Instituto Tecnológico y de Estudios Superiores de Monterrey

Escuela de Ingeniería y Ciencias

Ingeniería en Ciencia de Datos y Matemáticas

Momento de Retroalimentación: Reto Privacidad y Seguridad de los Datos

Inteligencia artificial avanzada para la ciencia de datos II

María Fernanda Torres Alcubilla A01285041

Morales Ramón Michelle Yareni A01552627

Paola Sofia Reyes Mancheno A00831314

Federico Medina García Corral A01721441

Supervisado por:
Iván Mauricio Amaya Contreras, Ph.D.
Edgar Covantes Osuna, Ph.D.
Hugo Terashima Marìn, Ph.D.

Monterrey, Nuevo León. Fecha, 29 de octubre de 2023

1. Datos Anonimizados

Para poder mantener los datos anónimos, se optó por utilizar diferentes claves para la creación de usuarixs, tanto del alumnado como de docentes. Por ejemplo, en lugar de que alguien se registre utilizando su nombre completo, el sistema le va a indicar que únicamente se pueden introducir valores que contengan números y/o letras. En el caso de pruebas, se han utilizado algunas matrículas de ejemplo del Tecnológico de Monterrey para mantener anónima la identificación de los registros, ya que únicamente una persona con acceso al sistema de la institución puede verificar el nombre de lxs usuarixs. Este fue el primer acercamiento que se tuvo, y nos dió la flexibilidad de mantener la privacidad de lxs usuarixs de todas aquellas personas que no pertenecen al Tec de Monterrey.

Por otro lado, para poder generalizar el uso de la interfaz y el modelo para aplicaciones fuera del Tec de Monterrey, se ha optado por utilizar la herramienta de Firestore para almacenar la información de lxs usuarixs que se registren al sistema. Su funcionamiento consta de varios pasos, al ingresar a la plataforma por primera vez se debe crear un perfil con el correo institucional y automáticamente se verifica si este correo no se encuentra registrado, en tal caso, se toman las fotos del rostro de la persona para entrenar el modelo de reconocimiento facial, las cuales se guardan en Google Cloud Storage. Una vez registrado, de manera automática se identifica el rol de la persona (estudiante o docente) y dependiendo de este se dirige a la plataforma correspondiente.

El acceso a Google Cloud estará restringido para el público, con excepción del host y lxs usuarixs específicos que el host mismo asigne. Cabe recalcar que toda la información dentro de la nube está encriptada y administrada por Google, lo cual da un poco de flexibilidad y ventaja al host para no tener que encargarse de dicha tarea.

Una vez teniendo esto, lxs usuarixs serán almacenados en FireStore, donde la información como correos electrónicos y contraseñas estarán guardadas y hasheadas. En un futuro, se buscará cambiar la matrícula en los registros de las fotos de lxs usuarixs por el UID, el cual es creado automáticamente por Firestore, esto nos permite tener un mejor control de la seguridad ya que no se encuentra relacionado con la información sensible de lxs usuarixs. Igualmente, dentro de este, se utiliza algo llamado Seguridad de Capa de Transporte (STP) la cual sirve para proteger los datos a medida que viajan por Internet durante las operaciones de lectura y escritura [“Encriptación en reposo predeterminada”, 2023]. Por otro lado, para los datos que están estacionarios o en reposo, se encriptan en la capa de almacenamiento mediante el algoritmo AES-256. Para este, se

usa la biblioteca criptográfica Tink, que incluye el módulo con validación del estándar FIPS 140-2, conocido como BoringCrypto, para implementar la encriptación de manera coherente en Google Cloud.[“Encriptación del servidor”, 2023]

2. Pasos Comunes para Garantizar la Privacidad de los Datos

En la actualidad las empresas recopilan grandes cantidades de datos sobre lxs usuarixs. Estos datos abarcan desde información demográfica hasta detalles más específicos como “ingresos”, “historial de navegación” o “lugares visitados recientemente”.

Cuando se combinan, esta Información de Identificación Personal puede utilizarse para estimar la ubicación precisa de lxs usuarixs, los productos que adquiere con frecuencia, sus creencias políticas, entre otras cosas. Es por ello que se requiere de estrategias para mantener la seguridad y privacidad de los datos.

Algunas de las prácticas que garantizan la privacidad de los datos y el cumplimiento de normativas son las siguientes.

- Inventario de los datos. Para asegurar la confidencialidad de la información se debe comprender qué datos se poseen, cómo se gestionan y dónde se almacenan, después, se deben establecer procesos de recopilación y manejo de esta información; por ejemplo, determinar la frecuencia de análisis de los datos y su posterior clasificación. En este contexto, las regulaciones deben detallar de forma explícita las medidas requeridas para los diferentes niveles de privacidad de datos y seguridad. Además, las políticas deben incorporar procedimientos para llevar a cabo auditorías de estas medidas con el fin de garantizar una correcta implementación de las soluciones.

