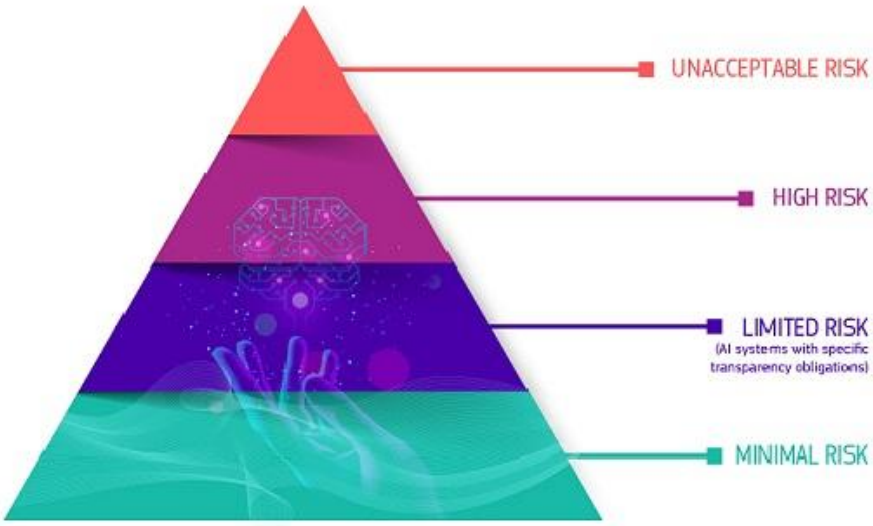


Figura 1: Especificación de los 4 niveles de riesgo contemplados en la Ley de IA

Dentro del último nivel, **riesgo inaceptable**, la Ley prohíbe ocho prácticas:

1. Manipulación y engaño perjudiciales basados en IA.
2. Explotación nociva de vulnerabilidades basada en la IA.
3. Puntuación social.
4. Evaluación o predicción del riesgo de infracción penal individual.
5. Recolección no selectiva de datos de Internet o material de CCTV (Circuito Cerrado de Televisión) para crear o ampliar bases de datos de reconocimiento facial.
6. Reconocimiento de emociones en lugares de trabajo e instituciones educativas.

7. Categorización biométrica para reducir determinadas características protegidas.
8. Identificación biométrica remota en tiempo real con fines policiales de acceso público.

Por otro lado, se considera de **alto riesgo** aquellos casos de uso que puedan plantear riesgos graves para la salud, la seguridad o los derechos fundamentales. Estos sistemas de alto riesgo están sujetos a obligaciones estrictas antes de que puedan comercializarse, como son:

- Tener unos adecuados sistemas de evaluación y mitigación de riesgos.
- Alta calidad de los conjuntos de datos que alimentan el sistema para minimizar los riesgos de resultados discriminatorios.
- Registro de la actividad para garantizar la trazabilidad de los resultados.
- Documentación detallada que proporcione toda la información necesaria sobre el sistema y su finalidad para que las autoridades evalúen su cumplimiento.
- Información clara y adecuada al implementador.
- Medidas adecuadas de supervisión humana.
- Alto nivel de robustez, ciberseguridad y precisión.

Por otro lado, en base al Artículo 5 del RGPD, se deben seguir los siguientes principios del tratamiento de datos personales:

1. Licitud, lealtad y transparencia: los datos deben ser tratados legal, leal y transparentemente para con el interesado.
2. Limitación de la finalidad: los datos se recogerán con fines determinados, explícitos y legítimos y no serán tratados ulteriormente de manera incompatible con dichos fines.
3. Minimización de datos: los datos deben ser adecuados, pertinentes y limitados a lo necesario en relación con los fines para los que se tratan.
4. Exactitud: los datos deben ser exactos y, si fuera necesario, actualizados.
5. Limitación del plazo de conservación: los datos se conservarán durante no más tiempo del necesario para los fines del tratamiento.
6. Integridad y confidencialidad: deben tratarse de forma que se garantice una seguridad adecuada, incluida la protección contra el tratamiento no autorizado o ilícito y contra su pérdida, destrucción o daño accidental.
7. Responsabilidad proactiva: el responsable del tratamiento debe ser capaz de demostrar el cumplimiento de estos principios.

Adicionalmente, se ha considerado como referencia la “Guía para la gestión del riesgo y la evaluación de impacto en tratamiento de datos personales”. Esta guía establece un marco metodológico para la identificación, análisis y mitigación de riesgos asociados al tratamiento de datos personales. En el marco del proyecto CAPSUL-IA, donde se manejan volúmenes significativos de datos, la valuación proactiva del riesgo es necesaria para garantizar un tratamiento seguro, transparente y conforme con el principio de responsabilidad proactiva recogido en el artículo 5.2 del RGPD.

Para la elaboración, en caso de ser necesario, de cláusulas informativas, se ha consultado como referencia la “Guía para el cumplimiento del deber de informar” elaborada por la Agencia Española de Protección de Datos (AEPD). Esta guía ofrece un marco práctico

para asegurar el cumplimiento del deber de información establecido en los artículos 12 a 14 del RGPD, promoviendo la transparencia y la comprensión por parte de los interesados.

En cuanto a los aspectos éticos, las directrices éticas para una IA fiable elaboradas por la Comisión Europea, establecen que la IA debe desarrollarse, desplegarse y usarse respetando los siguientes principios éticos:

- Respeto de la autonomía humana
- Prevención del daño
- Equidad
- Explicabilidad

En el contexto de CAPSUL-IA, la IA se desarrolla, pero no se despliega ni se utiliza en entornos productivos. Por tanto, los aspectos éticos relevantes para el proyecto son aquellos relacionados con la equidad y la explicabilidad. De los cuatro aspectos éticos anteriores, se derivarán una serie de requisitos a cumplir por la IA en cuestión, que son los siguientes:

1. Acción y supervisión humanas
2. Solidez técnica y seguridad
3. Gestión de la privacidad y de los datos
4. Transparencia
5. Diversidad, no discriminación y equidad
6. Bienestar ambiental y social
7. Rendición de cuentas.

Si nos circunscribimos al desarrollo de sistemas basados en IA que se hace en CAPSUL-IA, es decir, excluyendo el despliegue y uso de la IA, los requisitos que deben tenerse en cuenta son los siguientes:

- (2) Solidez técnica. El propio trabajo que se realiza en la elaboración de las cápsulas, con el detalle de los procesos seguidos y los resultados obtenidos satisfacen estos requisitos.
- (3) Gestión de la privacidad y de los datos. Este requisito se considera explícitamente en este entregable para cada una de las cápsulas.
- (4) Transparencia. Mediante la aportación del detalle de cómo han sido diseñadas y entrenadas cada una de las cápsulas, permitimos que cualquier usuario pueda comprender cómo toma las decisiones la cápsula en cuestión.
- (5) Diversidad, no discriminación y equidad. Este requisito se considera explícitamente en este entregable para cada una de las cápsulas.
- (7) Rendición de cuentas. Los aspectos de auditabilidad deben considerarse durante el desarrollo de la IA. El hecho de usar modelos abiertos y detallar los procesos de entrenamiento y verificación de las cápsulas facilita que puedan ser auditadas cuando sean desplegadas.

Otro documento que se ha tenido en cuenta en el procesamiento de los datos ha sido la Observación general núm. 25 (2021) relativa a los derechos de los niños en relación con el entorno digital, teniendo presente la exposición que puedan tener los menores de edad dentro de sistemas de visión por computador. Dentro de este documento, destacan, en el ámbito que nos concierte sus puntos 67, 71 y 88 que dicen lo siguiente:

- Punto 67: [...] Las amenazas a la privacidad de los niños pueden provenir de la reunión y el procesamiento de datos por instituciones públicas, empresas y otras organizaciones, así como de actividades delictivas como el robo de la identidad. Esas amenazas también pueden surgir como resultado de las propias actividades de los niños y de las actividades de los miembros de la familia, sus iguales u otras personas [...].
- Punto 71: [...] cerciorarse de que el niño o, según su edad y el grado de evolución de sus facultades, el padre o el cuidador, den su consentimiento informado, libre y previo al procesamiento de datos [...].
- Punto 88: [...] Los Estados partes deben tener presente que esos riesgos pueden verse propiciados por el diseño y la utilización de tecnologías digitales, por ejemplo, si estas permiten revelar la ubicación de un niño a un posible maltratador [...].

2. Descripción de los datos tratados

En esta sección se introducen los conjuntos de datos utilizados en el desarrollo de las cápsulas. Estos conjuntos de datos son accesibles de forma pública y el objetivo de su introducción es establecer el tipo de datos que representan, el volumen de estos datos, qué riesgos podrían incurrir en cuanto a la filtración de datos personales de los individuos reflejados en los datos, si los hubiera, y qué técnicas se han utilizado para proteger la información y derechos de estos individuos.

Por otro lado, se consideran aspectos relacionados con la inclusión y diversidad a lo largo de todo el tratamiento de datos.

2.1. Datos para cápsula de visión

2.1.1. Datos necesarios

La cápsula de visión debe ser agnóstica a los datos, esto es, debe ser capaz de procesar cualquier tipo de datos y realizar adecuadamente la tarea escogida. No obstante, para la realización de las pruebas, se han seleccionado conjuntos de datos alineados con la resolución del demostrador que integra la cápsula de visión, es decir, el demostrador de conteo de personas.

Uno de los conjuntos de datos que se han recogido para la cápsula de visión se trata de un conjunto de imágenes etiquetadas con múltiples personas. De esta forma, el modelo será capaz de reconocer a las personas. Algunos de los conjuntos de datos estudiados que cumplen este objetivo serían:

- **PersonPath22:** Este conjunto de datos está pensado para aplicaciones de rastreo de múltiples personas. Está formado por videos de una amplia diversidad en lo referente a ángulos de cámara, escenografía, condiciones ambientales y diferentes condiciones de iluminación.

Figura 2: Distribución de los datos según varios atributos

Estos datos cuentan, además, con diferentes etiquetas para representar personas en movimiento, personas quietas, en fondo, sobre vehículo abierto, dentro de vehículo, personas ocluidas, reflejos o masificaciones.

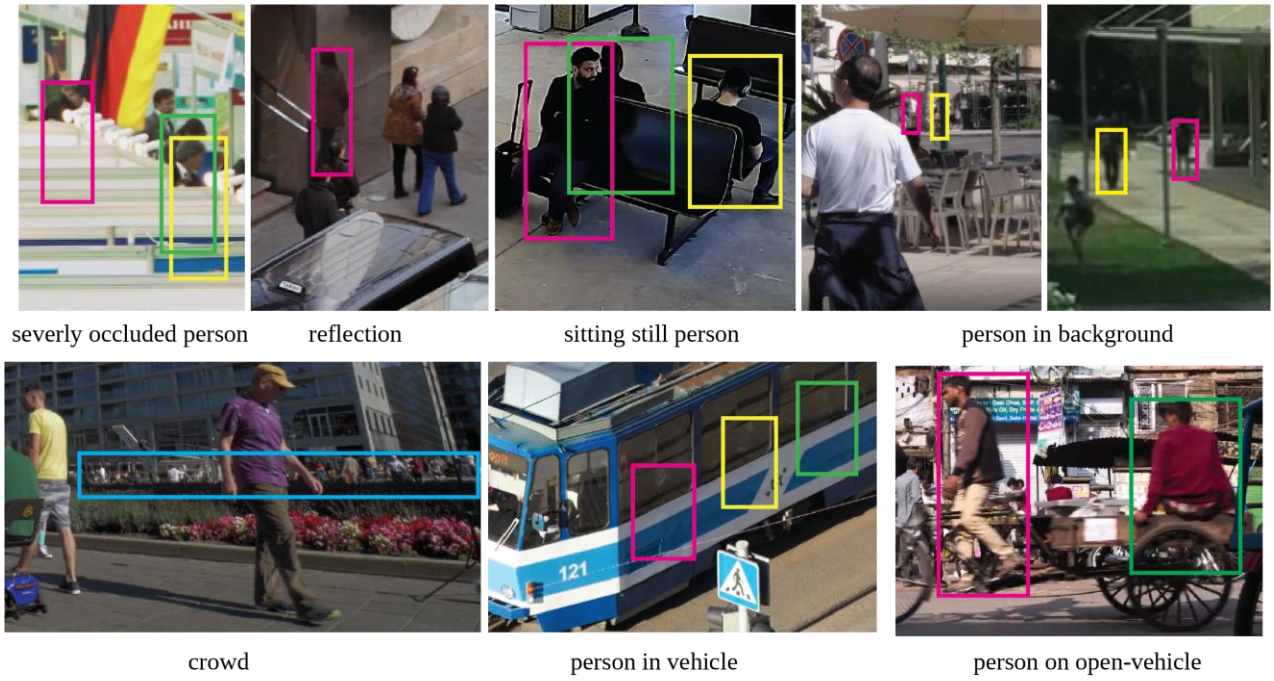


Figura 3: Ejemplos de imágenes etiquetadas y las diferentes anotaciones que tiene el conjunto de datos

- **CrowdHuman:** este conjunto de imágenes está pensado para lograr mejorar los resultados obtenidos en aplicaciones de detección de múltiples personas. En este caso, se contienen una media de 23 personas por imagen, en una gran variedad de escenarios.

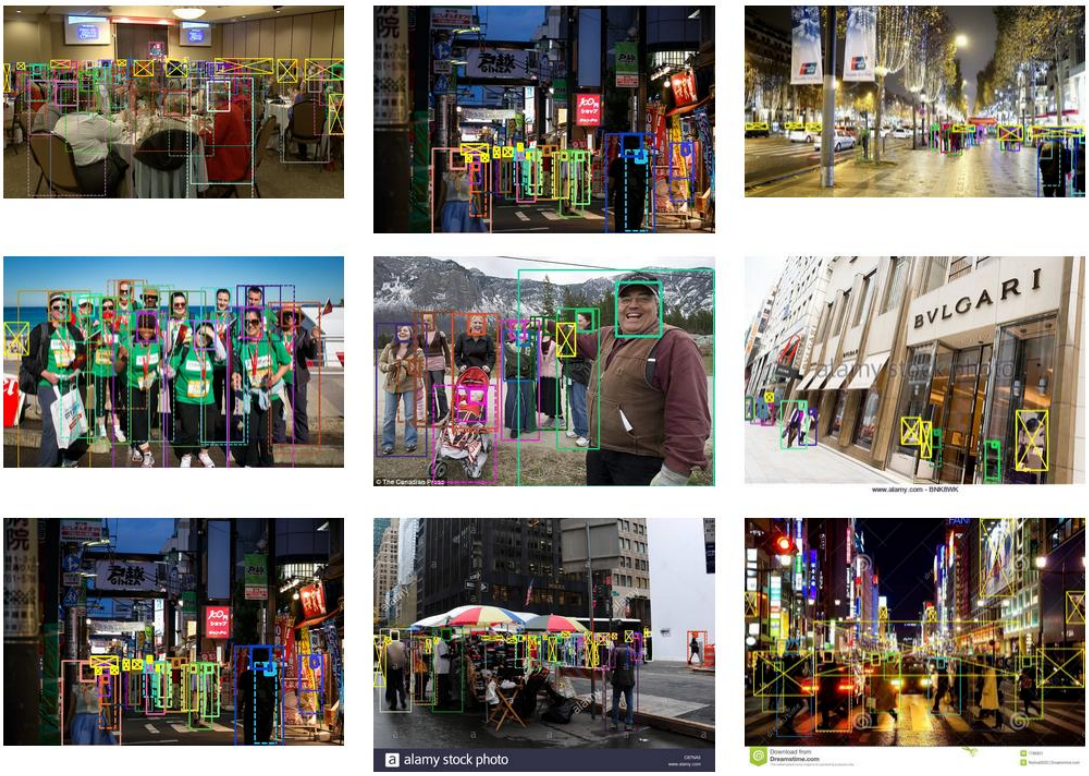


Figura 4: Ejemplos de imágenes etiquetadas y las diferentes anotaciones que tiene el conjunto de datos

Cada una de las personas cuenta con 3 etiquetas: el bounding box de su cabeza, el bounding box visible y el completo. Estos dos últimos se diferencian en que el visible únicamente muestra la zona visible del objeto, mientras que el completo trata de etiquetar el objeto completo.

Figura 5: Ejemplo de etiquetado para múltiples personas solapadas

Para la tarea de clasificación de objetos, se ha encontrado un conjunto de datos capaz de cumplir con el objetivo:

- **EuroSAT:** Este conjunto de datos está formado por un total de 27.000 imágenes y 10 clases distintas, y se trata de un problema de clasificación de terreno a partir de imágenes satelitales. Cada imagen tiene un tamaño de 64*64 píxeles y puede pertenecer a una de las siguientes clases:
  - Cultivo anual

- Bosque
- Vegetación
- Carretera
- Industrial
- Pradera
- Cultivo permanente
- Residencial
- Río
- Mar/Lago

Figura 6: Ejemplos del conjunto de datos utilizado para la clasificación.

