Práctica Profesional Supervisada

Alanís, Gastón Emmanuel

Estrategias de Protección de Datos Personales : Análisis de Riesgos y Propuestas de Implementación en Redes Sociales y Pequeños Proyectos

Instituto de Ingeniería y Agronomía

Carrera: Ingeniería en Informática

2025

Universidad Nacional Arturo Jauretche

Instituto de Ingeniería y Agronomía

Ingeniería en Informática

PRÁCTICA PROFESIONAL SUPERVISADA

Informe

Estrategias de Protección de Datos Personales: Análisis de Riesgos y Propuestas de Implementación en Redes Sociales y Pequeños Proyectos

---

Gastón Emmanuel Alanís

Resumen

Este estudio abordará la problemática de la exposición de datos personales en redes sociales y pequeños proyectos, destacando los riesgos de privacidad y la falta de concientización. Se compararán normativas como la Ley 25.326 en Argentina y el GDPR en Europa, identificando sus alcances y limitaciones. Además, se analizarán herramientas de anonimización como k anonymity y privacidad diferencial, evaluando su aplicabilidad en contextos accesibles. Finalmente, se desarrollarán estrategias prácticas y recomendaciones para mejorar la protección de datos en pequeños proyectos y entornos educativos, fomentando una mayor seguridad digital.

Palabras Clave: Protección de datos personales, privacidad digital, anonimización, redes sociales, normativas de privacidad, Ley 25.326, GDPR, seguridad de la información, gestión de datos, pequeñas empresas, ciberseguridad, filtración de datos, concientización, herramientas de anonimización, privacidad diferencial, k-anonymity.

Abstract

This study will address the issue of personal data exposure on social networks and small projects, highlighting privacy risks and the lack of awareness. Regulations such as Argentina's Law 25.326 and the GDPR in Europe will be compared, identifying their scope and limitations. Additionally, anonymization tools such as k-anonymity and differential privacy will be analyzed, assessing their applicability in accessible contexts. Finally, practical strategies and recommendations will be developed to enhance data protection in small projects and educational environments, promoting greater digital security.

Índice general

1. Introducción……...7

2. Marco Teórico ……10

2.1 Artículos clave de la Ley 25.326 y (GDPR)……..11

2.2 Conceptos generales……..14

2.3 Herramientas para Implementación en Pequeños Proyectos…..44

3. Mejores Prácticas de Protección de Datos……..56

4. Estudios de Caso y Lecciones Aprendidas…….……63

5. Protocolo General de Protección de Datos Personales……..……72

6.Conclusiones y Recomendaciones……77

6.1 Trabajos futuros……...79

7. Bibliografía y Anexos……..……84

Indice de figuras

Figura 1 Ventajas y limitaciones de los algoritmos de anónimizacion…...19

Figura 2 Diagrama de flujo k-anonymity……..21

Figura 3 Diagrama de flujo Differential privacy…….23

Figura 4 Diagrama de flujo toketizacion……..24

Figura 5 Diagrama de flujo Data Masking……25

Figura 6 Implementación de algoritmos de anonimizacion……26

Figura 7 Evaluación de la eficiencia de las técnicas de anonimización….43

Figura 8 Diagrama de flujo Aplicación de Técnicas de Protección de Datos……76

índice de tablas

Tabla 1 Comparación con normativas internacionales……...15

Tabla 2 Ejemplo k-anonymity…….21

Tabla 3 Ejemplo k-anonymity……..21

Tabla 4 Ejemplo Differential privacy…….22

Tabla 5 Ejemplo toketizacion…….24

Tabla 6 Ejemplo Data Masking……25

Tabla 7 Comparativa entre herramientas de anonimización……..45

Tabla 8 Ejemplo para aplicar Amnesia……52

Tabla 9 Datos anonimizados……...53

Capítulo I

Introducción:

En este capítulo aborda cómo la exposición de información personal en redes sociales representa un riesgo creciente para la privacidad y la seguridad, tanto a nivel individual como global. Señala que el uso masivo de estas plataformas en países como España, Estados Unidos, México, Alemania, China y Argentina ha facilitado el acceso a datos sensibles. En particular, destaca la situación en Argentina y la necesidad de generar conciencia y estrategias colectivas para proteger la privacidad en el entorno digital.