- Minimizar la recopilación de datos. Asegurar que las políticas establezcan que se adquieran únicamente los datos esenciales. Si se llega a recopilar más información de la necesaria, se incrementa la carga de responsabilidad. Reducir esta dimensión también contribuirá a la optimización de los recursos de ancho de banda y espacio de almacenamiento.

- Transparencia con lxs usuarixs. El consentimiento de lxs usuarixs es importante para la recopliación de datos por lo que es necesario notificarles cuando se recopilará su información y

para qué, así como incluir opciones para habilitar o desabilitar el acceso a sus datos.

- Llevar a cabo análisis de riesgos y pruebas de seguridad. con el análisis se pueden identificar, evaluar y proponer planes de mitigación de riesgos, además, se puede complementar con las pruebas de seguridad, programadas con anterioridad cada cierto tiempo, esto para verificar si las implementaciones funcionan correctamente.

- Capacitación del personal. Además de la seguridad de la red, se debe asegurar que las personas que tratan con los datos o se encuentran dentro de la red conozcan medidas y tengan las mejores prácticas en cuanto a la seguridad y privacidad de los datos.

3. Proceso para la Trata de Datos en el Dataset

Para poder trabajar con la trata de datos registrados, se debe de recordar que únicamente el host tendrá acceso a estos al igual que aquellos individuos que este dé los permisos adecuados a cada unx. Para asegurar esto, se debe de considerar hacer un contrato en donde se establezcan todos los detalles sobre cómo uno puede acceder a los datos, cuándo y para qué. Esto último con la intensión de mantener confidenciales los datos y en donde se pueda tomar acción legal en caso de que alguno rompa con dicho contrato.

En el caso actual, los datos estarán en el servicio de Google Cloud en regiones fuera de la UE, además, se hace la suposición que la trata de datos solo se dará fuera de la UE y no se harán uso de datos de personas residentes de estos países. Esto debido a que al entrar a la Unión Europea, en base al Reglamento General de Protección de Datos (GDPR), en caso de que se utilice el programa en dicha región, los desarrolladores que recopilen o procesen datos de usuarios "a gran escala", recopilen o procesen ciertos tipos de datos sensibles deben designar un Oficial de Protección de Datos (DPO) o a un representante de la Unión Europea. Es por esto mismo que los servidores en donde se almacenarán los datos por el momento serán fuera de la Unión Europea ya que no hemos contactado a un DPO ni a un representante de la UE.

Igualmente, se debe de considerar que la red en donde se establecerá debe de ser una red privada para no comprometer los datos a un ataque de manera más fácil por ser pública. De la misma forma, recalcando lo mencionado anteriormente, la única persona que tendrá acceso a la información confidencial será el host y todo aquel que este asigne para poder acceder a los datos,

después de firmar un contrato legal debido para mantener la confidencialidad de los datos.

4. Access Request Management

Algo muy importante para mantener la seguridad de los datos antes, durante y después de su uso, es la rendición de cuentas. Esta se refiere a que, además de especificar quiénes pueden acceder a la información y cuál es la razón, se necesita regular los permisos y registrar la actividad de estas personas. Esto con el fin de poder realizar auditorías continuas del acceso a la información y detectar cualquier comportamiento anómalo.

Para poder lograr este objetivo, existen varias herramientas y aplicativos que permiten gestionar permisos de acceso y llevan registro de estos. No obstante, tomando en cuenta que este proyecto será montado sobre Google Cloud, es posible hacer uso de las herramientas que esta plataforma de servicios en la nube proporciona.

En primer lugar, se puede activar la Aprobación de Acceso en el proyecto de Google Cloud donde se encuentren todos los servicios que utilizamos. Esta Aprobación de Servicio permite dar y quitar permisos a usuarixs dentro del proyecto. [“Descripción general de la aprobación de acceso”, 2023] Con esto, se asegura el acceso al proyecto únicamente a las personas encargadas de su correcto funcionamiento.

Posterior a esto, se utilizan los Registros de Auditoria, los cuales, al activarse registran qué usuarix accedió a qué servicio, y qué es lo que hizo en ese servicio en un determinado tiempo. Esta funcionalidad puede también registrar la lectura de datos, y no solo cuando hay un cambio a estos o los servicios como tal [“Información sobre los registros de auditoria”, 2023]. Esto permite tener un registro mucho más completo sobre lo que lxs usuarixs están haciendo dentro del proyecto, y por ende se puede asegurar el cumplimiento de las normativas del tratamiento de los datos.

Referencias

Descripción general de la aprobación de acceso. (2023, agosto). https://cloud.google.com/assured-workloads/access-approval/docs/overview?hl=es-419

Encriptación del servidor. (2023, agosto). https://cloud.google.com/firestore/docs/server-side-encryption?hl=es-419

Encriptación en reposo predeterminada. (2023, agosto). https://cloud.google.com/docs/security/encryption/default-encryption?hl=es-419

Información sobre los registros de auditoria. (2023, abril). https://cloud.google.com/compute/docs/logging/audit-logging?hl=es-419