Especificación tras el debate

El código tiene licencia GPL, pero el modelo de IA es propietario, ¿cómo lo vais a hacer para incluir el modelo en el código?

El modelo es propietario para diferenciarnos de la competencia. El código incluirá únicamente la lógica necesaria para interactuar con dicho modelo, sin exponer su arquitectura interna ni parámetros. Cualquier desarrollador podrá integrar y probar esta funcionalidad a través de nuestra API de desarrollo (con límites para el número de peticiones para evitar usos con fines comerciales), que proporcionará endpoints seguros y documentados para el envío de solicitudes y la obtención de respuestas del modelo.

Decís que la app tendrá un accuracy de un 95%, demasiado alto. ¿Cómo obtendrán un accuracy tan alto, alegando que reentrenan el modelo continuamente?

Tras analizar detenidamente Google Scholar, hemos encontrado bastantes papers que promueven esta tecnología, mayoritariamente en India. El siguiente paper: https://aclanthology.org/N15-2003.pdf - no es igual a nuestro proyecto, pero bastante similar.

Intentan predecir rutas seguras usando noticias en la ciudad de Delhi, India. Consiguen un accuracy de 83.64 % a la hora de localizar correctamente los crímenes a partir de las noticias. Utilizan análisis de discriminante (LDA) y named entity recognition (NER).

Podemos estimar que al usar modelos de estado del arte podremos conseguir el 95 % de accuracy, siendo este uno de los riesgos con mayor peso del proyecto.

Si hay varias rutas óptimas, la decisión final será en base a un algoritmo aleatorio. ¿Pero eso va a reflejar la realidad? ¿Haréis alguna auditoría de equidad algorítmica?

Somos conscientes de que no podemos favorecer una ruta frente a otra sin una justificación sólida. Por ello consideramos que introducir aleatoriedad en esta fase fomenta la "diversidad" y ayuda a evitar sesgos indebidos.

Antes de implementar la nueva versión del modelo en producción, el comité ético verificará la posible presencia de estos sesgos, que, en principio, no deberían producirse dado el diseño de la arquitectura algorítmica.

No dais información de la composición del Comité Ético (hay externxs?), de si rotará a futuro, de su poder y competencias más allá de menciones puntuales...

El Comité Ético (CE) es la máxima autoridad de gobernanza de GPS-Safe, un sistema de IA de Alto Riesgo (por seguridad física y datos sensibles) bajo el AI Act.

- Función principal: Garantizar la ética, la rendición de cuentas, la mitigación de riesgos y la auditoría continua de sesgos en el sistema.
- Composición: incluye al Product Owner (Pablo), un Representante Sénior de Ingeniería, y un Experto Externo en Ética/Legal. El Scrum Master (Carlota) actúa como consultor.

- Rotación: Los roles internos son continuos. El Experto Externo rota bienalmente para asegurar la objetividad y la actualización de criterios éticos.
- Poderes clave:
    1. Aprobación Ética: Aprueba cambios significativos en el pipeline de datos y modelos algorítmicos, incluido el despliegue de un nuevo modelo a producción.
    2. Intervención Humana: Posee la autoridad para detener de emergencia el sistema de recomendación de rutas seguras ante fallos graves o manipulación, dejando solo un mapa básico operativo.
    3. Conformidad: Garantiza la plena adhesión a los requisitos del AI Act.