1. Sobre la identificabilidad de los “esqueletos 3D” (art. 4.1 RGPD)

Los esqueletos digitales que genera **PrisonRisk-AI no** se consideran datos personales bajo el artículo 4.1 del RGPD, pues **no permiten identificar** directa ni indirectamente a una persona física.

Estas representaciones sólo contienen vectores espaciales y articulaciones (coordenadas abstractas X-Y-Z y trayectorias cinéticas). No incluyen rasgos biométricos, faciales ni texturas del cuerpo.

Adicionalmente:

- En la misma cadena de procesamiento se elimina el vínculo entre el vídeo original y el esqueleto digital (técnica conocida como *anonymization at source*).

- Ningún operario —ni el modelo— puede reconstruir la identidad a partir del esqueleto, pues las claves de correspondencia se eliminan inmediatamente tras la transformación.

- Se aplicarán auditorías técnicas de **reidentification risk** (riesgo de reidentificación) conforme a las guías del EDPB, en línea con las opiniones 05/2014 y 03/2019 del Grupo de Trabajo sobre Protección de Datos.

Si en algún momento se demostrase que existe posibilidad de reidentificar a una persona (por error o por avances tecnológicos), entonces los esqueletos serían tratados como **seudonimizados**, activándose **salvaguardas adicionales**: incluirían EIPD (evaluación de impacto), cifrado individual por persona, y restricciones de acceso.

2. Riesgo de exceso de confianza del personal (automation bias)

Se reconoce que usar sistemas de apoyo puede generar **sesgo de automatización** (que los usuarios confíen demasiado en la máquina). Para reducirlo, PrisonRisk-AI incorpora distintas medidas:

- Formación obligatoria anual para el personal, sobre uso responsable de sistemas de IA.

- Panel explicativo que obliga al funcionario a revisar las razones de cada alerta antes de actuar.

- Sistema de **doble confirmación**: cualquier acción basada en una alerta debe quedar registrada con la identidad del funcionario.

- Alertas de uso excesivo: el sistema detecta patrones donde se usa demasiado la herramienta (por ejemplo, que muchos casos se acepten sin revisión humana) y

alerta a supervisores.

Con esto se busca que la **supervisión humana** (requisito del artículo 14 del proyecto de ley europea de IA) sea efectiva, no meramente simbólica.

3. Supervisión y gobernanza

El sistema debe estar supervisado en tres niveles:

1. **Operativo**: el funcionario penitenciario de turno recibe alertas y decide acciones.

2. **Institucional**: el jefe de servicio o director del centro valida su uso correcto y revisa informes periódicos.

3. **Externo**: el Delegado de Protección de Datos (DPO) y un Comité Ético-Técnico auditan el desempeño, posibles sesgos y el cumplimiento con el RGPD y la ley de IA.

Ese comité debe incluir expertos en derecho, ética de IA, psicología penitenciaria y representantes de derechos humanos.

4. Responsabilidad frente a fallos del modelo

La responsabilidad se repartirá conforme a la futura **IA Act** (arts. 9-14) y la normativa española de función pública:

- **Proveedor (desarrollador de PrisonRisk-AI)**: responde si el fallo surge por un error técnico, deficiencia en el entrenamiento o incumplimiento de estándares.

- **Operador (Administración Penitenciaria)**: si el sistema se usa fuera del marco de los procedimientos definidos.

- **Funcionario**: sólo será responsable si se prueba negligencia grave o que actuó en contra del protocolo.

Se mantendrá un registro de auditoría con todas las alertas y decisiones humanas, para reconstruir la cadena de responsabilidades si es necesario.

5. Ansiedad y entorno de vigilancia excesiva (art. 25.2 CE)

Se reconoce el riesgo psicológico de sentirse constantemente observado. Para mitigarlo:

- El sistema **no aumentará** el número de cámaras; en cambio, optimiza las ya existentes, reduciendo vigilancia humana continua.

- Las alertas solo se activan en situaciones de riesgo real, evitando una percepción de vigilancia permanente.