2.1.2. Tipos de etiquetado

La tarea de clasificación de objetos necesita de un archivo *.csv*, de forma que indique a qué clase pertenece una imagen concreta. En casos como la clasificación multi-etiqueta, cada imagen puede tener una o más clases vinculadas.

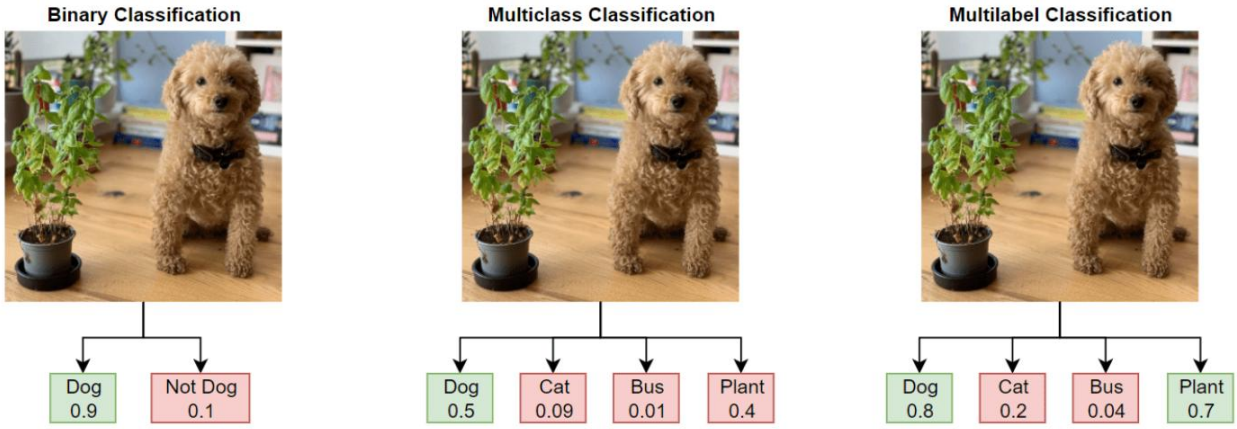


Figura 7: Una imagen se puede etiquetar para diferentes tipos de clasificación.

En la tarea de detección de objetos en el campo de la visión por computador, los datos a utilizar no solo deben reflejar en sus etiquetas el objeto presente, sino también su localización dentro del espacio de la imagen. En la evolución del campo han surgido diversos tipos de etiquetado y formatos.

Algunos de los formatos más comunes en los que se pueden encontrar las imágenes serían:

- Para lograr la tarea de detectar personas, se utiliza un modelo YOLOv8. Estos modelos siguen un etiquetado, conocido como etiquetado **YOLO**. Los bounding boxes en este formato contienen la clase, la X central del objeto normalizada, la Y central del objeto normalizada, el ancho del objeto normalizado y el alto del objeto normalizado.

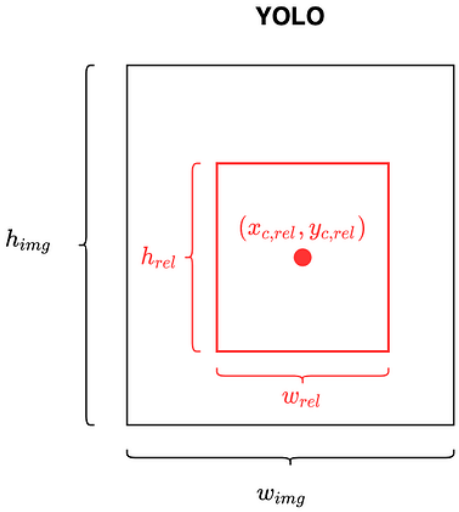


Figura 8: Formato de etiquetado YOLO

$$(x_c,\ rel,\ y_c,\ rel,\ w_{rel},\ h_{rel}) = \left(\frac{x_c}{w_{img}}, \frac{y_c}{h_{img}}, \frac{w}{w_{img}}, \frac{h}{h_{img}}\right)$$

- **COCO:** ofrece como bounding box para cada objeto el valor mínimo del eje X, el valor mínimo del eje Y, el ancho del objeto y el alto del objeto. De esta forma, la transformación a formato YOLO será:

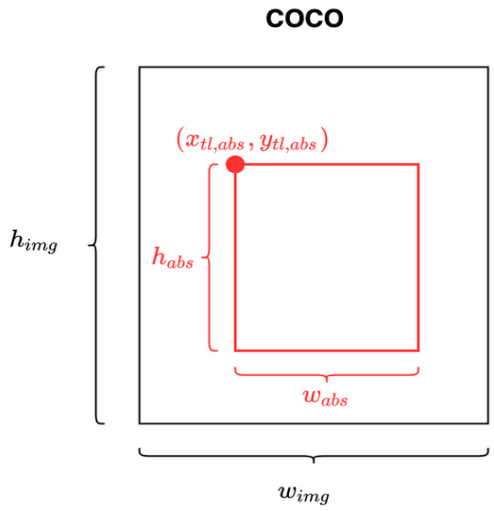


Figura 9: Formato de etiquetado COCO

$$(x_{c,rel}, y_{c,rel}, w_{rel}, h_{rel}) = \left(\frac{x_{min} + \frac{w}{2}}{w_{img}}, \frac{y_{min} + \frac{h}{2}}{h_{img}}, \frac{w}{w_{img}}, \frac{h}{h_{img}}\right)$$

- **Pascal VOC:** ofrece como bounding box para cada objeto el valor mínimo del eje X, el valor mínimo del eje Y, el valor máximo del eje X y el valor máximo del eje Y. De esta forma, la transformación a formato YOLO será:

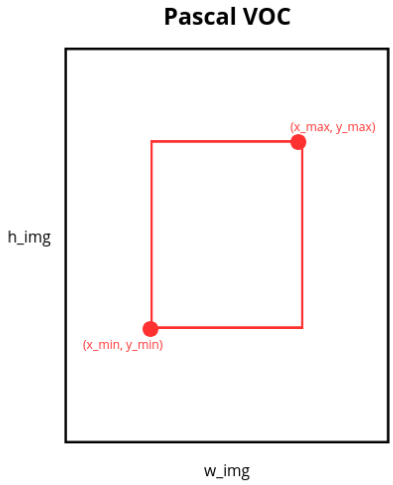


Figura 10: Formato de etiquetado Pascal VOC

$$(x_{c,rel}, y_{c,rel}, w_{rel}, h_{rel}) = \left(\frac{x_{min} + \frac{x_{max} - x_{min}}{2}}{w_{img}}, \frac{y_{min} + \frac{y_{max} - y_{min}}{2}}{h_{img}}, \frac{x_{max} - x_{min}}{w_{img}}, \frac{y_{max} - y_{min}}{h_{img}}\right)$$

- **CXCYWH:** ofrece como bounding box para cada objeto el valor central del eje X, el valor central del eje Y, el ancho del objeto y el alto del objeto. De esta forma, la transformación a formato YOLO será:

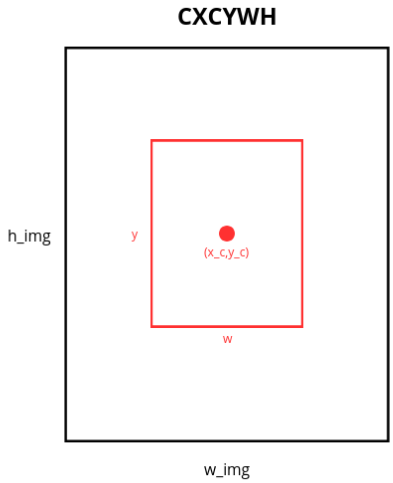


Figura 11: Formato de etiquetado CXCYWH

$$(x_{c,rel}, y_{c,rel}, w_{rel}, h_{rel}) = \left(\frac{x_c}{w_{img}}, \frac{y_c}{h_{img}}, \frac{w}{w_{img}}, \frac{h}{h_{img}}\right)$$

- **Albumentations:** ofrece como bounding box para cada objeto los valores normalizados del valor mínimo del eje X, el valor mínimo del eje Y, el valor máximo del eje X y el valor máximo del eje Y. De esta forma, la transformación a formato YOLO será:

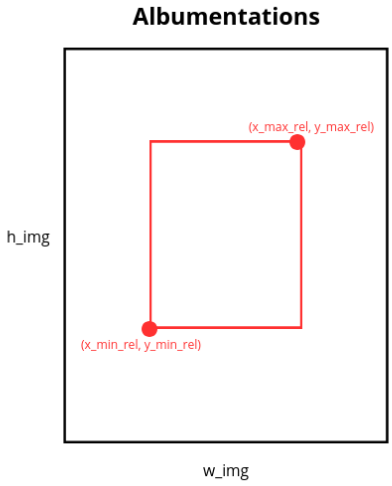


Figura 12: Formato de etiquetado Albumentations

$$(x_{c,rel}, y_{c,rel}, w_{rel}, h_{rel}) = \left(x_{min,rel} + \frac{w_{rel}}{2}, y_{min,rel} + \frac{h_{rel}}{2}, x_{max,rel} - x_{min,rel}, y_{max,rel} - y_{min,rel}\right)$$

Para el desarrollo actual se tendrán en cuenta los diferentes formatos que se han mostrado. En un futuro si fuera necesario se podrán incluir nuevos formatos.

2.1.3. Flujo de datos

Para garantizar que el usuario logre una buena solución tras el entrenamiento, las imágenes cargadas por el usuario deberán seguir una serie de preprocesados.

En primer lugar, al momento de subir las imágenes, el usuario deberá indicar qué tipo de etiquetado poseen las imágenes. Esta información es necesaria, ya que, si el formato de las etiquetas no corresponde con el utilizado por el modelo (YOLO), se aplicarán automáticamente las transformaciones necesarias para convertir las anotaciones al formato compatible.

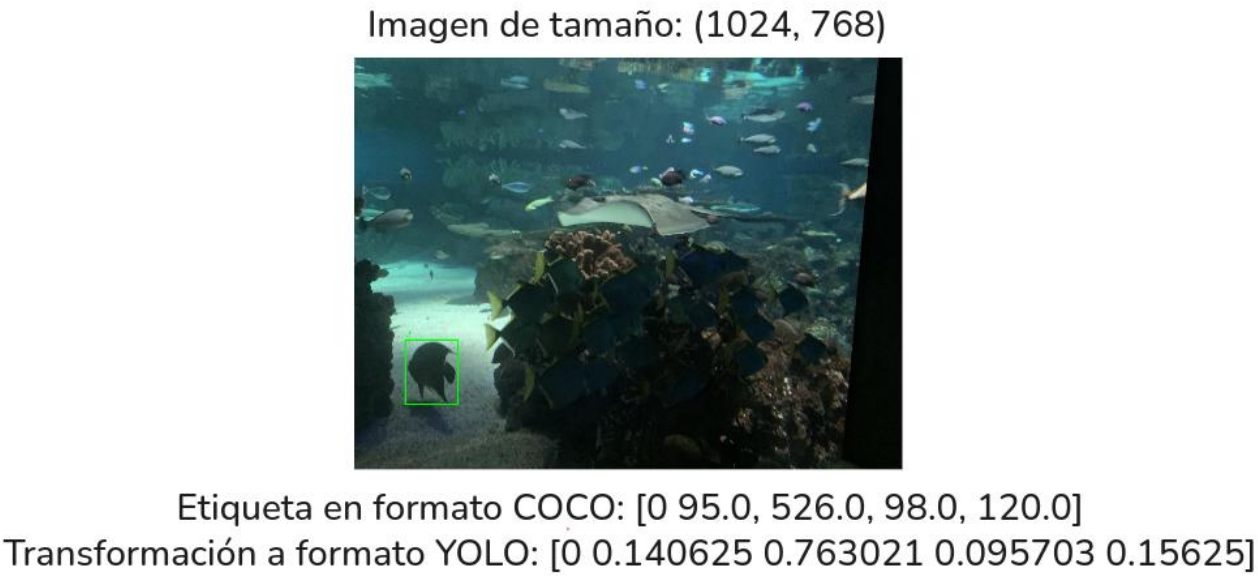


Figura 13: Ejemplo de trasformación de coordenados de COCO a YOLO

Al terminar con el procesado de las etiquetas, se comenzará con la evaluación de las imágenes. En primer lugar, se llevará a cabo una verificación para detectar imágenes duplicadas, etapa que evita la introducción de ruido en los modelos.

Seguidamente, se valorarán las métricas de entropía y borrosidad de las imágenes. Aquellas imágenes cuya entropía sea demasiado baja (cercana a 0) significará que contiene poca información y, por tanto, el modelo podría no ser capaz de detectar patrones en la imagen. Una menor entropía implica una menor variedad de colores en la imagen (menos información contiene la imagen). Por otro lado, la métrica de borrosidad ayudará a detectar aquellas imágenes con una entropía elevada pero que sean borrosas, por lo que dificultaría el aprendizaje de los modelos.

Una vez detectadas todas las imágenes que cumplan estas condiciones, se advertirá al usuario que existen imágenes problemáticas, indicando las correspondientes referencias.


Figura 14: Ejemplos de entropía y borrosidad

	Entropía	Borrosa(Sí/No)
Imagen 1	7.33	Sí
Imagen 2	7.76	No

Finalmente, para proteger la privacidad de las personas que puedan aparecer en las imágenes, se aplicará un modelo de detección facial. Gracias a este mecanismo, los rostros identificados serán ofuscados automáticamente antes de ser utilizados en el entrenamiento, protegiendo su identidad.

Figura 15: Ejemplo emborronado de caras para mantener la privacidad de las personas.

2.1.4. Ética

La cápsula de visión deberá tener en cuenta dos casos para asegurar los principios de inclusión y diversidad.

El primer caso se trata del uso de un modelo pre-entrenado. En este primer caso será responsabilidad del desarrollador asegurar que se cumple con los requisitos de equidad. Para ello, las imágenes con las que se trabaja provienen de distintas zonas geográficas a lo largo y ancho del globo, asegurando que distintas etnias, culturas y contextos queden

debidamente reflejados, contribuyendo a la minimización de sesgos y asegurando un enfoque más inclusivo en los resultados del modelo.

El segundo caso corresponde a los entrenamientos personalizados realizados por los propios usuarios. Aquí, la responsabilidad de mantener los principios de inclusión y diversidad recae en el usuario, ya que el modelo aprende con sus propios datos. Es decir, el modelo final puede volverse independiente del conjunto de datos original y, por tanto, reflejar únicamente las características de los datos proporcionados por el usuario.

Por esta razón, los usuarios deben asumir el compromiso de alimentar el sistema con información diversa, inclusiva y representativa. Es por ello por lo que en el anexo Manual de ética de datos se han descrito las pautas para que el usuario pueda considerarlas al entrenar sus propios modelos.

2.1.5. Datos robótica

Los datos de robótica vendrán en formato ROS. ROS es un conjunto de librerías y herramientas para la creación de aplicaciones de robótica. Es posible descomprimir un conjunto de imágenes en este formato y transformarlas en un formato manejable para los modelos como por ejemplo png, jpg, etc.

2.1.6. Aumento de Datos

En los datos tomados en exterior es frecuente tener un sesgo hacia condiciones meteorológicas particulares. Por ejemplo, es frecuente que las imágenes de día soleado estén sobrerrepresentadas respecto a las imágenes nocturnas, con lluvia, con nieve o con niebla. Esto tiene un impacto en el modelo de IA puesto que se crea un sesgo que, generalmente, llevan a que el modelo sea especialmente preciso para las condiciones sobrerrepresentadas en el entrenamiento, y su rendimiento sea pobre para las demás condiciones.

Para mitigar esta situación, se ha desarrollado una nueva solución de aumento de datos que, a partir de un conjunto de imágenes, cada una para unas condiciones meteorológicas concretas, genera imágenes equivalentes para las demás condiciones meteorológicas. En

particular, las condiciones consideradas son las siguientes: día soleado, noche, nieve, lluvia y niebla.

Los modelos de difusión son una potente familia de modelos generativos capaces de crear imágenes, audio y otras modalidades de datos con alta calidad. El objetivo de estos modelos es añadir ruido gradualmente a los datos durante un proceso de avance, y aprender a deshacer este proceso paso a paso en el proceso de retroceso para generar nuevos datos. La combinación de estos procesos permite al modelo deshacer el ruido generado y producir datos realistas.

Este proceso de eliminación de ruido se puede condicionar para guiar la generación hacia atributos o señales específicos. Por ejemplo, un enfoque común es el guiado sin clasificador, donde el modelo se entrena con y sin información de condicionamiento. Durante la inferencia, se puede aplicar una escala de guiado para ajustar la influencia del condicionamiento, lo que permite tareas como la generación de imágenes guiada por texto o la síntesis de datos específicos del dominio.

Construyendo sobre dichos procesos, el modelo Instruct Pix2Pix expande las capacidades de los modelos de difusión introduciendo el condicionamiento de imágenes y texto. Basándonos en la infraestructura Stable Difusion (un modelo de difusión diseñado para generar imágenes de alta resolución a partir de comandos de texto), Instruct Pix2Pix permite editar localizaciones concretas de las imágenes usando instrucciones textuales.

