Cegados por la IA

Anexo C: Análisis de Privacidad y Cumplimiento Normativo (GDPR & AI Act)

Índice

1. Afectación de la IA Act al proyecto

Bajo la EU AI Act, clasificamos "Cegados por la IA" como un Sistema de Alto Riesgo debido a:

- Afecta directamente a los derechos fundamentales de las personas (accesibilidad y autonomía de usuarios con discapacidad).
- Procesa datos biométricos e información visual de terceros, es decir, puede implicar riesgos para la privacidad o la dignidad de las personas grabadas.
- Puede influir en el comportamiento o decisiones del usuario, al ofrecer descripciones que guían su percepción del entorno.

Como sistema de alto riesgo, el proyecto deberá cumplir con las exigencias de la IA Act, entre las cuales destacan:

- **Documentación** técnica y trazabilidad del sistema: descripción del modelo, su funcionamiento, datos utilizados para el entrenamiento y pruebas de sesgos.

- **Supervisión humana efectiva**: el usuario debe conservar siempre el control y poder desactivar o corregir las descripciones, mediante testeos pre- lanzamiento y con evaluación de uso real post-lanzamiento.

- **Evaluación** de conformidad y **registro** en la base europea de sistemas de IA de alto riesgo.

- **Explicabilidad** y **transparencia**: el usuario debe conocer cómo y por qué se generan las descripciones.

2. Aplicación del GDPR y la LOPDGDD

El sistema trata información que pueden considerarse datos personales según el **Reglamento General de Protección de Datos (GDPR)** y la **Ley Orgánica 3/2018 de Protección de Datos Personales y Garantía de Derechos Digitales (LOPDGDD)**.

Sujetos afectados:

- Usuarios de las gafas: personas con discapacidad visual que utilizan el sistema.

- Terceros grabados: personas presentes en el entorno que podrían ser captadas por la cámara.

Tipos de datos tratados:

- Datos biométricos (rostros, rasgos faciales).

- Datos de localización (entorno físico, espacios visitados).

- Datos contextuales (acciones, objetos o personas descritas)

Bases legales para el tratamiento:

- Artículo 6.1.f del GDPR: interés legítimo del responsable del tratamiento en desarrollar una tecnología que mejore la accesibilidad.

- Artículo 9.2.g del GDPR: excepción para el tratamiento de datos sensibles con fines de interés público esencial y accesibilidad para personas con discapacidad.

3. Derechos y control de los usuarios

El usuario mantiene un control total sobre el tratamiento de sus datos y el funcionamiento del dispositivo.

Podrá:

- Activar o desactivar la cámara en cualquier momento.
- Elegir entre distintos niveles de descripción (por ejemplo, omitir rostros o detalles de personas).
- Activar un modo de privacidad reforzada, que elimina cualquier registro o configuración tras su uso.

El sistema informará, al iniciar su uso, de forma transparente y accesible, mediante mensajes de voz o texto, sobre qué datos se procesan, cómo y con qué finalidad.

4. Riesgos de privacidad identificados

1. Grabación involuntaria de terceros: personas que no han dado su consentimiento pueden ser captadas por la cámara.
2. Tratamiento de datos sensibles: rasgos faciales o comportamientos pueden ser procesados de forma inadvertida.
3. Fuga de información si el dispositivo fuera hackeado.
4. Errores de la IA o sesgos en la descripción (por ejemplo, interpretaciones subjetivas o discriminatorias).
5. Dependencia tecnológica: el usuario podría confiar ciegamente en la IA, reduciendo su atención al entorno real.

6. Estrategias de mitigación

- Privacidad por diseño → todos los datos se procesan en local sin almacenarse.
- Mecanismo de detección de zonas sensibles → pausa automática del procesamiento visual.
- Política de retención nula → no se guarda ningún dato personal tras finalizar el uso.
- Pruebas de sesgo e imparcialidad → validación previa con diferentes entornos culturales y demográficos.
- Formación al usuario → programa de adaptación que fomente la complementariedad con el bastón y la escucha activa.

7. Mapa de datos y principio de minimización

Hemos diseñado el sistema para que sea amnésico. A diferencia de las Big Tech, no queremos los datos de la gente.

- **Imagen Raw:** Vive milisegundos en la RAM. Se borra tras la inferencia.
- **Vectores Biométricos:** Se usan para decir "hay una persona", pero no *quién* es. No hay base de datos de caras.
- **Telemetría:** Solo guardamos logs de errores técnicos ("el sistema falló a las 14:00"), nunca el contenido visual.

Tabla de Minimización de Datos (GDPR):

Tipo de dato	Origen	Finalidad	Base legal (GDPR)	Retención	Medida de minimización
Imagen de vídeo	Cámara de las gafas	Descripción del entorno	Art. 9.2.g – accesibilidad Art. 6.1.f - Interés legítimo	0 segundos No se almacena	Procesamiento local + borrado inmediato
Datos de configuración	Usuario	Personalización de uso	Ejecución de contrato (Art 6.1.b)	Mientras dure la sesión	Logs de errores, datos no visuales, cifrados localmente
Telemetría técnica	Sistema	Mejora de rendimiento	Interés legítimo (Art. 6.1.f)	2 semanas	Sin datos personales identificables

Explicación de las Bases legales:

Imagen de vídeo del entorno (art. 6 GDPR):

- Interés legítimo (art. 6.1.f) → si el tratamiento es necesario para el funcionamiento del sistema y no vulnera derechos de terceros.
- Tratamiento de datos especiales (datos biométricos)

Datos de configuración:

- Ejecución de contrato (art. 6.1.b) → necesarios para prestar el servicio.