- Antes del despliegue final se elaborará un **informe de impacto psicosocial**, con psicólogos penitenciarios.

- La finalidad buscada es proteger la integridad física de internos y funcionarios, no instaurar control adicional, en conformidad con el principio constitucional de reeducación y reinserción (art. 25.2 CE).

6. Derechos de los presos

Los internos seguirán gozando de sus derechos según el RGPD y la legislación penitenciaria:

- **Derecho a información**: se informará mediante carteles y folletos sobre el uso del sistema de IA, su finalidad, el responsable y los medios para contactar con el DPO.

- **Derecho de acceso y reclamación**: pueden solicitar información (sobre los datos tratados, aunque ya anonimizados) y presentar reclamaciones ante la AEPD a través del DPO.

- **Derecho a revisión humana**: cualquier decisión con efectos reales sobre el régimen del interno debe ser revisada por un funcionario; la IA no puede decidir de forma autónoma.

7. Identificación sin perfiles personales

El sistema **no "castiga" individuos**: crea alertas contextuales (por ejemplo, riesgo en una zona) y el funcionario evalúa in situ. Solo en ese momento se puede vincular la alerta a una persona mediante observación humana, no mediante perfilado automático histórico.

De ese modo no es necesario conservar perfiles históricos de peligrosidad, lo que respeta los principios de proporcionalidad y minimización del artículo 5 del RGPD.

8. Persistencia de sesgos y medidas adicionales

Eliminar variables raciales no basta para eliminar sesgos. Por eso PrisonRisk-AI adopta estrategias adicionales:

- Evaluaciones continuas de equidad, usando métricas como **demographic parity**, **equalized odds** y *false positive rate balance* sobre grupos de prueba.

- Uso de datos sintéticos balanceados para neutralizar correlaciones espurias.

- Auditorías externas independientes anuales para detectar sesgos latentes.

- Un comité ético-técnico permanente con potestad para detener el modelo si se encuentra discriminación indirecta.

En resumen, la mitigación de sesgos es un proceso continuo, no algo que se logra una sola vez.

Referencias

Agencia Española de Protección de Datos (AEPD). (2022). *Guía sobre adecuación del RGPD a tratamientos con inteligencia artificial*. Madrid: AEPD.
Disponible en: https://www.aepd.es/guias/adecuacion-rgpd-ia.pdf

Agencia Española de Protección de Datos (AEPD). (2014). *Dictamen 05/2014 sobre técnicas de anonimización* (Grupo de Trabajo del Artículo 29).
Disponible en: https://www.aepd.es/documento/wp216-es.pdf

European Data Protection Board (EDPB). (2021). *Directrices 4/2019 sobre protección de datos por diseño y por defecto (versión 2.0)*.
Disponible en:

https://www.edpb.europa.eu/system/files/2021-04/edpb_guidelines_201904_dataprotection_by_design_and_by_default_v2.0_es.pdf

Agencia Española de Protección de Datos (AEPD). (2023). *Evaluación de la intervención humana en las decisiones automatizadas*. Blog de la AEPD.
Disponible en:

https://www.aepd.es/prensa-y-comunicacion/blog/evaluacion-de-la-intervencion-humana-en-las-decisiones-automatizadas

Microlab Hard. (2021). *El derecho a no ser objeto de decisiones automatizadas sin intervención humana según el RGPD*.
Disponible en:

https://microlabhard.es/el-derecho-a-no-ser-objeto-de-decisiones-automatizadas-sin-intervencion-humana-segun-el-rgpd/

RGPD - https://europa.eu/youreurope/business/dealing-with-customers/data-protection/data-protection-gdpr/index_es.htm

LOPD-RGPD - https://protecciondatos-lopd.com/empresas/cumplimiento-lopd/

IA ACT - https://maldita.es/malditatecnologia/20240916/ley-inteligencia-artificial-union-europea/

https://artificialintelligenceact.eu/es/ai-act-explorer/