Combinando un espacio latente preentrenado e incorporando una imagen e instrucciones textuales, Instruct Pix2Pix demuestra la flexibilidad de los modelos de difusión.

Algoritmo: Proceso de aumento de datos usando Instruct Pix2Pix

Datos de entrada: Un juego de datos *D* de imágenes sin condiciones meteorológicas adversas.

Datos de salida: Un juego de datos aumentado *D'* que incluye imágenes con distintas modificaciones inducidas por las condiciones meteorológicas.

Paso 1: Aplicación de los comandos

Aplicar el modelo Instruct Pix2Pix a cada imagen en D utilizando comandos de texto predefinidos e hiperparámetros para cada condición meteorológica deseada. En algunas condiciones climáticas es necesario aplicar varios comandos secuencialmente.

Paso 2: Generación de imágenes aumentadas

Generar las imágenes aumentadas D' que contienen las modificaciones para cada condición meteorológica deseada.

Paso 3: Filtrado

Filtrado manual de D' para eliminar:

1) Imágenes aumentadas donde faltan objetos debido a alucinaciones del modelo.
2) Imágenes aumentadas que muestran perturbaciones irreales o extremas.

Se integran las imágenes restantes en el juego de datos original, D' = D ∪ D', donde las imágenes con y sin condiciones meteorológicas adversas están incluidas.

Retornar D'

Se ha aplicado este proceso en un conjunto de 5.000 imágenes generadas con el simulador CARLA, así como con el juego de datos reales BDD100K. Los resultados son prometedores, pues la calidad de las predicciones en condiciones meteorológicas adversas mejora con poco impacto en las predicciones sin dichas condiciones meteorológicas.

Las siguientes tablas muestran la medida de precisión mAP50 para los modelos de detección de objetos FasterRCNN, YOLO-M y YOLO-N usando los datos originales de CARLA, y usando los datos aumentados con nuestra técnica.

Tabla 1: Comparación de resultados con el conjunto original y aumentado. Métrica mAP50

Condición meteorológica	FasterRCNN		YOLO-M		YOLO-N	
	Original	Aumentado	Original	Aumentado	Original	Aumentado
Día soleado	**0.881**	0.838	0.536	**0.550**	**0.422**	0.408
Niebla	0.635	**0.781**	0.395	**0.493**	0.302	**0.391**
Noche	0.556	**0.717**	0.336	**0.457**	0.266	**0.337**
Lluvia	0.680	**0.688**	0.380	**0.435**	0.236	**0.312**

Como se puede observar, la mejora de las predicciones con los datos aumentados es sostenible en todos los casos tanto para niebla, noche y lluvia. En el caso de los datos originales, al dejar de estar sobrerrepresentados, las predicciones empeoran ligeramente en algunos casos, aunque no siempre.

Para completar esta sección, se incluyen ejemplos de imágenes originales y aumentadas en distintas condiciones en casos exitosos y en casos donde el modelo produce alucinaciones, siendo estas últimas filtradas del juego de datos.

Tabla 2: Ejemplos de resultados

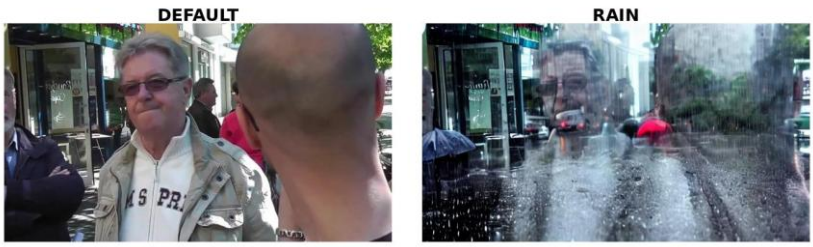
Aumento satisfactorio con niebla		
Aumento satisfactorio nocturno		

Aumento satisfactorio con lluvia

Aumento con alucinaciones con niebla

Aumento con alucinaciones nocturno

Aumento con alucinaciones con lluvia



Como tarea pendiente para el resto del proyecto se queda el automatizar el filtrado de las imágenes con alucinaciones.

2.1. Datos para cápsula de texto

Para el desarrollo de la cápsula de texto será necesario contar con un conjunto de datos diverso que permita analizar las diferentes herramientas y funcionalidades con los modelos LLMs. Todas las acciones realizadas para el conjunto de datos mostrado serán igualmente aplicables para cualquier conjunto de documentos que se utilice en la cápsula.

En lo referente al caso de uso de la Aplicación 3, se han utilizado documentos de texto disponibles en los portales web del Poder Judicial Español y en EUR-Lex, así como licitaciones de contratación pública.

Tabla 3: Distribución de los documentos recogidos

Fuente	Idiomas	Docs/Idioma (media)	Total documentos
Contratación pública	1	2019	2019
Poder Judicial	1	1149	1149
EUR-Lex	24	1219	29266

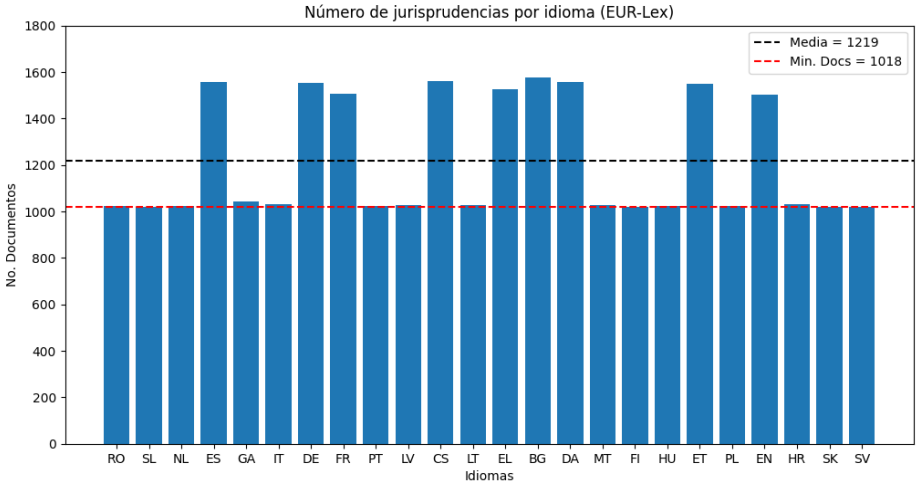


Figura 16: En el caso de documentos multilingües, se tiene al menos 1000 documentos de 24 idiomas

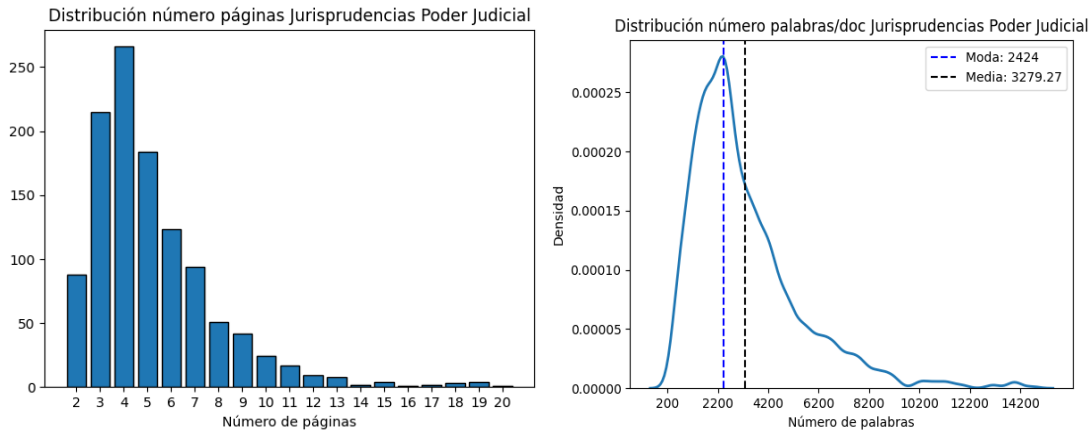


Figura 17: Estos documentos suelen tener, mayoritariamente, 4 páginas de extensión y una media de 2400 palabras por documento.

A continuación, se muestra como es el formato que sigue cada una de las fuentes que conforman el conjunto de datos completo:

- **Licitaciones de contratación pública:** cuenta con un total de 2019 documentos en español. El estilo de documentos es el siguiente:

Anuncio de licitación
Número de Expediente **CC170003**
Publicado en la Plataforma de Contratación del Estado el 30-12-2016 a las 09:58 horas.

Plataforma de Contratación del Sector Público

Contratación de servicios de transporte del correo por carretera en la Zona 5ª Alicante (3 lotes)

- Valor estimado del contrato 917.280 EUR.
- Importe 704.704 EUR.
- Importe (sin impuestos) 582.400 EUR.
- Plazo de Ejecución
  - 48 Mes(es)
- Clasificación CPV
  - 60160000 - Transporte de correspondencia por carretera.
- Tipo de Contrato Servicios
- Subtipo Transporte de correo por vía terrestre y por vía aérea

Proceso de Licitación

- Procedimiento Negociado con publicidad
- Tramitación Ordinaria
- Ofertar a A uno o varios lotes

Figura 18: Ejemplo de licitaciones de contratación pública.

- **Documentos del poder judicial español:** se cuenta con un total de 1149 documentos en español. El estilo de documentos es:

CONSEJO GENERAL DEL PODER JUDICIAL

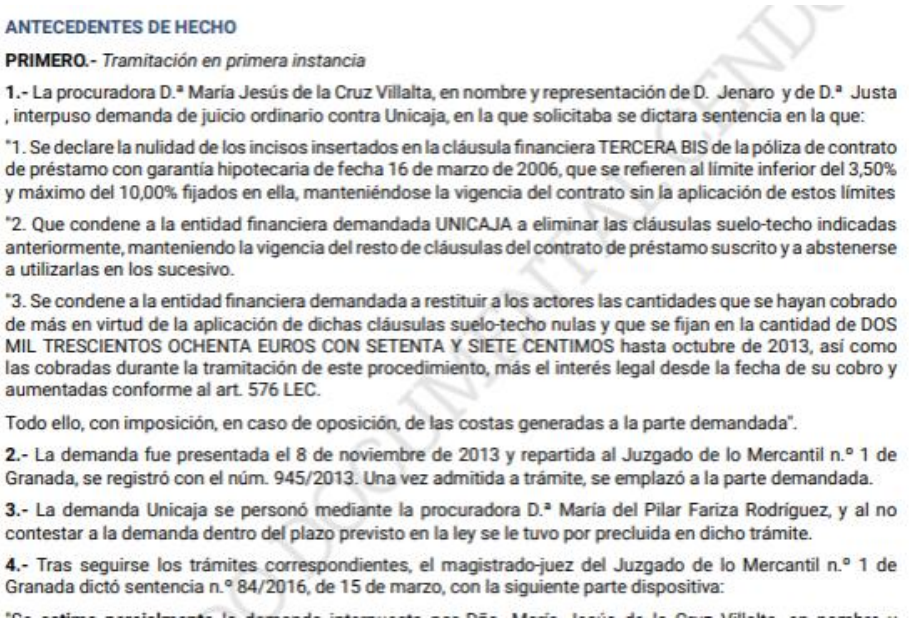
JURISPRUDENCIA

Roj: **STS 7/2020 - ECLI:ES:TS:2020:7**

Id Cendoj: **28079110012020100003**
Órgano: **Tribunal Supremo. Sala de lo Civil**
Sede: **Madrid**
Sección: **1**
Fecha: **08/01/2020**
Nº de Recurso: **1427/2017**
Nº de Resolución: **9/2020**
Procedimiento: **Recurso extraordinario infracción procesal**
Ponente: **PEDRO JOSE VELA TORRES**
Tipo de Resolución: **Sentencia**

Resoluciones del caso: **SAP GR 178/2017, ATS 5844/2019, STS 7/2020**

TRIBUNAL SUPREMO
Sala de lo Civil
Sentencia núm. 9/2020
Fecha de sentencia: 08/01/2020



ANTECEDENTES DE HECHO

PRIMERO.- *Tramitación en primera instancia*

1.- La procuradora D.ª María Jesús de la Cruz Villalta, en nombre y representación de D. Jenaro y de D.ª Justa , interpuso demanda de juicio ordinario contra Unicaja, en la que solicitaba se dictara sentencia en la que:

"1. Se declare la nulidad de los incisos insertados en la cláusula financiera TERCERA BIS de la póliza de contrato de préstamo con garantía hipotecaria de fecha 16 de marzo de 2006, que se refieren al límite inferior del 3,50% y máximo del 10,00% fijados en ella, manteniéndose la vigencia del contrato sin la aplicación de estos límites

"2. Que condene a la entidad financiera demandada UNICAJA a eliminar las cláusulas suelo-techo indicadas anteriormente, manteniendo la vigencia del resto de cláusulas del contrato de préstamo suscrito y a abstenerse a utilizarlas en los sucesivo.

"3. Se condene a la entidad financiera demandada a restituir a los actores las cantidades que se hayan cobrado de más en virtud de la aplicación de dichas cláusulas suelo-techo nulas y que se fijan en la cantidad de DOS MIL TRESCIENTOS OCHENTA EUROS CON SETENTA Y SIETE CENTIMOS hasta octubre de 2013, así como las cobradas durante la tramitación de este procedimiento, más el interés legal desde la fecha de su cobro y aumentadas conforme al art. 576 LEC.

Todo ello, con imposición, en caso de oposición, de las costas generadas a la parte demandada".

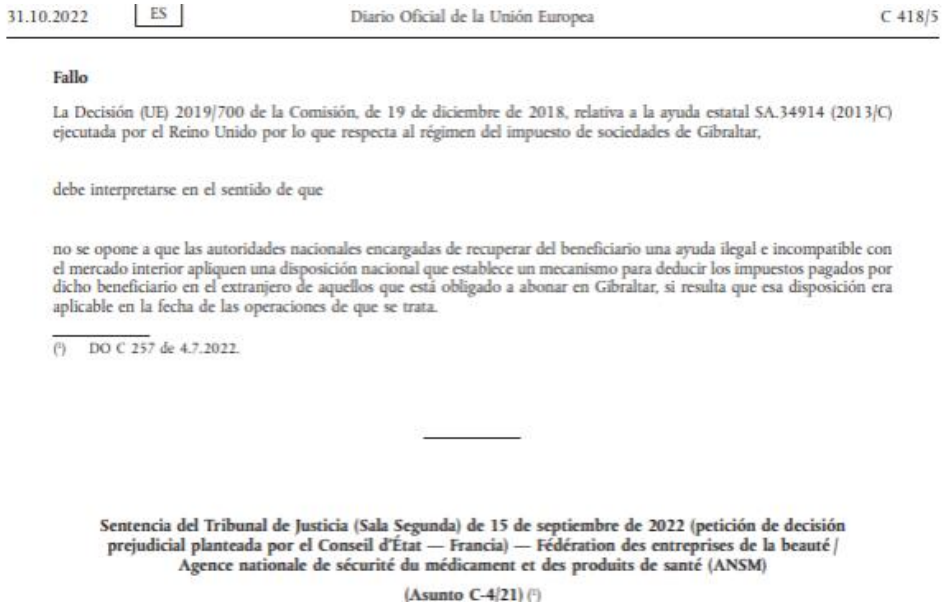
2.- La demanda fue presentada el 8 de noviembre de 2013 y repartida al Juzgado de lo Mercantil n.º 1 de Granada, se registró con el núm. 945/2013. Una vez admitida a trámite, se emplazó a la parte demandada.

3.- La demanda Unicaja se personó mediante la procuradora D.ª María del Pilar Fariza Rodríguez, y al no contestar a la demanda dentro del plazo previsto en la ley se le tuvo por precluida en dicho trámite.

