ITESM- Campus Puebla

Privacidad y Seguridad de los Datos

Inteligencia artificial avanzada para la ciencia Datos II

Integrantes Equipo 1:

Myroslava Sánchez Andrade A01730712
José Antonio Bobadilla García A01734433
Karen Rugerio Armenta A01733228
Alejandro Castro Reus A01731065

Fecha: 04/11/2022

En seguridad informática y en la seguridad de la información existen tres pilares:

- **Confidencialidad:** Otro término para la confidencialidad sería privacidad. Las políticas de la empresa deben restringir el acceso a la información al personal autorizado y garantizar que solo las personas autorizadas verán estos datos. Los datos se pueden dividir en secciones según el nivel de seguridad o sensibilidad de la información.

- **Integridad**: La integridad es precisión, consistencia y confiabilidad de los datos durante su ciclo de vida. Los datos deben permanecer inalterados durante la transferencia y no deben ser modificados por entidades no autorizadas.

- **Disponibilidad**: Dar mantenimiento a los equipos, realizar reparaciones de hardware, dar mantenimiento a los sistemas operativos y el software actualizados, así como crear respaldos, garantizar la disponibilidad de la red y los datos a los usuarios autorizados.

La confidencialidad está directamente relacionada con la privacidad de los datos y es un tema fundamental y de alto riesgo en la ciencia de datos, debido a que muchos de los sets de datos pueden contener información personal identificable. Algunas de las empresas que manejan grandes cantidades de datos, como Google, consideran que la información que se pueda usar por sí sola para identificar o ubicar con precisión a una persona, o para ponerse en contacto con ella de forma directa. Entre otros datos, incluye lo siguiente:

- Direcciones de correo electrónico
- Direcciones de correo postal
- Números de teléfono
- Ubicaciones precisas (por ejemplo, coordenadas GPS, salvo en los casos que se mencionan más abajo)
- Nombres completos (nombre y apellidos) o nombres de usuario

Además, en cada país existen leyes que regulan la colecta, el uso, almacenamiento y transferencia de datos personales, en especial los que son identificables. Por ejemplo, en la Unión Europea (UE) existe la GDPR, del inglés, *General Data Protection Regulation* o, en español, Reglamento General de Protección de Datos. Ésta reglamenta el procesamiento de los datos de usuarios individuales de la UE. Esa ley es aplicada en empresas con y sin

presencia física en territorio europeo, basta que el sitio de la empresa atienda a los usuarios de la unión para que se sometan a la reglamentación. Esto no solo afecta a empresas europeas, sino también a una empresa Mexicana que utilice datos de una organización o individuos de la UE, también está sujeto a apegarse a la GDPR.

De igual manera aplica con otras regulaciones, por ejemplo, como con las de Estados Unidos tales como HIPAA (*Health Insurance Portability and Accountability Act*, es decir, la Ley de Transferencia y Responsabilidad de Seguro Médico) que es una ley que protege la información de pacientes médicos, así como la ley PCI-DSS (*Payment Card Industry Data Security Standard*, es decir, el Estándar de Seguridad de Datos para la Industria de Tarjeta de Pago) el cual regula el uso de aplicaciones o plataformas donde se procesen pagos con información bancaria de los usuarios.

En México existe la Ley Federal de Protección de Datos Personales en Posesión de Particulares permite el tratamiento de datos personales por organizaciones y empresas en determinadas situaciones, pero siempre teniendo como objetivo preservar al usuario. Sin embargo, para eso es necesario que exista consentimiento o legítimo interés en el uso de los datos en cuestión. Los datos personales también pueden ser considerados “sensibles” cuando dan a conocer información íntima de la persona. Son datos sensibles aquellos que revelan el origen racial o étnico, estado de salud, ideología, opiniones políticas, orientación sexual, y otros. Para aplicar la ley que protege a los datos de los usuarios, son establecidos algunos principios por la Ley de Protección de Datos, son:

1. **Licitud**: deberá cumplirse la Ley de Protección de Datos, su Reglamento y cualquier otra regulación aplicable al tratar los datos.

