ANEXO 2.1. RIESGOS Y LÍMITES DEL PROYECTO

GPS-Safe se clasifica como un sistema de IA de alto riesgo según el IA Act, debido a su impacto directo en la seguridad física y el uso de datos personales y de contexto (ubicación, salud, movilidad).

1. Tabla de riesgos (mínimo 8 riesgos)

Se analizan los 8 riesgos más evidentes a los que se enfrenta GPS-Safe, especificando para cada uno de ellos: nombre; descripción; escala de gravedad de sus consecuencias si sucede; escala de probabilidad de que suceda; acciones de prevención (para que no suceda); acciones de contingencia (qué se hará si sucede).

Riesgo	Grav.	Prob.	Detalles
1. Sesgo de Datos Históricos (Técnico/Ético)	ALTA	MED	**Descripción**: El modelo de predicción perpetúa sesgos históricos de criminalidad, sobre-criminalizando ciertas zonas o grupos (similar a los problemas del algoritmo de Chicago). Se puede acusar a la empresa de "discriminación". **Prevención**: Uso de fuentes mixtas (datos oficiales y reportes). Aplicación de técnicas de normalización contextual, revisión humana experta y mecanismos de aleatoriedad controlada que reduzcan la dependencia de patrones históricos sesgados. **Contingencia**: Auditoría del Comité Ético. Ajuste manual de ponderaciones para corregir sesgos detectados durante el entrenamiento periódico.

Riesgo	Grav.	Prob.	Detalles
2. Falsos reportes (Técnico/Social)	MED	MED	**Descripción**: Usuarios pueden realizar reportes falsos, modificando las predicciones del modelo y evitando que las personas pasen por ciertas rutas. **Prevención**: Realizar un control de falsos reportes. Tener una variable de credibilidad de un reporte, por ejemplo, si hay varias personas que lo reportaron o tener en cuenta la credibilidad del usuario, si ha tenido un historial de reportes reales. **Contingencia**: Banear a usuarios que realizan demasiados reportes en un instante de tiempo y eliminar sus reportes falsos. Contrastar la información de los reportes de la app con la información de canales oficiales
3. Apagón del sistema (Seguridad)	ALTA	BAJA	**Descripción**: Los hackers pueden provocar un ataque de denegación de servicio, de modo que todos los usuarios de la aplicación no la pueden usar (“paraíso” para los delincuentes). **Prevención**: Cachear localmente las últimas rutas por las que has pasado, para reutilizarlas si se pierde conexión. Realizar el desarrollo del sistema siguiendo estándares de ciberseguridad. **Contingencia**: Intentar restablecer el funcionamiento del sistema en cuanto se pueda. Notificar por canales oficiales las imágenes estáticas de los mapas para que la gente pueda usar al menos algo para orientarse.

Riesgo	Grav.	Prob.	Detalles
4. Uso por Delincuentes (Seguridad)	ALTA	MED	**Descripción**: Los delincuentes utilizan la app para identificar zonas con menor presencia policial/ciudadana o patrones de movimiento de usuarios. **Prevención**: Los criterios de seguridad incluyen alta presencia ciudadana/policial, donde es poco probable que actúen. Inclusión de avisos preventivos. **Contingencia**: Proporcionar datos de patrones a las fuerzas de seguridad para asistencia y prevención del delito.
5. Fuga de datos personales (Privacidad/Legal)	ALTA	BAJA	**Descripción**: Los hackers entran en el sistema y consiguen acceso a los datos personales (rutas, direcciones de domicilio, reportes, edad, nombre), publicando dicha información. **Prevención**: Anonimizar los datos al máximo, de modo que la verificación de identidad sólo se realiza al principio. Realizar el desarrollo del sistema siguiendo estándares de ciberseguridad. **Contingencia**: No hay muchas acciones posibles, los datos ya están en la red. Lo único que se puede hacer es intentar bloquear el acceso a los datos en caso de intruso para evitar la fuga de alguna cantidad de datos y mejora de seguridad posterior del sistema.
6. Exceso de confianza del usuario en la app (Funcional/Social)	MED	ALTA	**Descripción**: El usuario confía ciegamente en la app y no toma precauciones personales al desplazarse. **Prevención**: Mostrar recordatorios y advertencias sobre responsabilidad personal. **Contingencia**: Enviar alertas o mensajes educativos en caso de incidentes.

Riesgo	Grav.	Prob.	Detalles
7. Sesgo Geográfico/Discriminación Económica (Social/Ético)	MED	ALTA	**Descripción:** La aplicación desvía constantemente el tráfico, perjudicando económicamente a comercios en calles percibidas como menos seguras. **Prevención:** Aplicación de mecanismos de equilibrio y corrección contextual, incluyendo un margen de aleatoriedad controlada o normalización local. Se prioriza la seguridad del usuario, pero se busca mínima desviación. **Contingencia:** El usuario puede excluir calles o zonas concretas. Dialogar con ayuntamientos para mejorar la seguridad en zonas sesgadas.
8. Responsabilidad Legal por Predicción Errónea (Legal)	ALTA	BAJA	**Descripción:** Un error en la recomendación de ruta provoca un accidente o incidente grave. **Prevención:** Transparencia y trazabilidad del sistema. Preferencia por IA explicable (XAI). Supervisión y pruebas continuas de la IA. **Contingencia:** El proyecto asume responsabilidad compartida (la decisión final recae en el usuario). Se incluye un botón para reportar errores de inferencia para aprendizaje inmediato.