En la sociedad actual, compartimos información personal constantemente, muchas veces con personas que no conocemos, lo cual pone en riesgo nuestra seguridad. Datos como fotos de nuestros hogares, números de teléfono o direcciones pueden volverse perjudiciales si llegan a manos equivocadas. Esta exposición no solo afecta nuestra privacidad, sino que también pone en peligro nuestra seguridad emocional, económica y física.

Dado que la práctica profesional supervisada no se desarrolló en una organización concreta, el presente trabajo se enfoca en el diseño e implementación de un protocolo aplicado, basado en el análisis conceptual y técnico realizado a lo largo de la carrera, con el fin de brindar una solución adaptada a entornos reales de bajo recurso.

Con el auge de las redes sociales y plataformas digitales, la exposición de nuestra vida privada es mayor que nunca. Se estima que más de 4.700 millones de personas utilizan redes sociales, lo que representa casi el 60% de la población mundial. Esta globalización digital ha generado una accesibilidad nunca antes vista a datos sensibles, facilitando su uso para quienes tengan malas intenciones. A nivel global, los usuarios de redes sociales en países como Estados Unidos, España, México, Alemania, China y Argentina interactúan y comparten información a diario, contribuyendo a la expansión de una sociedad digital interconectada.

Por ejemplo, en España, el 80% de los usuarios accede a las redes sociales diariamente, pasando un promedio de 42 minutos al día en estas plataformas. En Estados Unidos, el 91% de la población es usuaria de internet, y un alto porcentaje de ellos está presente en redes sociales. En México, un 81% de los usuarios de internet están activos en redes sociales. En Alemania, la cifra alcanza al 85%, y en China, más de 1.060 millones de personas ya utilizan redes sociales. Esta presencia masiva de usuarios en todo el mundo nos demuestra la importancia de tener un contexto global cuando hablamos de los riesgos asociados al manejo y la protección de nuestra información personal.

Es fundamental comprender que la privacidad y seguridad de nuestros datos no solo son cuestiones locales, sino que son parte de un fenómeno global. Las amenazas a nuestra seguridad en línea son universales, y el entendimiento de estas problemáticas a nivel mundial nos permite contextualizar mejor los casos específicos de cada país.

En el caso de Argentina, más del 68% de la población utiliza redes sociales, dedicando un promedio de 9 horas y 38 minutos diarios al acceso a internet. Sin embargo, la falta de conciencia sobre los riesgos asociados a la exposición digital en las redes sociales es un problema creciente. Un estudio indica que el 46% de los jóvenes argentinos de entre 16 y 24 años se siente abrumado por el uso de las redes sociales, mientras que el 42% de los adultos se muestra preocupado por la exposición de su información personal en línea.

Este contexto global resalta la necesidad urgente de una estrategia colectiva para proteger la privacidad de los usuarios, no solo en Argentina, sino en el mundo entero. La protección de la privacidad no solo resguarda nuestra integridad, sino que también garantiza nuestra seguridad emocional, económica y física. Este análisis busca concientizar sobre los peligros asociados a la exposición de información personal y ofrecer estrategias efectivas para proteger nuestra privacidad en la era digital, comenzando con un entendimiento global y luego enfocándonos en el caso específico de Argentina.

Capitulo II

Marco teórico.

El marco teórico aborda la Ley 25.326 de protección de datos personales en Argentina, sus principios clave y limitaciones frente a las nuevas tecnologías. Se comparan normativas internacionales como el GDPR y se destacan los desafíos que enfrentan pequeñas empresas para cumplir con la ley. También se explican conceptos técnicos como la anonimización y la pseudonimización, sus usos, ventajas y riesgos.

El marco legal en Argentina en materia de protección de datos personales está definido principalmente por la **Ley 25.326**, sancionada en el año 2000. Esta normativa establece principios fundamentales para garantizar la privacidad y el correcto tratamiento de la información personal de los ciudadanos. Uno de los pilares de la ley es el consentimiento informado, lo que significa que cualquier recopilación de datos personales debe contar con la autorización expresa de su titular. Además, la norma exige que la recolección y el uso de estos datos respondan a una finalidad específica, es decir, que solo puedan ser utilizados para el propósito previamente declarado y no para otros fines sin el consentimiento del usuario.

Otro aspecto clave de la ley es la seguridad, ya que obliga a quienes manejan datos personales a implementar medidas técnicas y organizativas que garanticen su protección, evitando accesos no autorizados, filtraciones o usos indebidos. Para asegurar el cumplimiento de estas disposiciones, se designó a la Agencia de Acceso a la Información Pública (AAIP) como la autoridad de control encargada de supervisar y hacer cumplir la normativa, además de recibir y gestionar denuncias relacionadas con el mal uso de datos personales.