2. **Consentimiento**: siempre que sea necesario deberá obtenerse el consentimiento de los titulares para el tratamiento de sus datos personales;

3. **Información**: debe ser de conocimiento del titular la información relacionada con el tratamiento de sus datos;

4. **Calidad**: todos los datos personales recolectados deben ser exactos, completos, pertinentes, correctos y actualizados de acuerdo con sus finalidades;

5. **Finalidad**: el tratamiento de datos debe cumplir las finalidades establecidas en el aviso de privacidad presentado al usuario;

6. **Lealtad**: proteger los intereses del titular y la expectativa razonable de privacidad al tratar los datos personales;

7. **Proporcionalidad**: sólo los datos personales necesarios, adecuados y relevantes en relación con las finalidades podrán ser tratados;

8. **Responsabilidad**: la empresa debe responder por el tratamiento de los datos personales recolectados.

Es por eso que las empresas Mexicanas están obligadas a:

1. Ofrecer un aviso de privacidad en tu sitio web
2. Tener el consentimiento del titular para la colecta, uso, transferencia y eliminación de los datos.
3. Permitir el acceso, rectificación, cancelación y oposición del titular

Asimismo, la Comisión Económica para América Latina y el Caribe (CEPAL), organismo dependiente de la Organización de las Naciones Unidas (ONU), describe la anonimización como el proceso de convertir los datos de tal manera que no se pueda identificar a individuos. Por tal razón la anonimización se considera como una herramienta fundamental para mitigar los riesgos que presentan la obtención y tratamiento masivo de los datos de carácter personal. El proceso de anonimización de datos consiste en identificar y ocultar la información sensible de las personas u organizaciones a las que se hace referencia, esto permite la divulgación del contenido sin vulnerar los derechos a la protección de datos de los individuos y organizaciones involucradas. Es por eso que es primordial el enmascaramiento de datos, el cual se trata de un proceso de transformación de los datos orientado a la protección, en el que es primordial el intentar mantener el realismo de los mismos. Se trata, a su vez, de un proceso que no permite posteriormente retroceder y recuperar los datos iniciales. A cualquier acción

de enmascaramiento se la podría definir como el reemplazo de toda la información que se pueda considerar sensible en cualquier base de datos que uno pueda tener.

El set de datos proporcionado proviene de una plataforma pública, es por ello que la información que se muestra están anonimizados y por lo tanto no fue necesario realizar técnicas de anonimización ya que no se tienen datos que permitan identificar a una persona u organización en específico. Si los hubiera tenido, hubiera sido imperativo despojar los identificadores personales obvios tales como nombres de una pieza de información, para crear un conjunto de datos en el que no hay identificadores de persona presentes. Algunos ejemplos serían:

- Eliminar direcciones específicas o coordenadas GPS y únicamente utilizar el código postal.
- Quitar los nombres de los individuos y en vez solo utilizar su género.
- Aplicar la ley del menor privilegio, que es darle a los usuarios/desarrolladores/administradores, únicamente el acceso necesario a los datos para realizar sus labores.

Es importante recalcar, que esto no puede aplicarse como una regla general ya que puede haber situaciones en donde, si el muestreo está desequilibrado, es posible deducir e identificar a un individuo u organización, como en los siguientes casos:

- Si una región geográfica contiene pocas empresas o fábricas, el código postal puede ser suficiente para perfilar la organización en cuestión.
- Si en un grupo hay un mayor número de individuos de cierto género, es posible perfilar al individuo que corresponde a la minoría.

Como una recomendación de mejora a la plataforma del socio formador, se propone el crear una página de inicio de sesión para que el acceso a los datos sea controlado otorgando la ley del menor privilegio y para tener una auditoría de quién y en qué momento accede a qué datos (logs/efemérides).