Telemetría técnica:

- Interés legítimo (art. 6.1.f) → mantenimiento, seguridad y mejora del sistema.

8. Clasificación según IA Act y cumplimiento GDPR/LOPDGDD

Aspecto	Evaluación / Clasificación
Nivel de riesgo IA Act	Alto riesgo (anexo III – accesibilidad y datos biométricos)
Requisitos IA Act	Evaluación técnica, registro en base europea
Base jurídica GDPR	Interés legítimo (art. 6.1.f) + finalidad de accesibilidad (art. 9.2.g)
Principios aplicables	Minimización, limitación de finalidad, privacidad por diseño
Derechos de los interesados	Acceso, rectificación, oposición, eliminación
Responsable del tratamiento	Empresa desarrolladora del sistema
Medidas técnicas	Procesamiento local, cifrado, borrado inmediato
Supervisión humana	Usuario mantiene control total sobre el dispositivo

9. Retención y borrado (política y procedimiento de purga)

El objetivo es garantizar que todos los datos personales captados por las gafas inteligentes de CegadosPorLaIA sean tratados de forma temporal, segura y

respetando la privacidad de usuarios y terceros, cumpliendo con GDPR y LOPDGDD.

Política de retención de datos:

- Procesamiento local y temporal:
    - Todas las imágenes captadas se procesan exclusivamente en el dispositivo, sin almacenamiento permanente.
    - No se transfieren datos personales a servidores externos ni a terceros.
- Datos de configuración y telemetría:
    - La información de configuración del usuario se mantiene solo durante la sesión activa y se elimina al finalizar el uso.
    - La telemetría técnica (información no personal sobre funcionamiento del sistema) se retiene un máximo de 2 semanas, únicamente para mejorar rendimiento y detectar fallos, sin contener información personal identificable.

Para garantizar el cumplimiento, implementamos tres niveles de borrado:

1. **Ciclo de Inferencia:** Sobrescritura de buffers frame a frame.
2. **Fin de Sesión:** Script de limpieza de caché al apagar las gafas.
3. **Botón del Pánico:** Modo de "Privacidad Reforzada" que corta alimentación de sensores y purga la memoria RAM instantáneamente.

10. Sesgos, equidad y colectivos afectados

1. Análisis de Usuarios y Dinámicas de Poder

El despliegue de unas gafas inteligentes de asistencia visual genera una reconfiguración de las relaciones de poder en el espacio físico:

- **Usuario Directo (Sujeto asistido):** Personas con discapacidad visual.
    - *Relación de poder:* Existe una **dependencia epistémica** crítica. El usuario delega en la IA la interpretación de la realidad. Si el

sistema alucina o sesga la descripción, el usuario no tiene (en ese instante) un canal visual alternativo para verificar la verdad, lo que le coloca en una situación de alta vulnerabilidad ante fallos del sistema.

- **Terceros No Usuarios (Sujeto pasivo):** Ciudadanía general.
    - *Relación de poder:* Asimetría de información y consentimiento. El tercero es analizado por una cámara sin haber dado su consentimiento explícito, desconociendo qué está "diciendo" la IA sobre él al usuario (ej. ¿está describiendo mi ropa? ¿mi actividad?).
- **La Organización (Sujeto controlador):** Nosotros como desarrolladores.
    - *Responsabilidad:* Definimos la ontología del mundo (qué objetos merecen ser nombrados y cómo). Esto conlleva el poder de visibilizar u ocultar realidades sociales a través de las etiquetas del modelo.

2. Mapa Ampliado de Colectivos Vulnerables

Además de los mencionados en la Model Card, se han analizado factores adicionales de vulnerabilidad que requieren atención específica:

A. Factores Socioeconómicos y Demográficos

1. **Renta y Clase Social (Acceso):** El hardware de alto rendimiento (gafas con chips neuromórficos) tiende a ser costoso. Existe el riesgo de crear una **"autonomía de dos velocidades"**, donde solo las personas ciegas con recursos pueden acceder a una navegación segura, mientras que las de rentas bajas dependen de ayudas tradicionales o tecnología obsoleta.
2. **Raza y Etnia (Visibilidad Algorítmica):** Históricamente, los modelos de visión por computador (entrenados en ImageNet o COCO) tienen tasas de error más altas ("Falsos Negativos") en personas de piel oscura, especialmente en condiciones de baja luz. El riesgo es que el sistema no detecte a un peatón racializado, provocando un choque o accidente.
3. **Edad (Brecha Digital y Cognitiva):** Personas mayores con pérdida de visión adquirida (DMAE, glaucoma). Este colectivo puede tener dificultades para interactuar con interfaces de voz rápidas o gestos táctiles complejos en la patilla de la gafa, sintiendo frustración o rechazo al sistema.