A pesar de su importancia, la Ley 25.326 presenta ciertas limitaciones en el contexto actual, ya que fue creada en una época en la que la digitalización y el desarrollo tecnológico aún no habían alcanzado la magnitud que tienen hoy en día. En particular, la ley no contempla de manera explícita tecnologías emergentes como el Internet de las Cosas (IoT) o la inteligencia artificial (IA), las cuales generan nuevos desafíos en términos de privacidad y seguridad de la información. Asimismo, en comparación con normativas más recientes como el Reglamento General de Protección de Datos (GDPR) de la Unión Europea, las sanciones establecidas en la legislación argentina son menos rigurosas, lo que limita su capacidad disuasoria frente a incumplimientos.

Debido a estos desafíos, en los últimos años se han planteado propuestas para modernizar la legislación y adecuarla a los estándares internacionales, con el objetivo de reforzar la protección de los datos personales en el país y garantizar un marco normativo acorde a las nuevas realidades tecnológicas.

2.1 Artículos clave de la Ley 25.326 y del Reglamento General de Protección de Datos (GDPR)

A continuación se presentan los artículos más relevantes de la Ley 25.326 de Protección de los Datos Personales de la República Argentina y del Reglamento General de Protección de Datos (GDPR) de la Unión Europea. Estos artículos constituyen los pilares normativos que regulan el tratamiento de datos personales y son fundamentales para comprender los principios, derechos y obligaciones que se abordan en este trabajo.

Ley 25.326 – Protección de los Datos Personales (Argentina)

- **Artículo 1 – Objeto**

  Establece que la ley tiene por objeto la protección integral de los datos personales asentados en archivos, registros, bancos de datos u otros medios técnicos, sean públicos o privados, para garantizar el derecho al honor y a la intimidad de las personas.

- **Artículo 2 – Definiciones**

  Incluye definiciones clave como “dato personal”, “dato sensible”, “archivo”, “tratamiento de datos”, “titular”, “responsable” y “cesionario”, las cuales permiten interpretar adecuadamente el alcance de la ley.

- **Artículo 5 – Consentimiento**

  Dispone que el tratamiento de datos personales requiere el consentimiento libre, expreso e informado del titular, salvo en los casos expresamente autorizados por la ley.

- **Artículo 6 – Información al titular**

  Obliga al responsable a informar al titular, de manera clara, sobre la finalidad del tratamiento, los destinatarios de los datos, el carácter obligatorio o facultativo de sus respuestas y sus derechos de acceso, rectificación y supresión.

- **Artículo 8 – Datos sensibles**

  Prohíbe, en general, el tratamiento de datos sensibles (como los relacionados con salud, orientación sexual, religión o afiliación sindical), salvo en circunstancias excepcionales como fines estadísticos o científicos que no permitan identificar al titular.

- **Artículo 9 – Seguridad de los datos**

  Establece que los responsables deben adoptar medidas técnicas y organizativas para garantizar la seguridad e integridad de los datos, evitando su adulteración, pérdida o acceso no autorizado.

- **Artículo 10 – Confidencialidad**

  Toda persona que intervenga en el tratamiento de datos personales está obligada a guardar secreto profesional, incluso después de finalizada su relación con el archivo.

- **Artículo 14 – Derecho de acceso**

Reconoce el derecho de toda persona a acceder gratuitamente a sus datos personales, previa acreditación de identidad, una vez cada seis meses, salvo que se demuestre un interés legítimo.

- **Artículo 16 – Rectificación, actualización y supresión**

  Permite a los titulares solicitar la rectificación, actualización o supresión de sus datos cuando sean inexactos, incompletos o no se ajusten a lo dispuesto en la ley.

- **Artículo 21 – Transferencias internacionales**

  Restringe la transferencia de datos personales a países que no ofrecen niveles adecuados de protección, salvo que se cuente con el consentimiento del titular o se cumplan excepciones previstas.

- **Artículo 29 – Autoridad de aplicación**

  Establece que la autoridad de aplicación es la Agencia de Acceso a la Información Pública (AAIP), organismo encargado de controlar el cumplimiento de la ley y dictar normas complementarias.