4.- Tras seguirse los trámites correspondientes, el magistrado-juez del Juzgado de lo Mercantil n.º 1 de Granada dictó sentencia n.º 84/2016, de 15 de marzo, con la siguiente parte dispositiva:

Figura 19: Ejemplos de documentos del poder judicial español

- **EUR-Lex:** este conjunto de datos tiene 29266 documentos en 24 idiomas distintos, siendo que cada uno de estos idiomas tiene mínimo 1000 documentos. El formato de documentos sería:



31.10.2022 ES Diario Oficial de la Unión Europea C 418/5

Fallo

La Decisión (UE) 2019/700 de la Comisión, de 19 de diciembre de 2018, relativa a la ayuda estatal SA.34914 (2013/C) ejecutada por el Reino Unido por lo que respecta al régimen del impuesto de sociedades de Gibraltar,

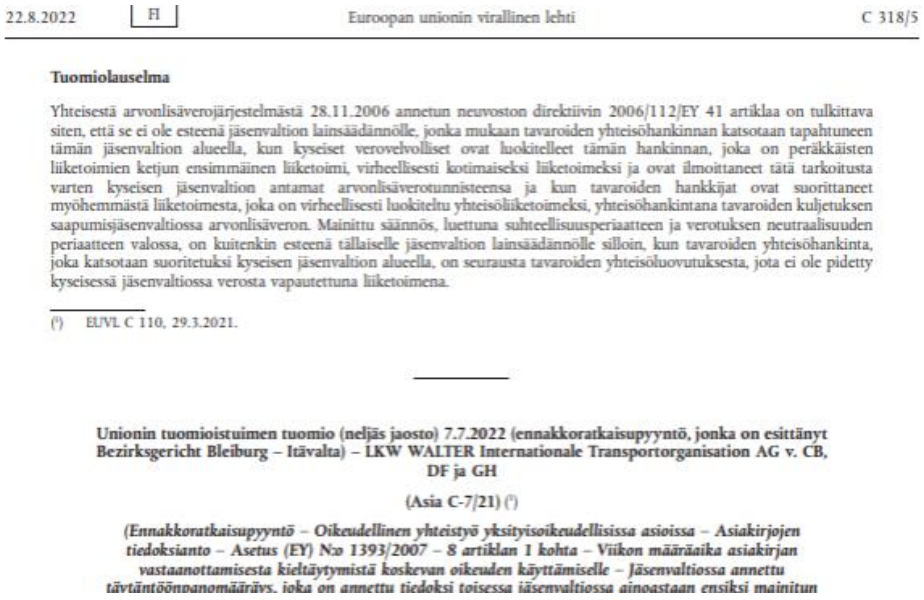
debe interpretarse en el sentido de que

no se opone a que las autoridades nacionales encargadas de recuperar del beneficiario una ayuda ilegal e incompatible con el mercado interior apliquen una disposición nacional que establece un mecanismo para deducir los impuestos pagados por dicho beneficiario en el extranjero de aquellos que está obligado a abonar en Gibraltar, si resulta que esa disposición era aplicable en la fecha de las operaciones de que se trata.

(*) DO C 257 de 4.7.2022.

Sentencia del Tribunal de Justicia (Sala Segunda) de 15 de septiembre de 2022 (petición de decisión prejudicial planteada por el Conseil d'État — Francia) — Fédération des entreprises de la beauté / Agence nationale de sécurité du médicament et des produits de santé (ANSM)

(Asunto C-4/21) (*)



22.8.2022 FI Euroopan unionin virallinen lehti C 318/5

Tuomiolauselma

Yhteisestä arvonlisäverojärjestelmästä 28.11.2006 annetun neuvoston direktiivin 2006/112/EY 41 artiklaa on tulkittava siten, että se ei ole esteenä jäsenvaltion lainsäädännölle, jonka mukaan tavaroiden yhteisöhankinnan katsotaan tapahtuneen tämän jäsenvaltion alueella, kun kyseiset verovelvolliset ovat luokitelleet tämän hankinnan, joka on peräkkäisten liiketoimien ketjun ensimmäinen liiketoimi, virheellisesti kotimaiseksi liiketoimeksi ja ovat ilmoittaneet tätä tarkoitusta varten kyseisen jäsenvaltion antamat arvonlisäverotunnisteensa ja kun tavaroiden hankkijat ovat suorittaneet myöhemmästä liiketoimesta, joka on virheellisesti luokiteltu yhteisöliiketoimeksi, yhteisöhankintana tavaroiden kuljetuksen saapumisjäsenvaltiossa arvonlisäveron. Mainittu säännös, luettuna suhteellisuusperiaatteen ja verotuksen neutraalisuuden periaatteen valossa, on kuitenkin esteenä tällaiselle jäsenvaltion lainsäädännölle silloin, kun tavaroiden yhteisöhankinta, joka katsotaan suoritetuksi kyseisen jäsenvaltion alueella, on seurausta tavaroiden yhteisöluovutuksesta, jota ei ole pidetty kyseisessä jäsenvaltiossa verosta vapautettuna liiketoimena.

(1) EUVL C 110, 29.3.2021.

Unionin tuomioistuimen tuomio (neljäs jaosto) 7.7.2022 (ennakkoratkaisupyyntö, jonka on esittänyt Bezirksgericht Bleiburg – Itävalta) – LKW WALTER Internationale Transportorganisation AG v. CB, DF ja GH

(Asia C-7/21) (1)

(Ennakkoratkaisupyyntö – Oikeudellinen yhteistyö yksityisoikeudellisissa asioissa – Asiakirjojen tiedoksianto – Asetus (EY) N:o 1393/2007 – 8 artiklan 1 kohta – Viikon määräaika asiakirjan vastaanottamisesta kieltäytymistä koskevan oikeuden käyttämiselle – Jäsenvaltiossa annettu tävtäntöönnanomäärävs, joka on annettu tiedoksi toisessa iäsenvaltiossa ainoastaan ensiksi mainitun

Figura 20: Ejemplo de documentos de EUR-Lex

Estos documentos se han recopilado para el diseño y testeo del caso de uso. Una vez finalizado este proceso, dichos documentos no se incluyen en el producto software final. En última instancia, serán los usuarios finales quienes proporcionen al modelo de lenguaje los documentos que deseen almacenar en la base de datos para su posterior consulta.

En otras palabras, el modelo es agnóstico a los datos. La aplicación ha sido diseñada teniendo en cuenta el caso de uso y su problemática —por ejemplo, la necesidad de encontrar documentos relevantes dentro de una base de datos potencialmente extensa— pero estos documentos no afectan al modelo en el sentido de que no se realiza un *fine-tuning* con ellos.

Dicho esto, dada la naturaleza de la aplicación, la responsabilidad final sobre qué datos se suministran a la base de datos recae en el usuario. En contextos como el legal, esto incluye también la responsabilidad de asegurarse de que los documentos indexados no contengan sesgos que puedan favorecer a determinados sectores de la sociedad o discriminar a grupos minoritarios. Si esto ocurriera, las respuestas del modelo también reflejarían dicho sesgo, ya que estarían basadas en información parcial o tendenciosa. Aun así, las respuestas que den los modelos utilizados cuentan con mecanismos de seguridad y que serán reflejados y expuestos en el propio desarrollo de la cápsula.

2.1.1. Evaluación

Para los conjuntos de texto, se han estudiado y aplicado métricas de evaluación, de forma que podamos saber qué tan difícil sea entender un texto, y podamos esperar la capacidad del modelo para entenderlo:

- **Readability:** permite saber qué tan difícil es un texto de entender. Esta métrica tiene su adaptación al inglés y al español:
  - **Flesch–Kincaid readability:** se trata de la dificultad que tiene entender un texto en inglés.

Flesch-Kincaid Score	Reading Level	School Level	Age Range (US)	Example Book
0 - 3	Basic	Kindergarten / Elementary	5 - 8	Hooray for Fish!
3 - 6	Basic	Elementary	8 - 11	The Gruffalo
6 - 9	Average	Middle School	11 - 14	Harry Potter
9 - 12	Average	High School	14 - 17	Jurassic Park
12 - 15	Advanced	College	17 - 20	A Brief History of Time
15 - 18	Advanced	Post-grad	20+	Academic Papers

Figura 21: Tabla de Flesch-Kincaid para determinar la dificultad de lectura

- **Fernández Huerta readability:** se trata de la dificultad que tiene entender un texto en español.

Puntuación	Nivel acamémico	Descripción
90-100	5º de primaria	Lectura muy fácil. Un estudiante medio de 11 años entiende el texto sin esfuerzo.
80-90	6º de primaria	Lectura fácil. Lenguaje conversacional.
70-80	1º de ESO	Lectura relativamente fácil.
60-70	2º-3º de ESO	Lenguaje llano. Un estudiante medio de 13 a 15 años entiende el texto sin esfuerzo.
50-60	4º de ESO - 2º Bach	Lectura relativamente difícil.
30-50	Grado universitario	Lectura difícil.
10-30	Máster universitario	Lectura muy difícil. Un universitario entiende mejor el texto.
0-10	Doctorado/Experto	Lectura extremadamente difícil.

Figura 22: Tabla de Fernández Huerta para determinar la dificultad de lectura en textos en español.

2.1.2. Flujo de datos

Para garantizar que el usuario logre una buena solución tras el entrenamiento, los documentos subidos por el usuario deberán seguir una serie de preprocesados.

En primer lugar, se llevará a cabo una verificación para detectar documentos duplicados, de esta forma, evitamos introducir ruido en los modelos durante el entrenamiento.

Seguidamente, se valorarán las métricas de los documentos. En este caso, valoraríamos la facilidad de lectura que tiene el texto, advirtiendo al usuario en caso de subir documentos demasiado complejos.

Finalmente, se aplicaría una ofuscación a los datos sensibles que pudiese contener el documento. Este procesado se basaría en anonimizar ciertas etiquetas o cifrar ciertos campos, que pudiesen tener este tipo de información sensible. Por ejemplo:

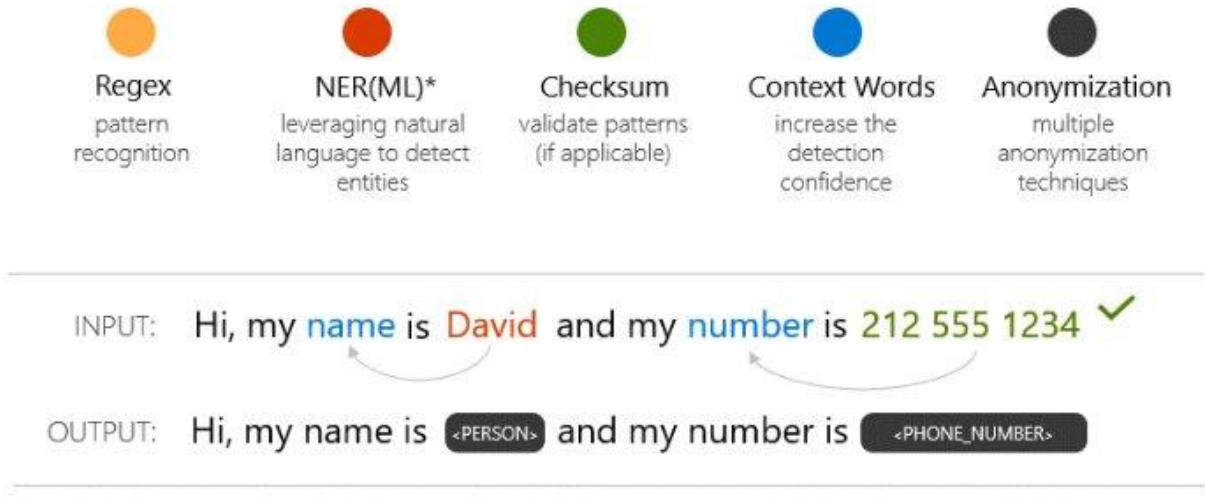


Figura 23: Ejemplo de anonimizar datos.

2.2. Datos para optimizador multicriterio

Aunque la cápsula estará desarrollada para procesar diferentes conjuntos de datos, es necesario disponer de uno o varios conjuntos de muestra para poder realizar las pruebas y la validación inicial del sistema. Por este motivo, se han seleccionado un conjunto de datos relacionados con el demostrador de la ACT6, que ofrecen un contexto operativo representativo y permiten evaluar adecuadamente el comportamiento del optimizador en escenarios reales y controlados.

El **optimizador multicriterio** se alimenta de un proceso de preparación y evaluación de datos AIS que convierte los ficheros brutos en conjuntos consistentes, representativos y auditables. Desde el inicio se asegura la representatividad de las fuentes para el problema a resolver, y se describen las limitaciones del conjunto (valores ausentes, redundancias, inconsistencias) junto con las medidas para mitigarlas.

El flujo de datos aplica salvaguardas que garantizan:

- **Coherencia temporal:** series ordenadas y continuas por embarcación.
- **Validez geoespacial:** exclusión de segmentos que crucen tierra o zonas no navegables.
- **Plausibilidad cinemática:** control de rangos (p. ej., velocidades físicamente viables).
- **Calidad intrínseca:** detección de datos faltantes, imputación controlada cuando procede, eliminación de duplicados y limpieza de registros inadecuados.
- **Balance de clases:** integración equilibrada entre trayectorias normales y **casos anómalos verosímiles** para cubrir huecos y evitar sesgos operativos.

Con estos datos depurados se generan representaciones de las trayectorias sobre las que el optimizador compara alternativas atendiendo a múltiples dimensiones, como la calidad del dato (completitud, consistencia y exactitud), o la eficiencia de procesamiento y robustez frente a ruido. Todas las etapas quedan trazadas y versionadas, con políticas de privacidad definidas antes del desarrollo, y con almacenamiento escalable y controlado que permite el análisis posterior y la auditoría.

En síntesis, el enfoque garantiza que los datos de entrada sean idóneos, a la vez que cumplen criterios de transparencia, reproducibilidad y gobernanza exigibles en entornos de evaluación pública.

2.2.1. Datos necesarios

Los datos utilizados para el optimizador multicriterio corresponden a datos del Sistema de Identificación Automática (AIS, por sus siglas en inglés), cuyo objetivo es permitir a embarcaciones comunicar información relevante como su posición, para que otras embarcaciones o instalaciones puedan conocerlas y evitar así colisiones.

Este sistema es de obligada implementación para embarcaciones con un arqueo bruto superior a 500GT, buques en viaje internacional con arqueo bruto superior a 300GT, y todos los buques de pasaje, independientemente de su tamaño.

Los datos cubren la totalidad del año 2024 y se estructuran en archivos CSV, uno por cada día del año. Cada CSV está disponible para descarga empaquetado en un archivo ZIP. El formato de nombre de estos archivos es AIS_<AÑO>_<MES>_<DIA>.zip, teniendo para el 1 de enero de 2024, el siguiente nombre AIS_2024_01_01.zip

El conjunto de datos completo en formato ZIP tiene un tamaño en memoria de 116.7GB. Este tamaño en disco se puede multiplicar x4 al descomprimir; un ZIP de este conjunto de datos con un peso de 200MB puede contener un CSV de aproximadamente 800MB.

Cada CSV contiene datos tabulares con 17 columnas:

- **MMSI:** Maritime Mobile Service Identity, es un número de 9 dígitos asignado a un transmisor de una embarcación para transferir información de registro a la guardia costera de los estados unidos para uso en situaciones de emergencia.
- **BaseDateTime:** El momento en que el mensaje fue transmitido, normalmente en UTC.
- **LAT:** La posición norte-sur de la embarcación.
- **LON:** La posición este-oeste de la embarcación.
- **SOG (Speed Over Ground):** velocidad de la embarcación relativa a la superficie terrestre.
- **COG (Course Over Ground):** dirección en la que la embarcación se está moviendo relativa a la superficie terrestre.
- **Heading**: Rumbo del buque en grados.
- **VesselName**: Nombre del buque.
- **IMO:** International Maritime Organization number - Número único e inmutable que identifica a cada barco a nivel mundial. Solo los barcos grandes (>300 toneladas) lo tienen.
- **CallSign:** Identificador único asignado a la embarcación, usado para comunicaciones de radio.
- **VeselType:** Tipo de embarcación. Valor entero que identifica el tipo, como por ejemplo 30-Barco pesquero o 52-remolcador.
- **Status:** valor entero que describe el estado de la navegación (fondeando, en ruta, pescando, etc.); siempre debe estar actualizado.
- **Length:** largo de la embarcación.
- **Width:** ancho de la embarcación.

- **Draft:** calado del barco, expresado en metro y refleja el calado actual o máximo de la embarcación.
- **Cargo**: Código de carga.
- **TransceiverClass:** Clase del transceptor AIS.

Variables núcleo para el optimizador. Para construir representaciones comparables entre configuraciones, el optimizador emplea principalmente MMSI, BaseDateTime, LAT, LON, SOG y, de forma opcional, COG/Heading. Estos campos permiten reconstruir trayectorias temporales y extraer rasgos cinemáticos robustos; el resto se conservan como metadatos (p. ej., tipología de buque, dimensiones, estado).

Cobertura y volumetría. La granularidad diaria de 2024 habilita muestreos por fecha y estratificación por regiones operativas. El volumen bruto (116,7 GB comprimido) hace aconsejable un filtrado selectivo, así como controles de valores ausentes, duplicados y rangos físicos plausibles.

2.2.2. Flujo de los datos

El flujo de tratamiento transforma los ficheros AIS diarios en trayectorias válidas y en representaciones multicanal estandarizadas sobre las que el optimizador multicriterio puede comparar configuraciones. Este pipeline se diseña para garantizar que los datos sean representativos del problema, estén limpios y consistentes, y puedan ser utilizados en un entorno reproducible y trazable.

Los objetivos principales son:

- Asegurar que las fuentes seleccionadas reflejan con fidelidad el dominio marítimo real.
- Identificar y corregir limitaciones inherentes a los datos AIS (valores ausentes, duplicados, errores de formato o ruido).
- Mantener coherencia temporal, espacial y cinemática en todas las trayectorias.
- Ampliar el conjunto con trayectorias anómalas verosímiles, para cubrir huecos y evitar sesgos en la comparación de configuraciones.
- Documentar todas las transformaciones realizadas, de modo que se garantice trazabilidad y reproducibilidad.
- Proteger la privacidad y asegurar un almacenamiento eficiente y seguro durante todo el ciclo de tratamiento.