2. Plausibilidad: partes frágiles o dudosas

- Entrenamiento de modelos predictivos: Requiere uno o varios modelos capaces de recibir tantas variables muy diferentes y proporcionar una precisión bastante alta (accuracy > 95%), evitando los falsos negativos (i.e. decir que una zona peligrosa es segura). Eso puede ser frágil, ya que la disponibilidad de datos de calidad y la capacidad real de generalización del modelo son inciertas (en cuanto a Horizonte 2029).

- Capacidad computacional y latencia: La interacción en tiempo real, como pasa con Google Maps, depende de servidores con alta potencia y baja latencia. Esto puede ser técnicamente viable, pero costoso y difícil de sostener a gran escala.

- Cobertura y conectividad: GPS-Safe necesitaría acceso al servidor en todo momento, por si hay que actualizar la ruta dinámicamente. En zonas con mala cobertura o limitaciones de red, la funcionalidad podría verse comprometida.

- Competencia con grandes plataformas: Si Google o Apple Maps incorporan esta funcionalidad en sus aplicaciones, el modelo de negocio GPS-Safe podría volverse inviable debido a su posición dominante y a los recursos que poseen, lo que representa un riesgo estratégico externo.

3. Líneas rojas (prohibiciones estrictas)

- En cuanto al uso de datos:
    - No guardar datos personales de forma explícita, usar técnicas de anonimizado de datos (incluso trabajadores de la empresa no pueden usar los datos para saber información sobre los usuarios).
    - No usar datos para fines distintos a la seguridad o a la navegación.
    - No usar la localización del usuario cuándo no está usando la app.
- No vender información de zonas que visita el usuario.
- No usar una única fuente de información.
- No desplegar un modelo de caja negra o modelo sesgado.
- No permitir el uso de la app a menores de 14 años.

4. Emergencias (paradas de emergencia)

El sistema de GPS-Safe debe garantizar la capacidad de interrupción o intervención humana.

Quién puede parar el sistema:

- El Product Owner (Pablo M. Rodríguez Sosa), asistido por el Comité Ético y el equipo de Ingeniería, es el responsable ejecutivo de solicitar una parada de emergencia (total o parcial).

Cuándo (Condiciones Medibles):

- Fallo grave y sostenido del sistema de predicción que derive en la recomendación de rutas con riesgo conocido o excesivamente alto.
- Pérdida de control o evidencia de manipulación masiva (p. ej., ataque de bots exitoso) que comprometa la fiabilidad del Modelo de predicción de riesgo urbano.
- Detección de una violación crítica del RGPD o una brecha de seguridad que afecte a los datos sensibles de los usuarios.

Cómo (Procedimiento Breve):

- El Product Owner, previa consulta con el Scrum Master (Carlota) y el Comité Ético, ordena la desconexión del motor de IA central o la suspensión temporal de la

funcionalidad de recomendación de rutas seguras, dejando activa únicamente la función de mapa básico (similar a un mapa sin IA).

- ❖ La aplicación mostrará un aviso claro sobre la suspensión y la causa.

5. Rendición de cuentas

- **Canal de reclamación**: feedback dentro de la app y botón para reportar errores de inferencia, canal de soporte con teléfono, correo electrónico y foros en la página web. Respuesta a reclamaciones formales: 15 días hábiles.
- **Revisión humana garantizada**: se podría usar tanto para comprobar los reportes dudosos como para comprobar la veracidad de las predicciones, pero es más fácil y escalable que los usuarios directamente reporten al finalizar la ruta (poniendo el número de estrellas o algo similar). El Comité Ético se encarga de la auditoría continua de riesgos.
- **Formas de reparación**:
    1. Corrección y Prevención: Uso del botón de reporte de errores para que la aplicación aprenda de los fallos detectados. Revisión y pruebas continuas del sistema por personas para resolver incidencias y asegurar que las rutas sean seguras. El usuario puede modificar la información que el sistema ha aprendido sobre él.
    2. Reversión: Opción de borrar completamente todos los datos personales del usuario sin demoras.
    3. Compensación: En el marco de la responsabilidad compartida, la compensación se aplicaría solo si se demuestra que el daño se derivó de una falta de transparencia, trazabilidad, o un fallo no auditado del sistema, según las leyes de responsabilidad civil.

Referencias

- Conversación con IA para la depuración y concisión de texto (siempre con supervisión humana): Conversación con Chat GPT