Reglamento General de Protección de Datos – GDPR (Unión Europea)

- **Artículo 5 – Principios del tratamiento**

  Enuncia principios fundamentales como la licitud, lealtad y transparencia del tratamiento; la limitación de la finalidad; la minimización de datos; la exactitud; la limitación del plazo de conservación; la integridad y confidencialidad; y la responsabilidad proactiva (*accountability*).

- **Artículo 6 – Licitud del tratamiento**

  Detalla las condiciones que hacen lícito el tratamiento de datos, como el consentimiento del interesado, el cumplimiento de una obligación legal, la ejecución de un contrato, la protección de intereses vitales o el interés legítimo del responsable.

- **Artículo 7 – Condiciones del consentimiento**

  Establece que el consentimiento debe ser libre, específico, informado e inequívoco, y que el interesado debe poder retirarlo en cualquier momento.

- **Artículo 17 – Derecho al olvido**

  Reconoce el derecho de las personas a solicitar la supresión de sus datos personales cuando ya no sean necesarios para los fines del tratamiento,

cuando se haya retirado el consentimiento, o cuando se haya producido un tratamiento ilícito.

- **Artículo 20 – Portabilidad de los datos**

  Otorga a los interesados el derecho a recibir sus datos en un formato estructurado y de uso común, y a transmitirlos a otro responsable del tratamiento sin impedimentos.

- **Artículo 25 – Privacidad desde el diseño y por defecto**

  Impone la obligación de aplicar, desde la fase de diseño de los sistemas y servicios, medidas técnicas y organizativas que garanticen la protección de los datos, limitando su tratamiento por defecto a lo estrictamente necesario.

- **Artículo 32 – Seguridad del tratamiento**

  Exige la implementación de medidas adecuadas —como cifrado o seudonimización— para garantizar un nivel de seguridad adecuado al riesgo, considerando el estado de la técnica y el coste de aplicación.

- **Artículo 35 – Evaluación de impacto**

  Obliga a realizar evaluaciones de impacto en protección de datos antes de iniciar tratamientos que puedan implicar un alto riesgo para los derechos y libertades de las personas, como los que utilizan tecnologías nuevas o implican vigilancia sistemática.

- **Artículo 83 – Sanciones administrativas**

  Establece la posibilidad de imponer multas de hasta 20 millones de euros o el 4% del volumen de negocios anual global del infractor, según cuál sea mayor, por incumplimientos graves del reglamento.

2.2. Conceptos generales

Legislación y Normativas Internacionales

En los últimos años, se han impulsado diversos proyectos para actualizar la Ley 25.326 y alinearla con el Reglamento General de Protección de Datos (GDPR) de la Unión Europea y otros estándares internacionales. Estas iniciativas buscan modernizar el marco normativo argentino para responder a los desafíos que plantean las nuevas tecnologías y garantizar un mayor nivel de protección para los ciudadanos en el entorno digital.

Entre las principales propuestas se destaca la inclusión de nuevos derechos, como el derecho al olvido, que permitiría a los usuarios solicitar la eliminación de su información personal en determinadas circunstancias, especialmente en entornos digitales. También se plantea la incorporación del principio de portabilidad de datos, que otorgaría a las personas la posibilidad de trasladar su información personal de una plataforma a otra sin restricciones, promoviendo así un mayor control sobre su propia información.

Además, las propuestas de actualización buscan reforzar las obligaciones de transparencia para las empresas y organismos que manejan datos personales, estableciendo mayores requisitos en cuanto a la obtención del consentimiento, el tiempo de almacenamiento de los datos y la implementación de medidas de seguridad avanzadas. Estas modificaciones permitirían una supervisión más estricta y sanciones más severas en caso de incumplimiento, acercando la legislación argentina a los estándares europeos y fortaleciendo la confianza de los ciudadanos en el manejo de su información personal.

Aspecto	**Argentina (Ley 25.326)**	**UE (GDPR)**	**EE.UU (CCPA)**
Alcance	Solo entidades locales	Empresas globales que operen en la UE	Empresas con ingresos > $25M en California
Multas máximas	Hasta ARS 100.000	Hasta 4% del volumen anual	Hasta $7.500 por violación
Derechos del titular	Acceso, rectificación	Acceso, rectificación, olvido, portabilidad	Acceso, eliminación, opt-out
Anonimazacion	Mencionada como medida de seguridad	Requiere irreversibilidad	Tratada como "información no personal"

Tabla 1 Comparación con normativas internacionales

Retos Legales y Éticos

Los pequeños proyectos y startups enfrentan importantes desafíos a la hora de cumplir con la legislación sobre protección de datos personales. Uno de los principales obstáculos es la ambigüedad en los estándares, ya que la normativa actual no siempre es clara en cuanto a qué técnicas de anonimización o minimización de datos son suficientes para garantizar el cumplimiento legal. Esto genera incertidumbre y dificulta la toma de decisiones sobre cómo manejar la información personal sin incurrir en posibles infracciones.