Descripción secuencial del pipeline

1. **Ingesta**
   - Entrada y conversión de ficheros de NOAA al formato pertinente.
   - Catalogación por fecha y fuente.
   - La selección de la fuente se justifica porque cubre un periodo completo y representa de forma adecuada la operativa marítima.

2. **Limpieza y validación básica**
   - Eliminación de registros duplicados o incompletos.
   - Corrección de formatos y estandarización de unidades.

- Validación de valores plausibles.

3. **Validación geoespacial**
   - Exclusión de segmentos incorrectos atendiendo a criterios geoespaciales.
   - Detección de incoherencias en posición o velocidad.

4. **Reconstrucción de trayectorias**
   - Agregación de mensajes por MMSI en ventanas temporales diarias.
   - Requisitos mínimos de densidad de puntos y cobertura temporal para considerar una trayectoria válida.
   - Aplicación de interpolación en casos justificados para mantener continuidad y representatividad.

Ejemplo aplicado de un día de datos (referencia 2024): En una jornada típica, el dataset bruto contenía **7.296.275 mensajes AIS individuales**, que tras una primera depuración básica quedaron en **7.276.501 registros válidos**. Esto implicó la eliminación de unos **19.774 puntos** por inconsistencias de formato, duplicados o incoherencias en los campos principales.

Adicionalmente, se identificaron y descartaron **1.402 puntos** con saltos imposibles en velocidad o posición (outliers cinemáticos), garantizando que las trayectorias retenidas fuesen físicamente plausibles.

En cuanto a las trayectorias, se partía de **14.837 tracks brutos** reconstruidos a partir de los mensajes. Tras la aplicación de filtros de densidad, continuidad temporal y plausibilidad cinemática, el conjunto se redujo a **2.439 trayectorias netas y válidas**.

5. **Ampliación con casos anómalos verosímiles**
   - Generación controlada de trayectorias anómalas.
   - Se simulan patrones anómalos realistas.
   - De esta manera se refuerza la diversidad del dataset y se evita que el optimizador trabaje con distribuciones sesgadas.

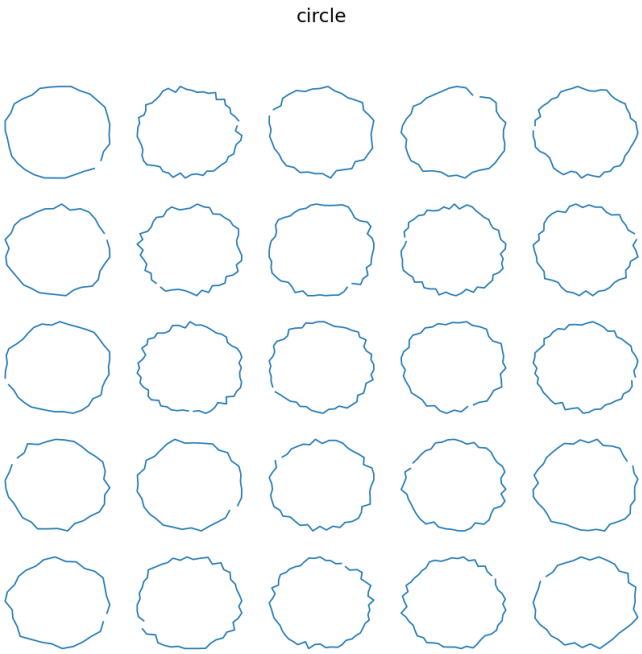


Figura 24: Trayectoria sintética tipo círculo (maniobra cerrada).

Figura 25: Ejemplos de trayectorias reales reconstruidas a partir de datos AIS (2024).

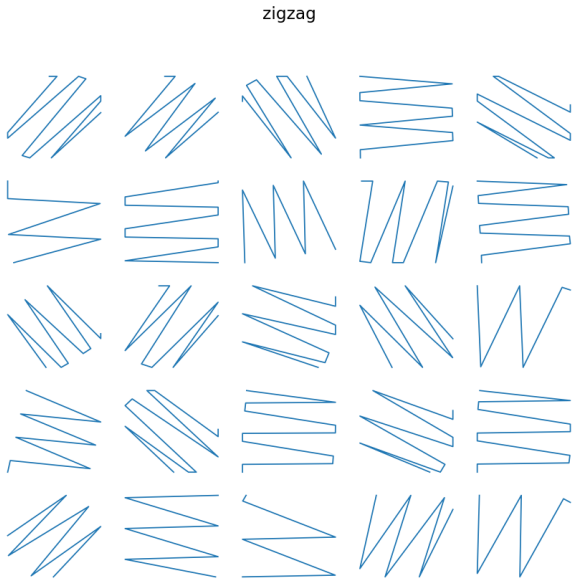


Figura 26: Trayectoria sintética en zigzag (comportamiento anómalo).

6. **Consolidación y particionado**
   - Unión de trayectorias reales y sintéticas en un conjunto equilibrado.
   - Estratificación temporal (meses, estaciones) y geográfica (zonas de operación) para garantizar representatividad.
   - El particionado permite separar claramente datos de entrenamiento, validación y prueba, manteniendo homogeneidad entre subconjuntos.

7. **Representación multicanal**
   - Conversión de trayectorias en representaciones tensoriales/imágenes con múltiples canales:
     - Geometría (LAT/LON).
     - Velocidad y variaciones direccionales (SOG, COG, Heading).
     - Evolución temporal.

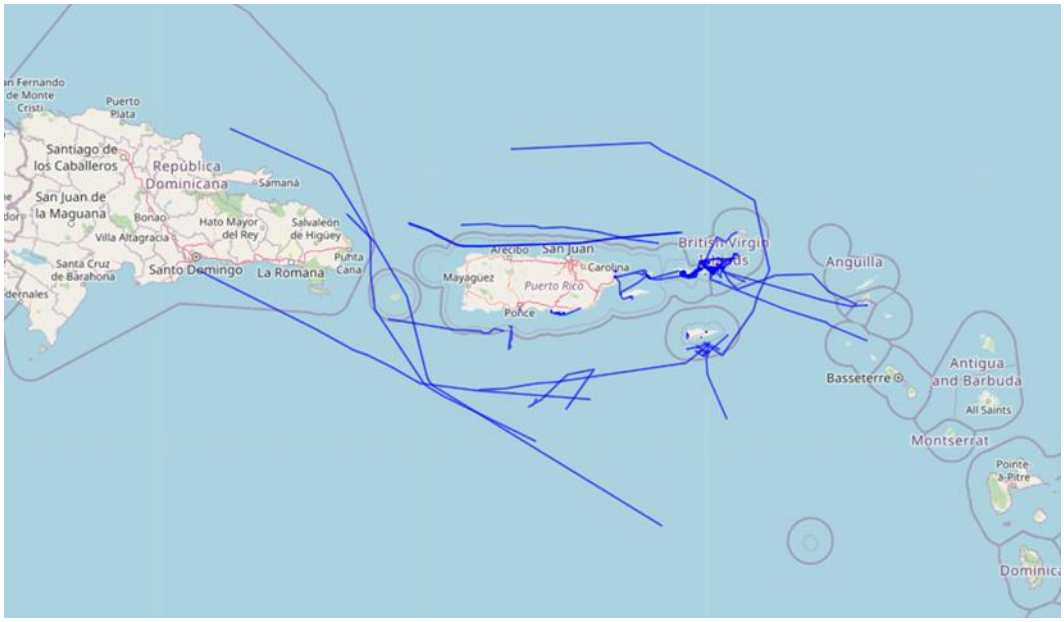


Figura 27: Validación geoespacial en Caribe (Puerto Rico/BVI). Azul: Trayectoria real.

Figura 28: Validación geoespacial en litoral de California (LA–Catalina). Azul: Trayectoria real, Rojo: Trayectoria sintética.

Tabla 4: Fases y artefactos generados

Fase	Entrada	Proceso clave	Salida/Artefacto
Ingesta	Ficheros AIS diarios (ZIP/CSV)	Descarga verificada y catalogación	CSV diarios catalogados
Limpieza/validación	CSV diarios	Depuración; rangos plausibles; coherencia temporal	Puntos limpios y ordenados
Validación geoespacial	Puntos limpios	Exclusión de tierra/zonas no navegables	Segmentos navegables
Reconstrucción de tracks	Segmentos navegables	Agregación por MMSI y ventana diaria	Trayectorias diarias válidas
Casos sintéticos	Trayectorias válidas	Generación de anomalías plausibles	Trayectorias sintéticas
Consolidación	Reales + sintéticas	Balance y estratificación	Dataset equilibrado
Representación	Dataset equilibrado	Rasterización/featurización multicanal	Tensores/imágenes estandarizadas

Tabla 5: Controles de calidad y criterios de aceptación

Control	Criterio (ejemplo)	Acción si no cumple
Duplicados/NaN	0 duplicados; sin nulos en columnas clave	Eliminar o imputar de forma trazable
Rango cinemático (SOG)	0–70 nudos; outliers justificados	Descartar o revisar segmentación
Coherencia temporal	Timestamps ordenados; sin solapes anómalos	Reordenar; excluir anomalías
Coherencia espacial	No cruce de tierra; dominio marítimo válido	Excluir tramos infractores
Densidad/cobertura	Mínimo de puntos y cobertura diaria representativa	Excluir trayectorias insuficientes
Balance de clases	Proporción real/sintético dentro de tolerancias	Reajustar composición
Trazabilidad	Registro de decisiones y cambios	Documentar y versionar

2.2.3. Ética

Los datos utilizados son de carácter público y son facilitados por el *Navigation Center* de la guardia costera de los Estados Unidos. Estos datos solamente ofrecen datos sobre la posición y el tipo de navegación que una embarcación ha realizado y no sobre los integrantes de la tripulación, por lo que no se facilitan datos personales de ningún tipo.

Se podría argumentar que estos datos reflejan patrones sobre las embarcaciones que podrían ser utilizados por actores maliciosos para establecer objetivos militares u organización de operaciones de asalto o robo de mercancías en alta mar. Sin embargo, estos posibles daños o perjuicios son mitigados por el hecho de que estos datos no se publican de forma inmediata ni en tiempo real y los patrones están sujetos a cambios.

En el momento de escribir este documento, no hay datos públicos publicados sobre el año en curso y los datos de 2024 se publicaron casi dos meses después de que finalizara el año, el 25 de febrero de 2025.

3. Finalidad del tratamiento

El tratamiento de los datos en el marco del proyecto CAPSUL-IA tiene como objetivo principal el desarrollo, evaluación y validación de soluciones tecnológicas basadas en inteligencia artificial, en particular machine learning y deep learning, aplicadas a casos de uso definidos dentro del contexto del proyecto. Estas actividades se enmarcan en iniciativas de investigación e innovación subvencionadas por organismos públicos, como el CDTI y el Ministerio de Ciencia e Innovación.

Las finalidades del tratamiento se alinean con el principio de minimización de datos, garantizando que únicamente se procesen aquellos datos estrictamente necesarios para alcanzar los fines legítimos del proyecto, conforme a lo dispuesto en el artículo 5.1.c del RGPD.

En concreto, las finalidades del tratamiento comprenden:

- El entrenamiento, ajuste y evaluación de modelos de IA para tareas de análisis visual, seguimiento de personas u objetos, y generación de mapas de calor basados en datos de vídeo e imagen pública (por ejemplo, datasets como PersonPath22).
- La creación y validación de bases de datos jurídicas utilizadas en el diseño de modelos de lenguaje (LLM) con capacidades de búsqueda inteligente y generación aumentada por recuperación (RAG), a partir de documentos legales provenientes de fuentes oficiales como EUR-Lex o el Poder Judicial Español.
- La mejora de técnicas de anonimización, aprendizaje federado y generación de datos sintéticos, como parte de la investigación sobre métodos que permitan preservar la privacidad en entornos de IA.

Cabe destacar que los datos empleados en las fases de diseño y testeo no forman parte del producto final y no son explotados comercialmente. La solución final está diseñada para que sea el usuario quien introduzca los datos que considere oportunos, de forma que se respete la soberanía y responsabilidad sobre el contenido gestionado.

En ningún caso se contemplan tratamientos con fines de elaboración de perfiles o decisiones automatizadas que produzcan efectos jurídicos sobre los interesados, salvo

que tales funcionalidades sean habilitadas por el usuario final bajo su responsabilidad y conforme al marco normativo aplicable.

El tratamiento de datos se lleva a cabo en entornos controlados, seguros y bajo protocolos que garantizan la confidencialidad, integridad y disponibilidad de la información. Todo ello, en cumplimiento estricto del Reglamento General de Protección de Datos (RGPD) y la Ley Orgánica de Protección de Datos Personales y garantía de los derechos digitales (LOPDGDD).

4. Medidas de Seguridad y Protección

4.1. Técnicas de privacidad

A continuación, se detallan diferentes técnicas aplicadas en el campo de la inteligencia artificial en materia de protección de datos. El documento de referencia en este apartado ha sido "*State-of-the-Art Report: The Potential of Privacy Tech & Computer Vision in Europe*", publicado en septiembre de 2021 por el KI Bundsverband (Asociación Alemana de Inteligencia Artificial) en colaboración con iconomy y la Oficina Federal de Seguridad de la Información (BSI) y que ofrece una visión integral sobre el emergente campo de la tecnología de privacidad (*Privacy Tech*) en Europa, con un enfoque particular en su aplicación dentro de la visión por computador.

Este informe destaca cómo, tras la implementación del RGPD de 2018, surgieron desafíos significativos para las empresas y organizaciones al intentar equilibrar el cumplimiento normativo con la innovación tecnológica. Sin embargo, este entorno regulatorio también ha catalizado el desarrollo de soluciones tecnológicas que integran la privacidad desde el diseño, transformando la privacidad en una ventaja competitiva.

4.1.1. Anonimización de datos profunda

El estado del arte en anonimización de datos para imágenes y videos es un área activa de investigación. Esta tecnología se centra en detectar identificaciones personales, como caras o matrículas, y generar un reemplazo sintético que refleja los atributos originales sin ser identificables. Este proceso se realiza mediante algoritmos de visión por computadora y técnicas de aprendizaje profundo.

Figura 29: Ejemplo de pixelado de caras.

La anonimización de datos profunda no solo protege las identidades, sino que también mantiene la información relevante para algoritmos de Machine Learning. Por ejemplo, la dirección de la mirada, la información sociodemográfica o las emociones reflejadas

pueden ser preservadas sin comprometer la privacidad del individuo. Esto beneficia la utilización de datos visuales para aplicaciones en las que se necesita analizar comportamientos humanos sin violar la privacidad, como en estudios de mercado, análisis de tráfico peatonal y sistemas de seguridad.

Un caso de uso de esta tecnología es su aplicación en el servicio de tren alemán para la anonimización de datos en el análisis de densidad de pasajeros. Mediante la anonimización de las imágenes capturadas en las estaciones y dentro de los trenes, se puede obtener información útil sobre el flujo de pasajeros y la ocupación de los vagones sin identificar a los individuos. Esto permite a la empresa optimizar sus servicios y mejorar la experiencia del usuario sin comprometer la privacidad.

Además, la anonimización de datos profunda se está utilizando en el ámbito de la salud para proteger la privacidad de los pacientes en imágenes médicas. Por ejemplo, en radiografías y resonancias magnéticas, se pueden anonimizar las características faciales mientras se preserva la información clínica relevante. Esto facilita la colaboración y el intercambio de datos entre instituciones médicas y de investigación sin riesgos de violación de la privacidad.

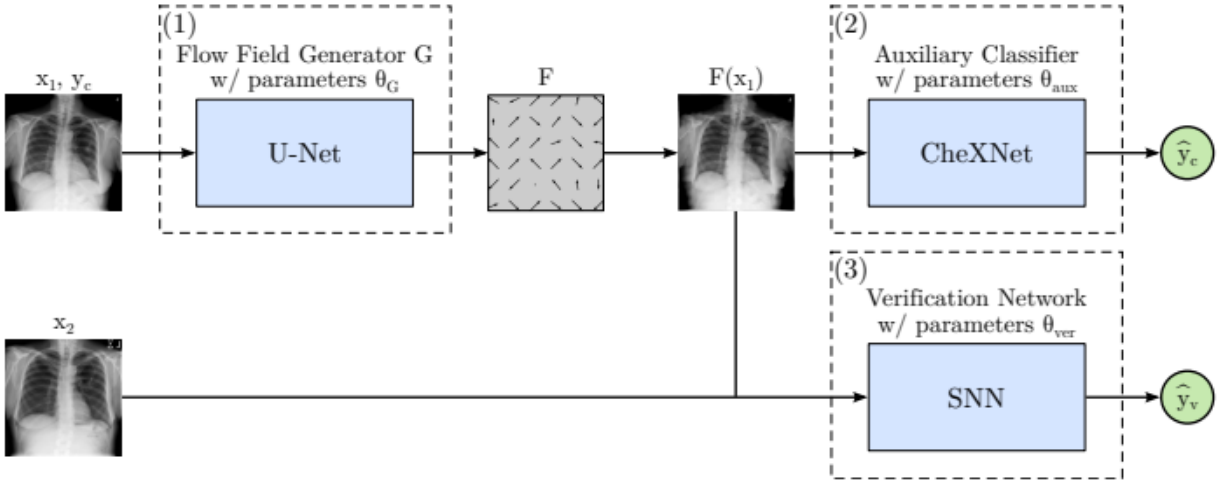


Figura 30: Ejemplo anonimización de datos médicos

Otra área de aplicación es la vigilancia y seguridad. Las cámaras de seguridad en espacios públicos pueden utilizar técnicas de anonimización para proteger la identidad de las personas mientras se monitorean actividades sospechosas. Esto cobra una importante relevancia en países con estrictas leyes de privacidad y protección de datos.