B. Factores del Entorno y Salud

4. **Entorno Geográfico (Rural vs. Urbano):** El sesgo de "mundo urbano". Los modelos suelen entrenarse con calles asfaltadas y señalética estándar. Un usuario en un entorno rural (caminos de tierra, ausencia de aceras, ganado suelto) podría recibir descripciones erróneas o nulas, dejándole desprotegido fuera de la ciudad.
5. **Discapacidad Auditiva o del Habla (Co-morbilidad):** Usuarios sordociegos o con dificultades en el habla. Si el sistema depende exclusivamente de *inputs* de voz o *outputs* de audio, excluye a quienes no pueden oír la descripción o dar comandos vocales.

C. Interseccionalidad (El efecto multiplicador)

La vulnerabilidad no es aditiva, es multiplicativa.

- *Ejemplo Crítico:* **Mujer mayor + Entorno Rural + Piel oscura.** En este caso, se suman: la menor precisión técnica del modelo para detectar su presencia (sesgo racial), la incapacidad del modelo para entender su entorno (sesgo geográfico) y la dificultad de la usuaria para corregir al sistema o gestionar fallos técnicos (brecha generacional). El sistema debe ser robusto específicamente en estos "casos de borde" demográficos.

3. Fuentes de Sesgo y Estrategias de Mitigación

Hemos auditado el ciclo de vida del producto para localizar el origen de los sesgos:

A. En los Datos (Origen)

- **Problema:** Los datasets *open source* masivos suelen estar occidentalizados y sobrerrepresentan a personas de piel clara y entornos de clase media.
- **Acción Correctiva:**
    - Enriquecimiento del dataset de validación (*Golden Set*) con imágenes de entornos rurales y diversidad étnica explícita.

    - Uso de **Datos Sintéticos** para generar escenarios de peligro (ej. obras en la calle) protagonizados por avatares de diversas etnias y géneros para equilibrar el entrenamiento.

B. En el Modelo (Procesamiento)

- **Problema:** Las funciones de pérdida suelen optimizar para la "mayoría". Un error del 1% es aceptable globalmente, pero inaceptable si ese 1% se concentra todo en una minoría étnica.
- **Acción Correctiva:**
    - Ajuste del **System Prompt** del LLM para forzar la neutralidad descriptiva y bloquear adjetivos subjetivos (*v.gr.* prohibir palabras como "pobre", "rico", "guapo", "feo", "sospechoso").
    - Penalización específica durante el *fine-tuning* para alucinaciones sobre atributos sensibles.

C. En el Producto (Interacción)

- **Problema:** Interfaz diseñada para un "usuario estándar" joven y tecnológico.
- **Acción Correctiva:**
    - **Subvenciones:** Acuerdos estratégicos con la ONCE y la Administración para que el precio no sea una barrera discriminatoria.
    - **Modo "Lectura Fácil":** Configuración simplificada que reduce la velocidad del habla y simplifica el vocabulario para usuarios mayores o con dificultades cognitivas.

4. Métricas de Equidad y Criterios de Validación

Para pasar a producción, el sistema debe superar umbrales estrictos de equidad matemática ("Fairness Constraints"):

1. **Paridad de Recall (Sensibilidad):**
    - La diferencia en la tasa de detección (*True Positive Rate*) de personas entre el grupo demográfico mejor detectado y el peor detectado (ej. hombre blanco vs. mujer negra) **no debe superar el 5%**.
2. **Tasa de Falsos Negativos en Seguridad:**
    - Se exige una tasa de **0%** de diferencia estadística significativa en la detección de peligros (coches, huecos) entre entornos rurales y urbanos en el dataset de prueba.
3. **Evaluación Contrafactual (Counterfactual Testing):**
    - Se realizan pruebas modificando digitalmente el tono de piel o género de una persona en la misma escena. La descripción generada por la IA debe permanecer semánticamente idéntica en cuanto a la acción descrita (ej. "persona esperando el bus" no debe cambiar a "persona merodeando" al cambiar el tono de piel).
4. **Validación Cruzada Cultural:**
    - Pruebas de usabilidad obligatorias con usuarios beta de al menos 3 perfiles demográficos distintos antes de cada *major release*.

11. Referencias

- **Documentos de los sprints:**
    - *Análisis de Privacidad y Cumplimiento Normativo.docx*
    - *Entrega - Sprint 1.docx* (Apartado de Privacidad)
- **Referencias externas (Normativa):**
    - EU AI Act (Reglamento Europeo de IA): https://artificialintelligenceact.eu/es/
    - Reglamento General de Protección de Datos (GDPR/RGPD): https://gdpr-text.com/es/
    - Ley de Propiedad Intelectual (BOE): https://www.boe.es/buscar/act.php?id=BOE-A-1996-8930

- **Uso de IA**
    - Apartado 1.2: Bases legales
    - Apartado 2: Bases legales
    - Sesgos, equidad y colectivos afectados: https://gemini.google.com/share/1c938eb66645