Otro reto significativo es el costo de cumplimiento, ya que muchas de las medidas exigidas por regulaciones como el GDPR de la Unión Europea pueden resultar costosas de implementar para proyectos de menor escala. Aspectos como la adopción de infraestructura segura, auditorías constantes, cifrado avanzado y la designación de un responsable de protección de datos representan una carga económica considerable, lo que pone en desventaja a startups y emprendedores frente a grandes corporaciones con mayores recursos.

La falta de guías específicas y de apoyo gubernamental para estos casos hace que muchas pequeñas empresas deban navegar en un entorno normativo complejo, en el que cumplir con la ley puede requerir inversiones significativas en asesoramiento legal y tecnológico. Esto resalta la necesidad de regulaciones más adaptadas a distintos niveles de negocio y la creación de herramientas accesibles que permitan a los pequeños proyectos cumplir con la normativa sin comprometer su viabilidad económica.

Actualización normativa y alineación internacional

En los últimos años, tanto Argentina como diversos organismos internacionales han impulsado nuevas directrices para fortalecer la protección de los datos personales frente al avance tecnológico. Esta actualización normativa resulta esencial para afrontar los desafíos emergentes derivados del uso masivo de datos en contextos como la inteligencia artificial, el Internet de las Cosas y la automatización de decisiones.

Reforma de la Ley 25.326 (proyecto 2023)

En 2023, la Agencia de Acceso a la Información Pública (AAIP) presentó un proyecto de reforma integral a la Ley 25.326 de Protección de Datos Personales, con el objetivo de actualizar su contenido y adecuarlo a los estándares internacionales. Entre las principales novedades del anteproyecto se incluyen la incorporación de nuevos derechos para los titulares de datos, como el derecho al olvido, la portabilidad y la limitación del tratamiento; definiciones más precisas sobre datos biométricos, datos sensibles y tratamientos automatizados mediante tecnologías emergentes; el refuerzo del principio de responsabilidad proactiva (accountability), que obliga a los responsables a demostrar el cumplimiento normativo y no solo a declararlo; así como un aumento del poder sancionatorio de la autoridad de control y la implementación de medidas más eficaces de supervisión. Estas reformas buscan restablecer la adecuación del régimen argentino ante la Unión Europea, lo cual es clave para facilitar las transferencias internacionales de datos y promover la

interoperabilidad regulatoria (Fuente: Agencia de Acceso a la Información Pública, AAIP, “Bases para la reforma de la Ley 25.326”, 2023).

Por otro lado, la norma internacional ISO/IEC 27701, publicada en 2019, representa una extensión de los Sistemas de Gestión de Seguridad de la Información (SGSI) definidos por ISO/IEC 27001, enfocándose específicamente en la gestión de la privacidad. Esta norma proporciona un marco estructurado para implementar un Sistema de Gestión de Información de Privacidad (SGIP), definiendo controles específicos según el rol (responsable o encargado del tratamiento), integrando requisitos de cumplimiento con normativas como el GDPR o la CCPA, y ofreciendo guías para aplicar principios clave de privacidad como la minimización de datos, el consentimiento informado, la retención limitada y la transparencia. Además, la ISO/IEC 27701 permite la certificación formal, lo que ayuda a las organizaciones a demostrar su compromiso con la privacidad. Para pequeñas empresas y startups, adoptar este estándar no solo mejora sus prácticas internas, sino que también genera confianza con socios, clientes e inversores, especialmente para operar en mercados internacionales.