Otros modelos como DeepPrivacy2, han sido entrenados para aplicar una anonimización de cuerpo completo, no solo sobre los rostros.

Figura 31: Ejemplo anonimización de cuerpo completo.

Esta clase de anonimización, además de servir como técnica para asegurar el anonimato de las personas que aparecen en las imágenes, ofrecen también la posibilidad de ser usadas como técnica de aumentado de datos, disminuyendo la necesidad de la recopilación de un mayor volumen de datos originales.

4.1.2. Aprendizaje federado

El Aprendizaje Federado (FL por sus siglas en inglés) es una técnica de entrenamiento de modelos descentralizada sin que los datos lleguen a compartirse. Un caso sería el de descargarse un modelo dado localmente y hacer un entrenamiento o ajuste sobre datos propios. Una vez hecho este ajuste, se actualiza el modelo o sus pesos y de esta forma los datos nunca llegan a ser compartidos y la información personal permanece privada.

Este enfoque persigue la oportunidad de capturar una variabilidad más elevada en los datos de entrenamiento. Por contra, conlleva un mayor riesgo de que los datos sean “memorizados” y tampoco se tiene una trazabilidad de los datos y por tanto dificulta la investigación de los resultados finales de un modelo.

Un caso de uso conocido de esta técnica es la predicción de palabras en los Google Keyboard donde el teléfono almacena información del contexto localmente para hacer sus sugerencias.

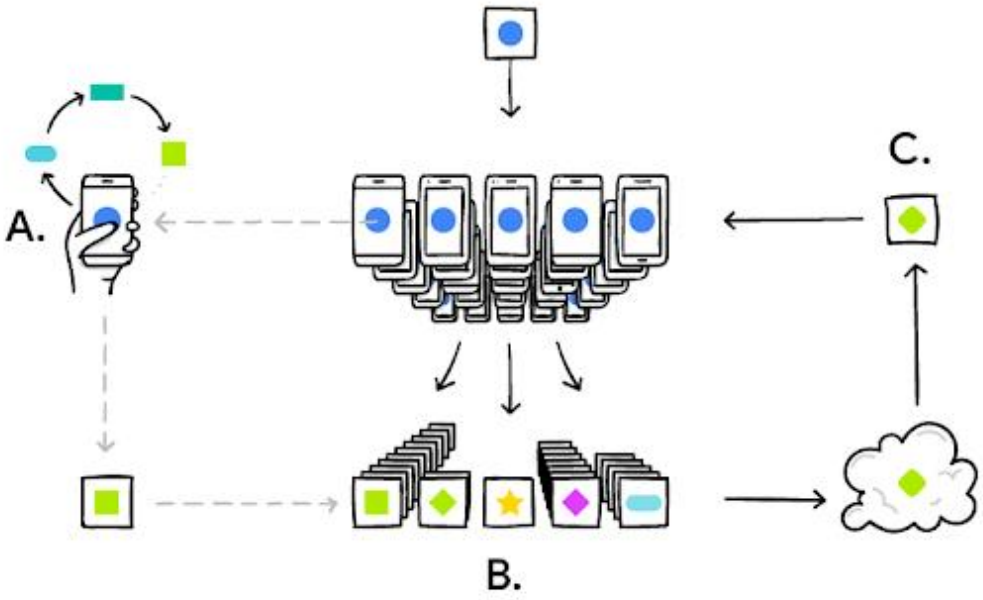


Figura 32: Esquema de la técnica de aprendizaje federado usada por Google para el Gboard de Android

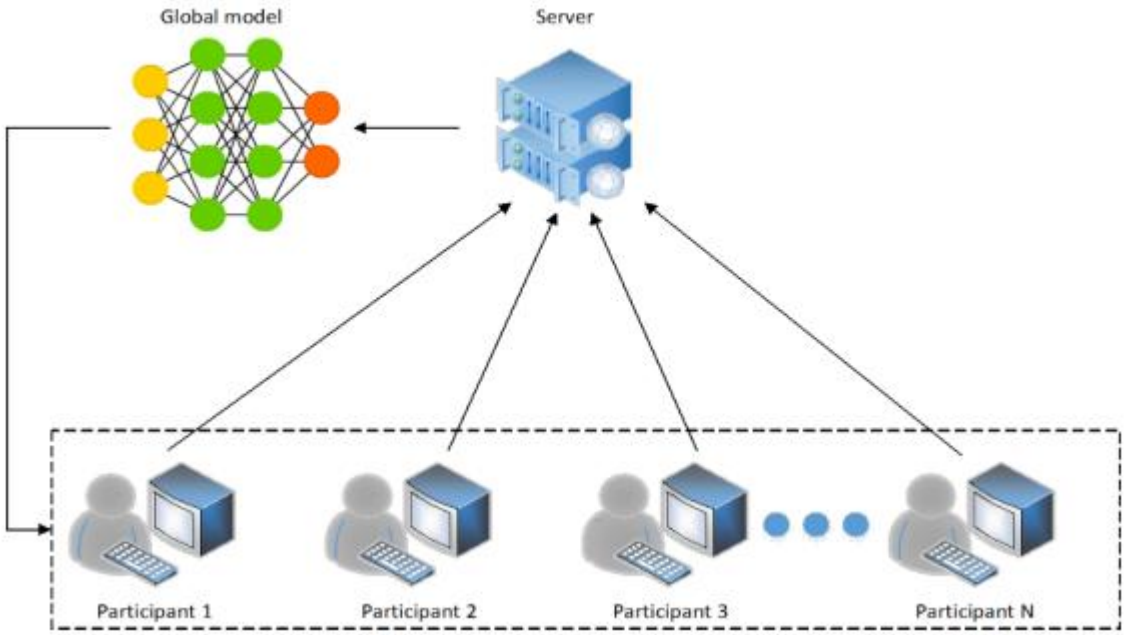


Figura 33: Esquema básico FL. Extraído de referencia de Springer. Valorar cuál es mejor, si ésta o la de google, pero no dejar las 2

FL se puede definir como una técnica de machine learning distribuido que de forma cooperativa ejecuta algoritmos en múltiples dispositivos edge o servidores bajo la

condición de que los datos no abandonan el entorno local. FL transfiere la tarea de entrenamiento a cada cliente local y la comunicación entre el cliente y el servidor se hace a través de parámetros y no de los datos directamente. El servidor solo participa en una agregación de parámetros para actualizar el modelo global.

El funcionamiento general es el siguiente; primero, cada cliente descarga el modelo inicial; posteriormente, cada cliente entrena el modelo sobre sus datos locales hasta que el modelo local converge; tercero, el cliente encripta los parámetros del modelo y los sube al servidor; por último, el servidor agrega los gradientes o parámetros y actualiza el modelo global.

La investigación actual sobre esta técnica trata de solventar tres principales cuellos de botella: las amenazas a la privacidad y seguridad, la heterogeneidad y elevado coste en comunicación entre dispositivos locales y servidores. Comparado con otras técnicas tradicionales de privacidad, FL se caracteriza por exponer ciertos parámetros y asumir que esos parámetros no reflejan ni revelan información sensible, aunque un creciente número de investigaciones han encontrado que esta hipótesis no es necesariamente objetiva, por lo que la investigación para la protección efectiva de la privacidad sigue siendo un punto importante en este campo.

4.1.3. Homomorphic Encryption

El problema del encriptado clásico de los datos es que estos necesitan ser desencriptados para ser usados, lo que los hace vulnerables. Con el encriptado homomórfico se pueden realizar tareas computacionales sobre datos encriptados sin tener que desencriptarlos.

Esta técnica protege los detalles sensibles de los datos, aunque aún permite que esta sea analizada y procesada. Además, requiere de unos recursos computacionales considerables y tiende a ser un proceso lento hasta el punto de que actualmente su uso no es práctico en determinadas aplicaciones.

4.1.4. Datos sintéticos

Los datos simulados son aquellos generados por un algoritmo o simulación como alternativa a datos del mundo real, con el fin de eliminar o minimizar la información personal presente en ellos.

Esta técnica mejora la diversidad de los datos en aquellos conjuntos de datos pequeños o limitados y es particularmente útil en tareas de visión por computador. Otra ventaja de este tipo de generación de datos es que se pueden compartir de forma pública sin afectar a la seguridad o privacidad de ninguna persona.

Por otro lado, la generación de datos sintéticos conlleva el desafío extra de que estos no siempre se asemejan a datos reales, caso más evidente cuando se trata de imágenes

donde a simple vista se puede apreciar. Además, este tipo de generación tiene sus límites ya que, en términos de diversidad, los modelos generativos no representan completamente la realidad.

4.2. Medidas de Seguridad y Protección en datos y documentos de texto

Los documentos recuperados, al ser de tipo legal, están sujetos a mostrar información sensible *per se*. Con esto en mente se han investigado y desarrollado estrategias para la ofuscación de información sensible. Concretamente se ha hecho uso de Presidio, una biblioteca desarrollada por Microsoft para dar soporte en la protección y ocultación de información en texto.

Esta biblioteca aplica esta anonimización en diferentes pasos:

1. **Expresiones regulares (Regex):** sirven para la identificación de patrones de texto como números de teléfono o direcciones de correo electrónico.
2. **Reconocimiento de Entidades Nombradas (NER por sus siglas en inglés):** es una técnica de procesamiento de lenguaje natural que identifica y categoriza entidades dentro de un texto como personas, organizaciones, localizaciones, fechas, horas o cantidades.
3. **Anonimización:** una vez detectadas las partes del texto sensibles, se anonimizan.

La anonimización se lleva a cabo bien sustituyendo la entidad por una etiqueta, estrategia que una vez aplicada, no es posible recuperar la información adicional; o bien, aplicando una clave de cifrado, cifrar esa información de forma que pueda ser recuperable.

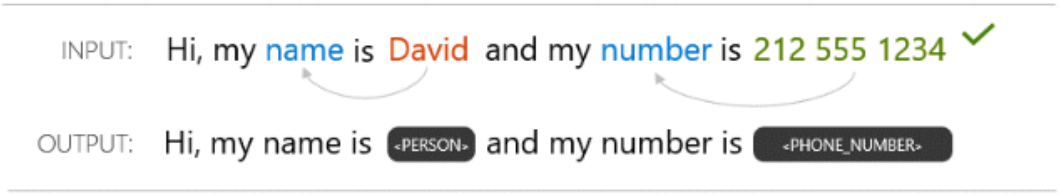


Figura 34: Anonimización de texto.

El punto donde se aplique esta anonimización depende en gran medida del caso de uso particular donde, dependiendo del contexto, se anonimicen los datos antes de ser guardados en la base de datos, o bien se aplique esta ofuscación en la respuesta generada por el modelo de lenguaje.

Por otra parte, y también dependiendo del contexto, esta ocultación de información se puede aplicar solamente a determinados usuarios en casos en las que una empresa o entidad, convenga en ocultar este tipo de información a empleados más nuevos o, en presentaciones ante potenciales clientes, demostrar la potencia de su modelo de negocio sin poner en peligro la filtración de información confidencial.

4.2.1. Control de redundancia y duplicidad de datos

Bajo el uso de la técnica Retrieval-augmentedgeneration (RAG), las bases de datos se construyen a base de fragmentos de texto denominados *chunks*. Estos fragmentos se realizan en base a una configuración previa y en principio estática e independiente al modelo de lenguaje que se use posteriormente para procesar el contexto recuperado de esa base de datos.

En el uso de esta técnica se ha tenido en cuenta el evitar almacenar fragmentos de texto duplicados en la base de datos. Hay que tener en cuenta que esto no se puede evitar con una comprobación de nombres de los documentos ya que estos están sujetos a cambios y no es algo fiable como descriptor del contenido.

La aproximación que se ha seguido consiste en generar un identificador único, de forma reproducible, a cada fragmento de texto. Esto se consigue utilizando algoritmos de criptografía que dada una entrada siempre generen el mismo resultado. Un ejemplo sencillo sería el uso de MD5 al que, dada como entrada el texto a almacenar, genera un código de 128 bits, representado típicamente como un número de 32 símbolos hexadecimales. Elegido el algoritmo específico, sería el *hash* generado el identificador que se le asignará a ese texto en la base de datos.

4.3. Medidas de Seguridad y Protección en datos en formato imagen y video

En lo referente a los datos destinados a modelos de visión, pese a ser conjuntos de datos públicos y evitar cualquier tipo de filtración de datos que pueda ser explotable en las aplicaciones finales y en los modelos distribuidos, se ha aplicado, en primer lugar, una ofuscación de los metadatos en imágenes y videos para ocultar, por ejemplo, la identificación de lugares que pudieran aparecer, en los nombres de los archivos.

Por otro lado, y con el fin de que los modelos entrenados no contengan ningún tipo de aprendizaje sobre personas concretas y evitar el posible uso malintencionado de los modelos para la identificación de nadie, se ha aplicado a los datos de entrenamiento un modelo de FaceDetection para detectar las zonas de la imagen donde aparecen caras para aplicar, posteriormente, un difuminado o emborronamiento en esa zona.

5. Almacenamiento y Procesado de datos

5.1. Propuesta de Arquitectura ETL

La adecuada ingesta y almacenamiento de datos es un componente esencial dentro del ciclo de tratamiento planteado en el proyecto CAPSUL-IA, tanto desde el punto de vista funcional como desde la perspectiva de trazabilidad, seguridad y posterior explotación de los datos.

La arquitectura diseñada combina distintas tecnologías especializadas, con el objetivo de garantizar la escalabilidad, trazabilidad y flexibilidad necesarias para el tratamiento de datos en contextos heterogéneos. Esta arquitectura está pensada para capturar información tanto de fuentes estructuradas como no estructuradas, y tanto en tiempo real como en procesos batch.

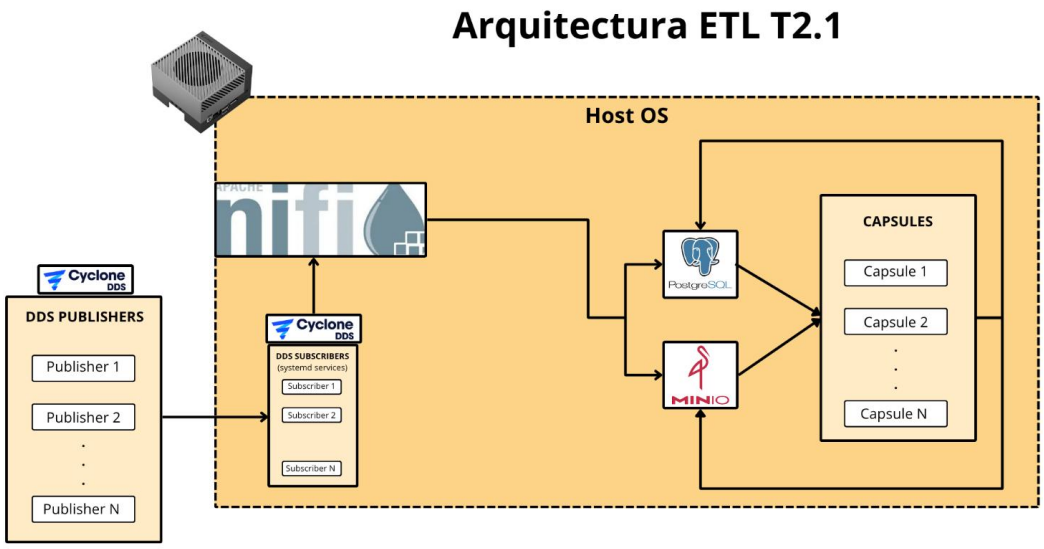


Figura 35: Esquema de la arquitectura para el almacenamiento de datos

En el sistema conviven múltiples puntos de entrada de datos. Por un lado, se dispone de un conjunto de **publicadores DDS** que transmiten información en tiempo real, como pueden ser sensores embarcados, dispositivos robóticos o servicios de monitorización industrial. Estos publicadores se comunican a través de **CycloneDDS**, un middleware basado en el estándar DDS (Data Distribution Service), que permite una comunicación eficiente, desacoplada y determinista entre productores y consumidores de datos en entornos distribuidos.

