



Tecnológico de Monterrey

Campus Querétaro

Privacidad y Seguridad de los Datos

Luis Ángel Cruz García

A01736345

Inteligencia artificial avanzada para la ciencia de datos II
Grupo 501

Introducción

A continuación se describe el análisis realizado de manera individual respecto a la privacidad y seguridad de los datos en el reto, para así poder aportar información y lograr un acuerdo en el equipo sustentado con diferentes puntos de vista.

Primero es importante mencionar el contexto de la problemática del Socio Formador, el cual es el CAETEC, que al pertenecer al Tecnológico de Monterrey permite una mayor flexibilidad al momento de compartir y permitir utilizar los datos, otro de los puntos importantes es que estos datos constan de fotografías donde principalmente se muestran camas de vacas vistas desde un plano cenital.

Al momento de etiquetar las imágenes, se encontraron un número alto de vacas, sin embargo, en algunas se llegaron a localizar objetos que forman parte del entorno, como llantas que son utilizadas en dichas camas o en algunos casos se pudieron apreciar trabajadores, aunque cabe recalcar que estos no se notaban en su totalidad.

Datos anonimizados

En cuanto a anonimización de los datos, lo cierto es que al tratarse de un plano cenital, no se suele captar información que pudiera comprometer a la privacidad de una persona, o en dado caso, de la institución, sin embargo, dentro de las técnicas para anonimizar datos podemos encontrar el pixelado, que permite mantener toda la información relevante, únicamente anonimizando datos sensibles como logos, placas de vehículos que pudieran estar trabajando, rostros de personas o inclusive identificadores de las vacas.

Otra de las formas de anonimizar es evitando completamente las fotografías en las que aparezca uno de estos datos sensibles, que de hecho fue un punto considerado por el equipo desde el momento en que se hizo el etiquetado de los datos. Sin embargo, una de las desventajas de esta técnica es que se pierden datos y para un dataset pequeño no es la mejor opción.

Existen más técnicas que permiten mantener más datos como el desenfoque o el uso de máscaras negras, sin embargo, al no contar con demasiadas imágenes con datos que requieran anonimización, en este caso su implementación no es un requerimiento.

Normativas

En cuanto a normativas, investigué principalmente que es lo que se encontraba implementado en México, debido a que es en donde se está llevando a cabo el

proyecto. Lo cierto es que en mi investigación no encontré normativas que fueran relacionadas directamente con nuestro caso, ya que es bastante específico y el uso de datos para este tipo de prácticas no es un tema muy explorado en México, sin embargo, si se pueden considerar elementos como los siguientes.

Primero, para las fotografías en donde se puedan ver involucradas personas, la Ley Federal de Protección de Datos Personales en Posesión de los Particulares (Cámara de Diputados del H. Congreso de la Unión, 2023), cuenta con principios y derechos para el manejo adecuado de datos personales en México, protegiendo la información identificable de personas físicas en posesión de particulares, que en este caso seríamos nosotros como equipo. Específicamente en el capítulo II y III, siendo en este último en que se describen los derechos ARCO, que es un acrónimo de Acceso, Rectificación, Cancelación y Oposición, en los que se especifican los lineamientos de los ciudadanos respecto al uso de sus datos personales.

Por otra parte, la Ley Federal de Derecho de Autor (LFDA) regula los derechos de autor en México, protegiendo obras intelectuales, científicas, literarias y artísticas, así como los derechos de los creadores sobre sus obras (Cámara de Diputados del H. Congreso de la Unión, 2023). La LFDA es importante si el uso de las fotografías implica derechos de autor o propiedad intelectual, por ejemplo, si las imágenes fueran de un fotógrafo o una empresa que exige reconocimiento de autoría. Lo cual en nuestro caso no aplica ya que las fotografías fueron tomadas específicamente para el proyecto y se nos brindó la autorización para su uso, sin embargo, en caso de que se quiera escalar el proyecto se tendría considerar para no caer en problemáticas legales.

Por último, se podría considerar el ganado como parte de la propiedad privada del Socio Formador, por lo que el Código Civil Federal en su título de la propiedad (Cámara de Diputados del H. Congreso de la Unión, 2023), habla que solo se podría hacer uso de dichos elementos bajo disponibilidad del S.F., por lo que lo más acertado sería realizar un documento firmado por los interesados en donde se declare este permiso.