Estándar ISO/IEC 27701:2019

La norma internacional ISO/IEC 27701, publicada en 2019, es una extensión de los Sistemas de Gestión de Seguridad de la Información (SGSI) definidos por la ISO/IEC 27001, enfocándose en la gestión de la privacidad. Esta norma proporciona un marco estructurado para implementar un Sistema de Gestión de Información de Privacidad (SGIP), que incluye la definición de controles específicos según el rol que se desempeñe, ya sea responsable o encargado del tratamiento de datos. Además, integra requisitos de cumplimiento con diversas normativas internacionales, como el GDPR, la CCPA de California y otros marcos regulatorios. También ofrece guías para aplicar principios fundamentales de privacidad, tales como la minimización de datos, el consentimiento informado, la retención limitada de la información y la transparencia en el manejo de los datos personales. Una ventaja importante de esta norma es que permite la certificación, lo que brinda a las organizaciones la posibilidad de demostrar formalmente su compromiso con la privacidad. Para pequeñas empresas y startups, la adopción de este estándar no solo mejora sus prácticas internas, sino que también fortalece la confianza con socios, clientes e inversores, especialmente cuando se busca operar en mercados internacionales. (Fuente: International Organization for Standardization (ISO), ISO/IEC 27701:2019 – Security techniques — Extension to ISO/IEC 27001 and ISO/IEC 27002 for privacy information management).

Otros marcos de referencia internacional

Además del GDPR europeo y la reforma nacional, existen otros marcos internacionales que pueden orientar a las organizaciones en el diseño de políticas de privacidad robustas y efectivas. El NIST Privacy Framework, desarrollado por el Instituto Nacional de Estándares y Tecnología de Estados Unidos, proporciona un enfoque flexible y adaptable que ayuda a las organizaciones a identificar, evaluar y gestionar los riesgos relacionados con la privacidad, independientemente de su tamaño o sector. Este marco se basa en prácticas de gestión de riesgos y se puede integrar con otros sistemas de seguridad y cumplimiento. Por otro lado, las Directrices de Privacidad de la Organización para la Cooperación y el Desarrollo Económico (OCDE) establecen principios fundamentales reconocidos globalmente para la protección de la información personal, tales como la limitación del propósito para el uso de datos, la calidad y exactitud de estos, la transparencia, la seguridad y la responsabilidad de los responsables del tratamiento. Estas directrices han sido base para muchas legislaciones nacionales en materia de privacidad. Además, el Convenio 108+ del Consejo de Europa es el primer tratado internacional jurídicamente vinculante dedicado a la protección de datos personales y la privacidad, promoviendo estándares modernos y actualizados para responder a los desafíos tecnológicos actuales. Argentina adhirió a este convenio en 2019, lo que implica un compromiso formal con la armonización de sus normativas de protección de datos a nivel regional y con la cooperación internacional para proteger la privacidad de sus ciudadanos frente a flujos transfronterizos de información personal. Estos marcos complementan y fortalecen los esfuerzos locales, contribuyendo a una gestión integral y alineada con las mejores prácticas globales en materia de privacidad.

Conceptos Técnicos de Anonimización

La anonimización es un proceso mediante el cual se eliminan o modifican datos personales de manera irreversible, impidiendo cualquier posibilidad de vincular la información con un individuo específico. Este método es utilizado para proteger la privacidad cuando los datos deben ser compartidos o analizados sin comprometer la identidad de las personas. Una vez que los datos han sido anonimizados correctamente, ya no están sujetos a las regulaciones de protección de datos, ya que dejan de considerarse información personal.

Ventajas y limitaciones de los algoritmos de anónimizacion

Figura 1 Ventajas y limitaciones de los algoritmos de anónimizacion

Riesgos Técnicos y de Reidentificación

Aunque los datos estén anonimizados, siempre existe la posibilidad de que alguien logre identificar a una persona si cruza esa información con otros datos disponibles. Por ejemplo, saber la edad, el sexo y la profesión puede ser suficiente para descubrir quién es alguien si esos datos coinciden con otros registros públicos.

Al eliminar o modificar partes de los datos para proteger la privacidad, puede perderse información importante. Esto puede hacer que los análisis no sean tan precisos o que los resultados no reflejen bien la realidad. Es difícil lograr que los datos sean seguros sin perder utilidad. Algunos tipos de datos, como los registros médicos, tienen relaciones internas que son importantes (por ejemplo, la conexión entre paciente y tratamiento). La anonimización puede romper ese contexto, lo que dificulta realizar estudios o análisis que dependan de esas relaciones. Algunas técnicas de anonimización, sobre todo las más avanzadas, necesitan mucho poder de procesamiento. Si se trabaja con grandes volúmenes de datos, el proceso puede tardar bastante y requerir muchos recursos, lo cual es un desafío para organizaciones con infraestructura limitada.