Por otro lado, se ha integrado **Apache NiFi** como sistema de ingesta principal para datos heterogéneos. NiFi no solo recibe datos desde los suscriptores DDS desplegados como servicios del sistema, sino que también puede conectar directamente con fuentes externas mediante APIs REST, ficheros, sistemas de mensajería, colas Kafka, bases de datos o incluso almacenamiento en la nube. De este modo, NiFi actúa como puerta de entrada unificada para datos de distinta naturaleza y procedencia.

Además de su capacidad para capturar datos, NiFi implementa un pipeline completo de procesamiento ETL (**Extract, Transform, Load**):

- **Extracción (Extract)**: Captura datos desde fuentes diversas —como sensores, logs, bases de datos, APIs o colas de mensajes— con control de frecuencia, priorización de flujos y tolerancia a fallos.
- **Transformación (Transform)**: Permite realizar operaciones intermedias como filtrado, enriquecimiento, limpieza, agregación, validación, conversión de formatos o anonimización de datos, todo ello mediante un entorno visual y configurable.
- **Carga (Load)**: Una vez transformados, los datos se enrutan hacia los sistemas de almacenamiento persistente, que en este caso son PostgreSQL para datos estructurados y MinIO para datos no estructurados u objetos binarios.

Los datos procesados se almacenan en dos sistemas principales. Por un lado, **PostgreSQL**, que actúa como base de datos relacional para información estructurada y transaccional, permitiendo consultas complejas y agregaciones sobre grandes volúmenes de datos. Por otro, **MinIO**, un sistema de almacenamiento de objetos compatible con el protocolo S3, que se utiliza para alojar datos no estructurados como imágenes, vídeos o archivos generados por sensores, ofreciendo una solución escalable y eficiente para grandes volúmenes de información binaria.

Finalmente, las **cápsulas de IA** acceden tanto a PostgreSQL como a MinIO para recuperar los datos necesarios en cada fase del entrenamiento, inferencia o validación. Además, las cápsulas también se encargan de almacenar los resultados generados (predicciones, métricas, etc.), permitiendo que toda la información relevante quede registrada y accesible para futuros usos, completando así el ciclo de vida del dato. Esta separación clara entre capas de ingesta, procesamiento y almacenamiento permite una mayor mantenibilidad del sistema, facilita su auditoría y mejora su capacidad de adaptación a nuevos requerimientos.

A diferencia de arquitecturas monolíticas o altamente acopladas, esta solución modular basada en estándares abiertos ofrece ventajas significativas en términos de resiliencia, extensibilidad y gestión del ciclo de vida del dato. La combinación de CycloneDDS y NiFi permite capturar datos desde entornos robóticos, edge o industriales de forma natural, mientras que PostgreSQL y MinIO **aseguran una persistencia eficaz** adaptada a la diversidad de formatos de datos manejados por el sistema.

5.2. Validación de Arquitectura ETL

Con el objetivo de validar la arquitectura ETL propuesta, se ha desarrollado un flujo completo en Apache NiFi que sirve como prueba de concepto (PoC) de la misma. Este flujo permite mostrar de manera práctica cómo los distintos componentes del sistema interactúan entre sí, desde la recepción de los datos en bruto hasta su transformación, almacenamiento y uso por parte de las cápsulas de IA.

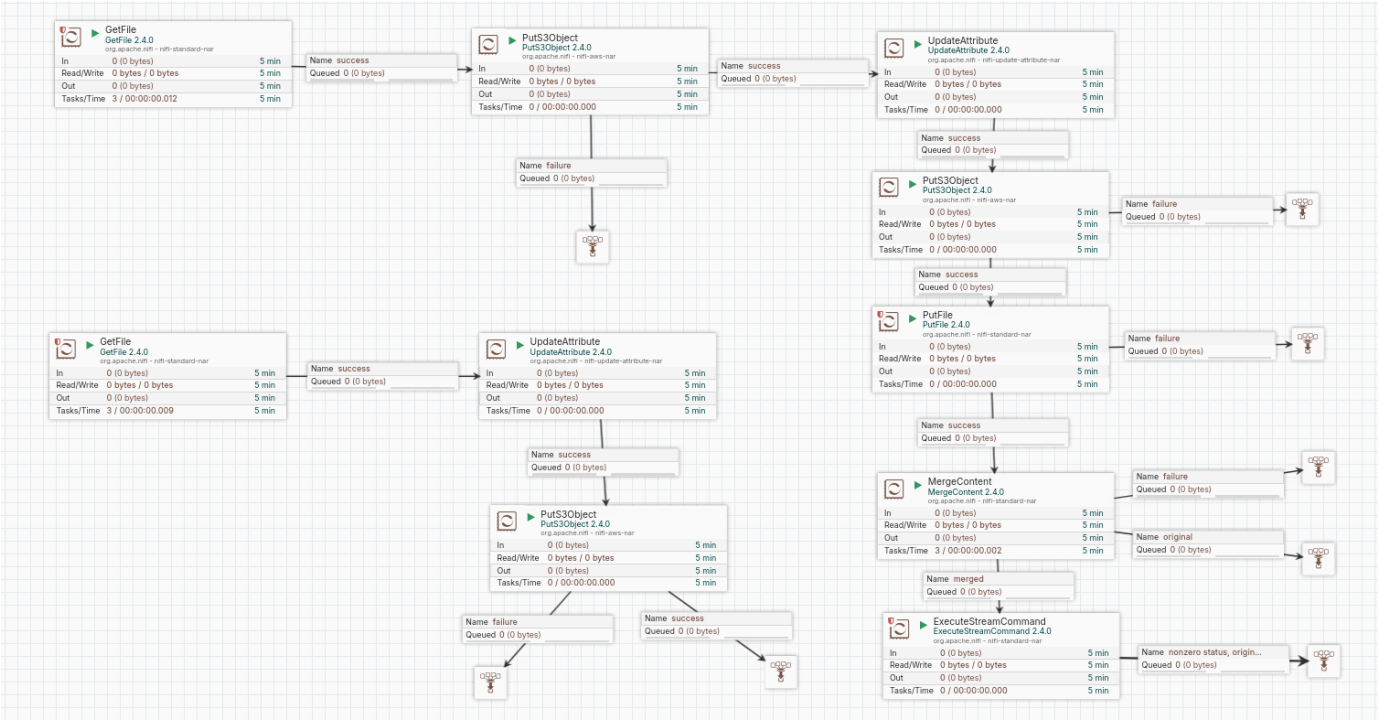


Figura 36: Flujo de Apache NiFi.

La PoC se centra en el tratamiento de imágenes para su posterior uso por parte de modelos de inferencia, reproduciendo un caso real de flujo de datos dentro del sistema de CAPSUL-IA. Este flujo de NiFi consta de las siguientes partes:

- **Ingesta de los datos:** a través del processor GetFile, se capturan las imágenes en bruto desde un directorio de la plataforma. Esta captura se realiza de forma continua, es decir, se recogen todos los nuevos archivos en cuanto están presentes en el directorio mencionado.
- **Procesamiento de los datos:** mediante los processors UpdateAttribute, se normalizan los nombres de las imágenes para así mantener la consistencia de estas en etapas posteriores.
- **Almacenamiento de los datos:** mediante los processors de PutS3Object, tanto las imágenes en bruto como las imágenes procesadas se guardan en diferentes rutas dentro de un mismo bucket de MinIO, manteniendo así una separación lógica de los datos para su posterior uso.
- **Ejecución de modelo de inferencia:** mediante el processor PutFile, las imágenes procesadas se copian a un volumen compartido con un contenedor Docker que ejecuta la cápsula de visión, la cual está basada en el modelo de Person Detection. Después, mediante el processor ExecuteStreamCommand, el contenedor ejecuta la inferencia de forma automática con las imágenes procesadas.
- **Almacenamiento de los resultados:** las imágenes generadas por la cápsula de visión son recogidas y procesadas por los processors GetFile y UpdateAttribute, respectivamente, y almacenadas en una nueva ruta del bucket de MinIO por medio del processor PutS3Object.

Esta PoC confirma que cada uno de los elementos que forman parte de la arquitectura ETL pueden integrarse e interactuar entre sí de manera coherente, segura y modular. Y aunque se trata de un flujo simplificado, consigue mostrar la viabilidad técnica de la arquitectura propuesta y servir como referencia para el diseño de los flujos que serán necesarios para cada caso de uso en específico.

6. Conservación y Eliminación de Datos

En cumplimiento del principio de limitación del plazo de conservación recogido en el artículo 5.1.e del Reglamento General de Protección de Datos (RGPD), los datos tratados en el marco del proyecto CAPSUL-IA serán conservados únicamente durante el tiempo estrictamente necesario para el cumplimiento de las finalidades específicas para las que fueron recogidos. Esta conservación se realizará siempre bajo condiciones que garanticen la seguridad, confidencialidad e integridad de la información.

Durante el desarrollo del proyecto, los datos permanecerán almacenados en entornos controlados y protegidos mediante medidas técnicas y organizativas adecuadas. Los conjuntos de datos utilizados para el entrenamiento y validación de modelos, como es el caso de conjuntos de datos públicos (por ejemplo, PersonPath22), serán conservados mientras su uso sea necesario para el desarrollo, evaluación y auditoría de los modelos diseñados en el marco del proyecto. En cuanto a los documentos legales empleados en fases de prueba —como los extraídos de EUR-Lex o del portal del Poder Judicial español— su almacenamiento será temporal, estrictamente ligado a la fase de diseño y testeo. Estos documentos no formarán parte del producto final ni se integrarán en ninguna base de datos accesible al usuario final.

Finalizado el ciclo de vida útil de los datos, y una vez alcanzadas las finalidades previstas, se procederá a su eliminación o anonimización irreversible, de conformidad con la normativa vigente. Los datos personales o pseudonimizados serán suprimidos utilizando procedimientos de borrado seguro, incluyendo técnicas de sobreescritura de archivos o eliminación mediante herramientas certificadas. Por su parte, aquellos datos que hayan sido debidamente anonimizados, y que por tanto no permitan la identificación de los individuos, podrán conservarse con fines científicos o de validación futura del modelo, siempre garantizando su carácter no identificativo.

En los casos en que se hayan empleado plataformas en la nube, se asegurará que los proveedores contratados ofrezcan mecanismos contrastables de destrucción de datos y cumplan con las exigencias del RGPD, incluyendo la disponibilidad de certificaciones y auditorías que validen la supresión efectiva de la información almacenada. Todas estas actuaciones serán debidamente registradas y verificadas internamente, en coherencia con las recomendaciones de la Agencia Española de Protección de Datos (AEPD) sobre la gestión responsable del ciclo de vida del dato.

7. Acrónimos y Abreviaciones

RGPD	Reglamento General de Protección de Datos
LOPDGDD	Ley Orgánica de Protección de Datos Personales y garantía de los derechos digitales
AEPD	Agencia Española de Protección de Datos
CCTV	Circuito Cerrado de Televisión
AIS	Sistema de Identificación Automática
RAG	Retrieval-augmentedgeneration
LLM	Large Language Model.
FL	Federal Learning
DDS	Data Distribution Service
ETL	Extract, Transform, Load

8. Referencias

[1] Abadi, M., Chu, A., Goodfellow, I., McMahan, H. B., Mironov, I., Talwar, K., & Zhang, L. (2016). Deep Learning with Differential Privacy. *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security - CCS'16*. https://doi.org/10.1145/2976749.2978318

[2] Acar, A., Aksu, H., Uluagac, A. S., & Conti, M. (2017, October 5). *A Survey on Homomorphic Encryption Schemes: Theory and Implementation*. ArXiv.org. https://doi.org/10.48550/arXiv.1704.03578

[3] AEPD. (2021, June). *Gestión del riesgo y evaluación de impacto en tratamientos de datos personales*. https://www.aepd.es/guias/gestion-riesgo-y-evaluacion-impacto-en-tratamientos-datos-personales.pdf

[4] *AUTOMATIC IDENTIFICATION SYSTEM USCG AIS Encoding Guide*. https://www.navcen.uscg.gov/sites/default/files/pdf/AIS/AISGuide.pdf

[5] Blum, M. G. B., Nunes, M. A., Prangle, D., & Sisson, S. A. (2013). A Comparative Review of Dimension Reduction Methods in Approximate Bayesian Computation. *Statistical Science*, *28*(2). https://doi.org/10.1214/12-sts406

[6] Bonawitz, K., Ivanov, V., Kreuter, B., Marcedone, A., McMahan, H. B., Patel, S., Ramage, D., Segal, A., & Seth, K. (2017). Practical Secure Aggregation for Privacy Preserving Machine Learning. *Cryptology EPrint Archive*. https://eprint.iacr.org/2017/281

[7] Brendan, M. H., Moore, E., Ramage, D., Hampson, S., & Arcas, Blaise Agüera y. (2016). *Communication-Efficient Learning of Deep Networks from Decentralized Data*. ArXiv.org. https://arxiv.org/abs/1602.05629

[8] CatalyzeX. (2017). *Input Perturbation: A New Paradigm between Central and Local Differential Privacy*. CatalyzeX. https://www.catalyzex.com/paper/input-perturbation-a-new-paradigm-between

[9] *Comité de los Derechos del Niño*. (n.d.). Retrieved November 3, 2025, from https://doncel.org.ar/wp-content/uploads/2023/03/Obs.-25-del-Instituto-Interamericano.pdf

[10] Common Crawl. (2024). *Common Crawl*. Common Crawl. https://commoncrawl.org/

[11] *CONGRESO DE LOS DIPUTADOS XV LEGISLATURA BOLETÍN OFICIAL DE LAS CORTES GENERALES PROYECTO DE LEY*. (n.d.). Retrieved November 3, 2025, from https://www.congreso.es/public_oficiales/L15/CONG/BOCG/A/BOCG-15-A-52-1.PDF

[12] Danger, R. (2022, May 19). *Differential Privacy: What is all the noise about?* ArXiv.org. https://doi.org/10.48550/arXiv.2205.09453

[13] Dwork, C., & Roth, A. (2014). The Algorithmic Foundations of Differential Privacy. *Foundations and Trends® in Theoretical Computer Science*, *9*(3-4), 211–407. https://doi.org/10.1561/0400000042

[14] European Commission. (2019). *Ethics guidelines for trustworthy AI | Shaping Europe's digital future*. European Commission. https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai

[15] Ghojogh, B., Samad, M. N., Mashhadi, Sayema Asif, Kapoor, T., Ali, W., Karray, F., & Crowley, M. (2019). *Feature Selection and Feature Extraction in Pattern Analysis: A Literature Review*. ArXiv.org. https://arxiv.org/abs/1905.02845

[16] Hukkelås, H., & Lindseth, F. (2022). *DeepPrivacy2: Towards Realistic Full-Body Anonymization*. ArXiv.org. https://arxiv.org/abs/2211.09454

[17] Hukkelås, H., Mester, R., & Lindseth, F. (2019). DeepPrivacy: A Generative Adversarial Network for Face Anonymization. *ArXiv:1909.04538 [Cs]*. https://arxiv.org/abs/1909.04538

[18] Joshi, N. (2019, June 26). The Present And Future Of Computer Vision. *Forbes*. https://www.forbes.com/sites/cognitiveworld/2019/06/26/the-present-and-future-of-computer-vision/?sh=41c92fe3517d

[19] Konečný, J., McMahan, H. B., Ramage, D., & Richtárik, P. (2016). Federated Optimization: Distributed Machine Learning for On-Device Intelligence. *ArXiv:1610.02527 [Cs]*. https://arxiv.org/abs/1610.02527

[20] Konečný, J., McMahan, H. B., Yu, F. X., Richtárik, P., Suresh, A. T., & Bacon, D. (2017). Federated Learning: Strategies for Improving Communication Efficiency. *ArXiv:1610.05492 [Cs]*. https://arxiv.org/abs/1610.05492

[21] Lee, G. Y., Alzamil, L., Doskenov, B., & Termehchy, A. (2021). A Survey on Data Cleaning Methods for Improved Machine Learning Model Performance. *ArXiv:2109.07127 [Cs]*. https://arxiv.org/abs/2109.07127

[22] Li, J., Cheng, K., Wang, S., Morstatter, F., Trevino, R. P., Tang, J., & Liu, H. (2018). Feature Selection: A Data Perspective. *ACM Computing Surveys*, *50*(6), 1–45. https://doi.org/10.1145/3136625

[23] Li, Q., He, B., & Song, D. (2021). *Model-Contrastive Federated Learning*. ArXiv.org. https://arxiv.org/abs/2103.16257

[24] Li, T., Sahu, Anit Kumar, Talwalkar, A., & Smith, V. (2019). *Federated Learning: Challenges, Methods, and Future Directions*. ArXiv.org. https://arxiv.org/abs/1908.07873

[25] Link to external site, this link will open in a new tab, & Link to external site, this link will open in a new tab. (2023). Leveraging Searchable Encryption through Homomorphic Encryption: A Comprehensive Analysis. *ProQuest*, 2948. https://doi.org/10.3390/math11132948

