# **ESPECIFICACIÓN V2** | GRUPO PRISION RISK-AI

Ruta Github Prision Risk-AI con especificación de los puntos tocados en el debate:
sprint2/Ampliacion_Debate_Sprint1.pdf

### Riesgos y límites

| Riesgo | Descripción | Gravedad | Prob. | Prevención | Contingencia |
|---|---|---|---|---|---|
| Sesgo algorítmico | Asociaciones accidentales de datos | Alta | Media | Autorías de equidad; balanceo de datos | Reentrenamiento e informe |
| Falsos positivos | Alertas erróneas sobre internos | Alta | Media | Calibración y supervisión humana | Anulación registro y revisión ética |
| Falsos negativos | No detectar conflictos reales | Alta | Media | Modelos redundantes | Vigilancia reforzada temporalmente y revisión del algoritmo |
| Fuga de datos | Acceso no autorizado | Alta | Baja | Cifrado AES-256; control de roles | Notificación DPO/AEPD y bloqueo |
| Manipulación | Evasión intencional del sistema por un preso | Media | Media | Multimodalidad y detección de anomalías | Reentrenamiento y alerta interna |
| “Lock-In” tecnológico | Dependencia de proveedor | Media | Media | Software libre y APIs abiertas | Migración a alternativa libre |
| Fallo crítico | Error en sensores / software | Alta | Media | Redundancia y monitorización | Modo manual y revisión técnica |
| Riesgo psicosocial | Ansiedad por vigilancia constante | Media | Media | Comunicación y evaluación psicosocial | Ajustes operativos o suspensión |

- **Plausibilidad (2029)**
  Las tecnologías (esqueletos, XAI, multimodalidad) serán maduras para 2029. Persisten riesgos en reidentificación y ataques adversariales, mitigables con supervisión continua y evaluaciones EIPD iterativas.

- **Líneas rojas**
  ❌ No se tomarán decisiones automáticas sobre sanciones, o clasificación.
  ❌ No se almacenarán datos biométricos identificables.
  ❌ No se usarán los resultados del sistema como prueba judicial o disciplinaria directa.
  ❌ No se comercializarán ni transferirán los datos fuera del sistema penitenciario español.

- **Emergencias y paradas**
  ➢ Solo el Director o el DPO técnico pueden detener el sistema.
  ➢ Se activa “modo seguro” ante brechas o falsos positivos >20%.
  ➢ Se bloquea el sistema, la base de datos y se notifica al Ministerio del Interior.

- **Rendición de cuentas**

➢ Canal digital de reclamaciones con respuesta en 15 días hábiles.
➢ Revisión por comité mixto (IA, ética, supervisión penitenciaria).
➢ Medidas: eliminación de registros erróneos, revisión de modelo y compensación administrativa.
➢ Responsables: Ministerio del Interior y proveedor IA (corresponsables, art. 26 RGPD).

Ampliación en Github: sprint2/Ampliacion_Riesgos_y_Limites.pdf

## Propiedad Intelectual

**Licencias por componente**

- Código fuente: Licencia GNU GPL v3, que garantiza acceso al código y obliga a compartir mejoras, favoreciendo la auditoría ética. Se descartan MIT (demasiado permisiva), Apache 2.0 (riesgo de apropiación comercial) y LGPL (copyleft débil). El código operativo en los centros no se publica por motivos de seguridad.
- Modelo de IA: Distribuido con pesos abiertos (open weights) bajo cláusulas éticas (uso no militar ni comercial). Los modelos entrenados localmente quedan restringidos a uso institucional.
- Datos: Datasets generados internamente y protegidos por copyright; uso permitido solo en el marco del proyecto.
- Documentación: Licencia Creative Commons BY 4.0, para permitir reutilización con atribución.

**Modelo de negocio basado en software libre**

Se adopta un modelo “Open Core ético”, combinando núcleo libre (GPL) y servicios profesionales (auditorías, soporte, personalización) como fuente de ingresos sostenibles. Este enfoque equilibra transparencia y sostenibilidad económica, evitando dependencia tecnológica (lock-in).
Modelos descartados:

- Donaciones / mecenazgo, por inestabilidad financiera.
- Publicidad o monetización de datos, incompatible con la ética penitenciaria y el RGPD.
- Licenciamiento dual, por riesgo de fragmentación entre versiones abiertas y propietarias.

**Marcas y Naming**

El nombre “PrisonRisk-AI” ha sido elegido por su claridad y asociación directa con la finalidad del sistema: evaluación de riesgos en entornos penitenciarios mediante IA.
Se realizó una búsqueda de colisiones en la OEPM, EUIPO, Google y GitHub, sin resultados conflictivos en el ámbito tecnológico o penitenciario.
El logo se basará en una identidad visual sobria (tonos azules y grises, tipografía sans-serif), transmitiendo confianza y transparencia institucional.
Guía de uso: el nombre y logotipo sólo podrán emplearse con fines informativos o institucionales, prohibiéndose su uso comercial o en contextos no autorizados.

**Plan de Entrenamiento sin Violación de Copyright**

Los datasets se generan internamente mediante:
Simulaciones sintéticas de interacciones penitenciarias y comportamientos de riesgo.
Datos anonimizados procedentes de cárceles piloto, bajo acuerdo con el Ministerio del Interior.
Datos públicos de estudios de comportamiento, siempre con licencias abiertas.
Estrategia legal y técnica:
Todos los datos externos se emplean con licencias de uso explícitas.
Se aplican técnicas de data augmentation y generación sintética para evitar dependencia de material protegido.
Cumplimiento garantizado del RGPD y el AI Act en todo el ciclo de entrenamiento.

Ampliación en Github: sprint2/Ampliacion_Propiedad_Intelectual.pdf

Las referencias de cada apartado están puestas en sus respectivas ampliaciones.