Ahora bien, lo cierto es que a nivel internacional si se cuenta con normativas más específicas, de donde destaco el Reglamento General de Protección de Datos (2016) de la Unión Europea que establece un marco avanzado para la protección de datos que podría servirnos como referencia adicional. Este reglamento exige un consentimiento explícito para el tratamiento de datos personales, lo que resalta la importancia de obtener autorización no solo para capturar imágenes, sino también para su uso en el proyecto. Además, el GDPR promueve el principio de "privacy by

design", sugiriendo que la privacidad debe ser considerada desde las etapas iniciales de desarrollo de un proyecto, lo cual podría guiarnos a implementar medidas adecuadas desde el inicio, lo cual es algo que hicimos con la omisión de fotografías.

Otro aspecto clave del GDPR es el derecho de acceso, que permite a los individuos conocer qué datos se tienen sobre ellos y cómo se están utilizando. En el contexto de nuestro proyecto, esto implica que, si alguna persona identificable en las fotografías lo solicita, debemos estar preparados para proporcionar información clara y precisa sobre el uso de sus datos.

Además, el GDPR establece normas sobre la minimización de datos, lo que significa que solo deberíamos recopilar y procesar aquellos datos que sean estrictamente necesarios para los fines del proyecto.

Proceso para trabajar con el dataset

Para trabajar con el dataset, primero se etiquetaron las imágenes, en donde se acordó omitir aquellas imágenes que contaran con algún elemento que requiriera ser anonimizado, como personas y detalles personales o de la institución. También nos encontramos en constante comunicación con el Socio Formador para informarle sobre el uso de las imágenes, que de hecho ya tenían esa función asignada desde que fueron tomadas, por lo que no resultó en ningún inconveniente.

En cuanto a futuro, lo ideal es continuar con la transparencia que hemos tenido con todos los involucrados, incluyendo profesores, además, en caso de que el proyecto escalara, se podría incluir un plan que registre todas las tecnologías que requieren acceso directo a los datos, como herramientas de cómputo en la nube.

Bitácora de buenas prácticas

Para la bitácora me pareció una buena opción utilizar software dedicado al desarrollo de proyectos como Notion o Jira, sin embargo, también es posible realizarla en documentos de Google, que podría ser una forma más cómoda para todos los integrantes.

Es de importancia ya que con esto tendremos evidencia de la forma en la que se trataron los datos y el proyecto en sí a lo largo de todas las etapas y con relación a cada integrante.

Ahora bien, los puntos de la bitácora deben ser discutidos con el equipo, ya que todos nos debemos de comprometer en cumplirlos, sin embargo, dentro de estas prácticas me parece importante abarcar puntos como el consentimiento de datos informado a través de documentación, la anonimización ya mencionada

anteriormente, contar con un plan de capacitación que permita un aprendizaje lo más eficiente posible para el Socio Formador, mantener claridad en el proceso tanto con los otros miembros como con el S. F. y por último, contar con métricas y reportes que muestren evaluaciones de nuestros avances y sus mejoras en caso de requerirse.

A continuación se adjunta el link a la bitácora en nuestro repositorio:

<https://github.com/Mirabay/Vakas/blob/main/docs/data%20security/Bit%C3%A1cora%20de%20Buenas%20Pr%C3%A1cticas.xlsx>

Política de datos y regulaciones de privacidad acordadas por el equipo

Como equipo definimos un proceso detallado para trabajar con los datos brindados por el Socio Formador, el cual se adjunta a continuación:

https://github.com/Mirabay/Vakas/blob/main/docs/data%20security/Data_Privacy_And_Security_TC.pdf

Por otra parte, a continuación se muestran las política de datos y regulaciones de privacidad que se decidieron en conjunto:

https://github.com/Mirabay/Vakas/blob/main/docs/data%20security/Poli%C3%A9tica%20de%20Datos%20y%20Regulaciones%20de%20Privacidad_TC.pdf

Referencias

1. Cámara de Diputados del H. Congreso de la Unión. (2023). Ley Federal de Protección de Datos Personales en Posesión de los Particulares. Recuperado de <https://www.diputados.gob.mx/LeyesBiblio/pdf/LFPDPPP.pdf>
2. Cámara de Diputados del H. Congreso de la Unión. (2023). Ley Federal de Derechos de Autor. Recuperado de <https://www.diputados.gob.mx/LeyesBiblio/pdf/LFDA.pdf>
3. Cámara de Diputados del H. Congreso de la Unión. (2023). Código Civil Federal. Recuperado de <https://www.diputados.gob.mx/LeyesBiblio/pdf/CCF.pdf>
4. General Data Protection Regulation (GDPR). (2016). Regulation (EU) 2016/679 of the European Parliament and of the Council. Recuperado de <https://gdpr-info.eu/>