[26] Liu, B., Lv, N., Guo, Y., & Li, Y. (2023, January 3). *Recent Advances on Federated Learning: A Systematic Survey*. ArXiv.org. https://doi.org/10.48550/arXiv.2301.01299

[27] Liu, X. (2025). *Two-level Data Staging ETL for Transaction Data*. ArXiv.org. https://arxiv.org/abs/1409.1636v2

[28] Liu, Y., Carr, A., & Sun, L. (2024). *Empirical Perturbation Analysis of Linear System Solvers from a Data Poisoning Perspective*. ArXiv.org. https://arxiv.org/abs/2410.00878

[29] McMahan, B., & Ramage, D. (2017, April 6). *Federated Learning: Collaborative Machine Learning without Centralized Training*. Google Research. https://research.google/blog/federated-learning-collaborative-machine-learning-without-centralized-training-data/

[30] *Microsoft Presidio*. (n.d.). Microsoft.github.io. https://microsoft.github.io/presidio/

[31] Narayanan, A., & Shmatikov, V. (2024). *Robust De-anonymization of Large Datasets (How to Break Anonymity of the Netflix Prize Dataset)*. https://arxiv.org/pdf/cs/0610105

[32] Packhäuser, K., Gündel, S., Thamm, F., Denzinger, F., & Maier, A. (2022). *Deep Learning-based Anonymization of Chest Radiographs: A Utility-preserving Measure for Patient Privacy*. ArXiv.org. https://arxiv.org/abs/2209.11531

[33] Penedo, G., Malartic, Q., Hesslow, D., Cojocaru, R., Cappelli, A., Alobeidli, H., Pannier, B., Almazrouei, E., & Launay, J. (2023). *The RefinedWeb Dataset for Falcon LLM: Outperforming Curated Corpora with Web Data, and Web Data Only*. ArXiv.org. https://arxiv.org/abs/2306.01116

[34] *Principios de Protección de Datos*. (n.d.). https://privacyinternational.org/sites/default/files/2018-11/Parte%203%20-%20Principios%20de%20Proteccio%CC%81n%20de%20Datos_web.pdf

[35] *Principios éticos del tratamiento del dato*. (n.d.). Retrieved November 3, 2025, from https://datos.gob.es/sites/default/files/doc/file/OdD/Principioseticosdeltratamientodeldato.pdf

[36] *REGLAMENTO (UE) 2024/1689 DEL PARLAMENTO EUROPEO Y DEL CONSEJO*. (2024). https://eur-lex.europa.eu/legal-content/ES/TXT/PDF/?uri=OJ:L_202401689

[37] Sandoval-Segura, P., Singla, V., Geiping, J., Goldblum, M., Goldstein, T., & Jacobs, D. W. (2022). *Autoregressive Perturbations for Data Poisoning*. ArXiv.org. https://arxiv.org/abs/2206.03693

[38] Solove, D. (2007). “I’ve Got Nothing to Hide” and Other Misunderstandings of Privacy. *San Diego Law Review*, *44*(4), 745. https://digital.sandiego.edu/sdlr/vol44/iss4/5/

[39] Sweeney, L. (2000). *Simple Demographics Often Identify People Uniquely*. Sweeney. https://dataprivacylab.org/projects/identifiability/paper1.pdf

[40] Wen, J., Zhang, Z., Lan, Y., Cui, Z., Cai, J., & Zhang, W. (2022). A survey on federated learning: challenges and applications. *International Journal of Machine Learning and Cybernetics*, *14*(2), 513–535. https://doi.org/10.1007/s13042-022-01647-y

[41] Zhou, Y., Tu, F., Sha, K., Ding, J., & Chen, H. (2024). *A Survey on Data Quality Dimensions and Tools for Machine Learning*. ArXiv.org. https://arxiv.org/abs/2406.19614

[42] Zhu, Y., Yu, X., Chandraker, M., & Wang, Y.-X. (n.d.). *Private-kNN: Practical Differential Privacy for Computer Vision*. Retrieved November 3, 2025, from https://openaccess.thecvf.com/content_CVPR_2020/papers/Zhu_Private-kNN_Practical_Differential_Privacy_for_Computer_Vision_CVPR_2020_paper.pdf

[43] (2023). Unesco.org. https://unesdoc.unesco.org/ark:/48223/pf0000385841

9. Anexos

9.1. Manual de ética de datos

El presente Manual de Ética de Datos sirve como guía para usuarios de las cápsulas que no necesariamente tengan conocimientos técnicos. Está orientado a quienes necesiten preparar y alimentar datos en sistemas basados en modelos de lenguaje (incluyendo técnicas como *Retrieval-Augmented Generation*, RAG) y modelos de visión por computador. El objetivo es asegurar que el manejo de los datos siga buenas prácticas éticas, garantizando que la información utilizada por las diferentes cápsulas sea pertinente, segura y responsable.

9.1.1. Principios de ética de datos

Para asegurar un uso ético de los datos conviene seguir una serie de principios. Estos ayudan a evaluar si los datos que vamos a proporcionar a los modelos son adecuados y respetuosos con las personas y con el propósito del proyecto. A continuación, se detallan los principios de relevancia, concreción, cobertura, calidad, veracidad, minimización, equidad y transparencia, junto con su significado y aplicación práctica. Dichos principios han sido seleccionados basándonos en las recomendaciones de la Oficina del Dato española, y de Privacy International, que son agencias de referencia a nivel nacional e internacional en este campo.

- **Relevancia**: Solo utilizar datos pertinentes y relacionados con el fin que se persigue. Los datos deben tener un propósito claro dentro del sistema, evitando recopilar o cargar información que no aporte valor a la tarea o, en caso de, por ejemplo, datos legales, que ya no estén en vigor. Por otro lado, si se está construyendo un asistente jurídico, serían relevantes las leyes, sentencias y doctrinas legales, pero no datos personales de un usuario.
- **Concreción (Especificidad)**: Asegurar que los datos se recopilan y usan con un propósito específico y bien definido. Este principio está relacionado con la limitación de la finalidad: los datos deben servir para un objetivo concreto y conocido de antemano, no para usos generales o indeterminados. También implica que la información proporcionada sea precisa y específica, evitando ambigüedades.
- **Cobertura (Integridad y exhaustividad)**: Procurar que los datos cubran suficientemente el espectro de información necesario para el propósito del sistema, sin lagunas importantes que puedan sesgar o limitar los resultados. Cobertura implica cierta completitud: dentro de lo éticamente posible, los datos deben ser lo más completos y representativos posible del dominio en cuestión. Esto evita que la omisión de información relevante lleve al modelo a inferencias incorrectas o a respuestas parciales. En otras palabras, si estamos alimentando al sistema con datos sobre un tema, debemos incluir las distintas facetas o puntos de vista importantes de ese tema. Por ejemplo, para un modelo de visión que clasifica que tiene que detectar personas, hay que asegurarse de ofrecer datos de donde aparezcan personas de diferente etnia, género, condiciones lumínicas, número de personas en la imagen, etc. de manera que se eviten sesgos sistemáticos, aunque este aspecto se analiza en detalle bajo el principio de equidad.

- **Calidad**: Velar por la calidad intrínseca de los datos. Esto abarca varios atributos: precisión, consistencia, integridad, actualidad y pertinencia. En resumen, datos de calidad son datos correctos, bien estructurados, completos y relevantes. Se debe evitar incluir información errónea, duplicada, inconsistente o desactualizada. En el mundo del aprendizaje automático hay un dicho que reza "*Garbage in, Garbage out*" que refleja que la calidad y rendimiento de los sistemas de inteligencia artificial depende de la calidad de los datos fuertemente.
- **Veracidad (Exactitud y veracidad):** Asegurarse de que los datos son verdaderos, confiables y verificables. La veracidad está relacionada con la calidad, pero merece mención especial: implica confirmar que la información corresponde con la realidad y proviene de fuentes fidedignas. Cargar datos falsos, rumores o contenidos no verificados es éticamente inaceptable, pues llevará al modelo a generar respuestas engañosas. Siempre que sea posible, conviene validar la veracidad de los datos contra una fuente independiente antes de incorporarlos. Por ejemplo, si añadimos un artículo informativo al repositorio que usa el modelo, conviene comprobar que los hechos allí descritos estén respaldados por fuentes oficiales o expertos. Recordemos que los datos éticos provienen de fuentes creíbles y no incluyen rumores ni contenido dañino.
- **Minimización**: Aplicar el principio de minimización de datos, muy ligado a la privacidad: significa recopilar y utilizar solo los datos estrictamente necesarios para cumplir el propósito definido, evitando la tentación de almacenar datos "por si acaso". La minimización reduce la exposición innecesaria de las personas y limita los riesgos si ocurre una filtración, además de simplificar el cumplimiento de la normativa. En la práctica, debemos preguntarnos siempre: "¿Necesitamos realmente este dato para lograr nuestra meta?". Si la respuesta es no, es mejor no recogerlo. No toda información disponible debe ser recolectada, incluso si técnicamente es posible; hay que sopesar necesidad vs. Intrusión.
- **Equidad (Justicia y no discriminación):** Garantizar que el uso de los datos sea equitativo y no produzca discriminaciones. La equidad en datos implica que el conjunto de datos no esté sesgado de forma que perjudique sistemáticamente a un grupo o favorezca injustamente a otro. Se busca que los datos representen adecuadamente la diversidad relevante (por ejemplo, datos demográficos equilibrados si corresponde, o casos diversos en un conjunto de datos de entrenamiento) y que no refuercen estereotipos o prejuicios. Los algoritmos pueden parecer imparciales, pero tienden a reflejar los patrones del conjunto de datos con que fueron entrenados. Si los datos históricos contienen sesgos sociales, económicos, de género, etc., el modelo podría reproducir o incluso amplificar esas desigualdades.
- **Transparencia**: Mantener la transparencia en el manejo de los datos y en el funcionamiento del sistema de IA. La transparencia tiene varios niveles: por un lado, significa ser claro con las personas (usuarios, clientes, o cualquier involucrado) sobre qué datos se están recopilando, cómo se utilizan, con qué fin, durante cuánto tiempo y con quién se comparten. Esta información debe comunicarse de forma accesible, comprensible para cualquier persona independientemente de su conocimiento técnico. Por otro lado, también implica que los propios datos tengan procedencia conocida y documentada (saber de dónde provienen, fecha de obtención, etc.), lo cual ayuda a su trazabilidad y a generar confianza en los resultados. En la práctica, ser transparente podría traducirse en acciones como: proporcionar una nota informativa o políticas de privacidad sencillas que expliquen

el uso de datos; documentar las fuentes de los datos (por ejemplo, “estos documentos fueron obtenidos del portal X el día Y”); y alertar sobre cualquier limitación o margen de error del sistema. La transparencia empodera a los usuarios a tomar decisiones informadas sobre su interacción con la IA y refuerza la responsabilidad de quienes desarrollan el sistema.

9.1.2. Prácticas recomendadas en la gestión de datos (selección, organización y actualización)

Además de los principios anteriores, es útil seguir buenas prácticas concretas al seleccionar, organizar y actualizar los datos que alimentarán a un modelo de lenguaje o de visión. Estas prácticas ayudan a mantener la ética y la eficacia del sistema a lo largo del tiempo:

- **Selección de datos apropiados:** El primer paso es elegir cuidadosamente qué datos incorporar. Se deben aplicar los principios de relevancia y calidad: seleccionar datos pertinentes al ámbito del modelo, de fuentes confiables y que cumplan con estándares de calidad (exactitud, integridad, actualidad). Es preferible fuente creíble a cantidad: por ejemplo, mejor usar 100 documentos verificados y de alta calidad que 1000 documentos de procedencia dudosa. De hecho, la UNESCO recomienda recopilar tantos datos de alta calidad como sea posible, provenientes de variedad de fuentes creíbles y éticas, evitando datos basados en rumores o contenido perjudicial. Igualmente, se debe descartar información irrelevante o redundante que solo “haría bulto” en el conjunto de datos. Si se detectan datos desactualizados o erróneos durante la selección, se debe considerar eliminarlos o reemplazarlos por datos más recientes y correctos.
- **Actualización periódica:** Los datos deben mantenerse actualizados con el tiempo. Una práctica recomendable es establecer un calendario o protocolo de revisión. Por ejemplo, decidir que cada 3 meses se revisarán las fuentes para incorporar nueva información (nuevas leyes, estudios recientes, datos del último trimestre, etc.) y para retirar o marcar como obsoletos aquellos datos que ya no sean válidos. La actualización continua es esencial, sobre todo en ámbitos dinámicos, para evitar que el modelo se quede desfasado. También conviene estar atento a eventos que puedan requerir actualizaciones inmediatas: p. ej., si entrenó un modelo con datos hasta 2022 y en 2023 ocurre un hecho muy relevante (como un cambio normativo importante o un nuevo descubrimiento científico), habría que añadir esa información nueva lo antes posible. Al actualizar, se deben seguir los mismos criterios éticos que en la selección inicial: usar fuentes confiables para las novedades y documentar los cambios realizados (qué se añadió, modificó o eliminó y por qué).

9.1.3. Revisión y validación de fuentes de datos

No basta con tener muchos datos: es esencial que provengan de fuentes confiables y válidas. La revisión y validación de las fuentes de datos es una tarea ética clave para asegurar la calidad y fiabilidad del sistema de IA:

- **Verificación de la fuente:** Para cada conjunto de datos o documento que se planee usar, se debe preguntar: "¿De dónde viene esta información? ¿Es una fuente conocida y confiable?". Siempre que sea posible, se debe priorizar fuentes primarias u oficiales. Por ejemplo, cifras estadísticas de un informe del gobierno o un organismo internacional suelen ser más confiables que las de un blog anónimo. Si la fuente es secundaria (alguien que cita a otro), se debe intentar rastrear la referencia original. En el caso de investigaciones académicas, se deben usar publicaciones revisadas por pares en lugar de artículos no revisados. En el caso de noticias, se deben usar medios reconocidos y contrastados. Cabe recordar que la ética de datos exige usar fuentes creíbles y éticas, evitando material no verificado o tendencioso.
- **Corroboración cruzada:** Una buena práctica es validar la información importante en más de una fuente. Si todas sus fuentes independientes coinciden, es más probable que el dato sea correcto. Si se encuentran discrepancias, se debe investigar más. Por ejemplo, si se va a introducir en el sistema la afirmación "El río X tiene Y kilómetros de longitud" y se encuentran dos fuentes con números distintos, se debe averiguar cuál es la más actual o autorizada (quizá la fuente oficial de hidrografía). En especial, para hechos críticos (dosificación de un medicamento, interpretación de una ley, etc.), doble verificación puede evitar errores graves.
- **Actualidad de la fuente**: Se debe comprobar la fecha de las fuentes. Una fuente confiable pero antigua podría no ser válida actualmente. Por ejemplo, un artículo médico de 2010 puede estar desactualizado por investigaciones más recientes. Siempre que sea posible, se debe usar la versión más actualizada de la información (ej.: la edición vigente de una norma legal, la última guía clínica, la base de datos actual de productos). Si por alguna razón se incluyen datos históricos (porque son necesarios para contexto), hay que asegurarse de marcarlos como históricos para que no se confundan con el estado actual.
- **Imparcialidad y sesgo de la fuente:** Se debe evaluar si la fuente pudiera tener algún sesgo o agenda. Por ejemplo, un estudio financiado por una empresa puede tender a resaltar conclusiones favorables a su producto. Esto no invalida automáticamente la fuente, pero conviene tenerlo en cuenta y, de ser posible, complementarla con otras más neutrales. Del mismo modo, una base de datos puede estar incompleta o inclinada hacia cierto grupo (pensemos en encuestas en línea que suelen captar más a ciertos sectores poblacionales). Ser consciente de estos sesgos ayuda a compensar o ajustar el conjunto de datos para que el modelo no herede perspectivas parcializadas.
- **Documentación de las fuentes:** Se debe mantener un registro de qué fuentes alimentaron al sistema. Esto puede ser tan detallado como un listado de referencias bibliográficas, URLs o identificadores de dataset, o incluso un simple documento donde se consigne: "Fuente X (dirección o referencia) utilizada para datos de tipo Y". Esta documentación de fuentes no solo es buena práctica para transparencia, sino que también facilita actualizar o retirar datos si alguna fuente pierde vigencia o se descubre un problema con ella (por ejemplo, si se supo que un determinado conjunto de datos tenía errores, se puede identificar rápidamente y subsanar).
- **Pruebas piloto o de muestreo:** Si es viable, se deben hacer pequeñas pruebas con una muestra de los datos para validar su comportamiento. Por ejemplo, si se integra una nueva fuente, hay que hacer consultas de ejemplo al modelo para ver qué tipo de respuestas genera basadas en esa información, y verificar si son correctas. Esto puede descubrir problemas de calidad o malinterpretaciones a tiempo. En un modelo de visión, si se añaden 1000 imágenes nuevas de entrenamiento, se deben verificar

algunas predicciones en casos conocidos para asegurarse de que no se degradó la precisión.