Por otro lado, la **pseudonimización** es una técnica que reemplaza identificadores directos, como nombres o números de documento, por alias o códigos que permiten reducir el riesgo de exposición de los datos. A diferencia de la anonimización, este proceso es reversible si se cuenta con la clave o los datos necesarios para volver a identificar a la persona. Un ejemplo de pseudonimización sería sustituir un nombre real como "Pablo Pérez" por un código genérico como "Usuario_XY12", permitiendo mantener cierto nivel de privacidad mientras se conserva la posibilidad de reidentificación en caso necesario. Ambas técnicas son herramientas clave en la protección de datos

personales y su aplicación depende del nivel de seguridad requerido y del marco normativo que rija el tratamiento de la información.

Técnicas de anonimización

Las técnicas de anonimización son fundamentales para proteger la identidad de las personas cuando se comparten o procesan datos. Estas estrategias buscan transformar la información de manera que no sea posible identificar a los individuos, manteniendo al mismo tiempo su utilidad para análisis o toma de decisiones. Su importancia radica en que permiten cumplir con regulaciones de privacidad, evitar filtraciones y reducir el riesgo de reidentificación en entornos cada vez más expuestos al intercambio masivo de datos.

k-anonymity:

La **k-anonymity** es una técnica de anonimización que busca garantizar que cada registro dentro de un conjunto de datos no pueda ser distinguido de al menos *k-1* registros adicionales, reduciendo así el riesgo de identificación individual. Para lograrlo, se aplican métodos como la generalización, donde se reemplazan valores específicos por rangos más amplios (por ejemplo, en lugar de registrar una edad exacta como "34 años", se agrupa dentro del intervalo "30-40 años"), y la supresión, que consiste en eliminar datos únicos o altamente identificables dentro del conjunto de información.

Si bien k-anonymity es una técnica útil para proteger la privacidad en bases de datos estructuradas, presenta ciertas limitaciones. Es vulnerable a ataques de **homogeneidad**, en los cuales, si todos los registros dentro de un grupo tienen el mismo valor en un atributo sensible, un atacante aún podría inferir información confidencial. También puede ser susceptible a ataques de **conocimiento previo (background knowledge)**, donde un atacante con información externa podría cruzar datos para reducir la efectividad de la anonimización. A pesar de estas debilidades, k-anonymity sigue siendo una estrategia ampliamente utilizada en la gestión de privacidad de datos.

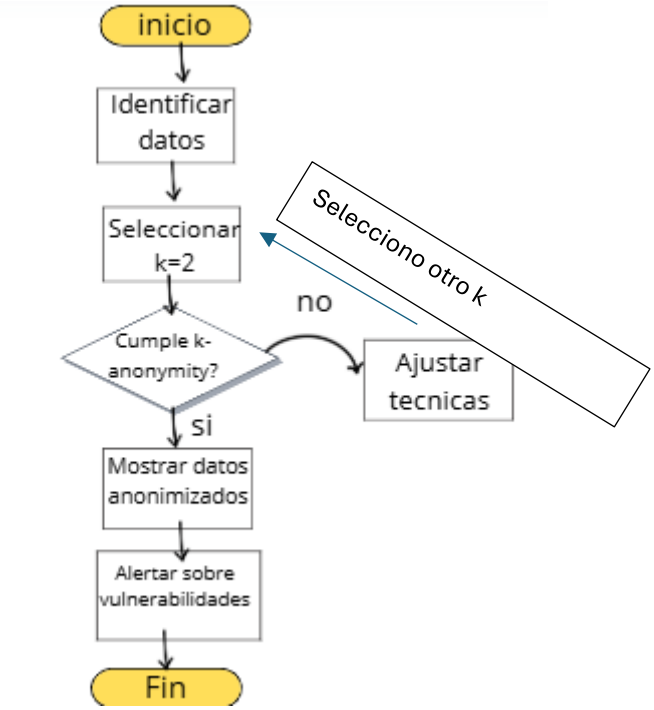

Figura 2 Diagrama de flujo k-anonymity

Ejemplo:

Suponiendo que, en una base de datos con información de empleados, incluyendo nombre, edad y código postal.

Nombre	**Edad**	**Código Postal**
Juan	34	1885
Luis	35	1886
Marcos	41	1886

Tabla 2 Ejemplo k-anonymity

Si alguien sabe que juan tiene 34 años o que su código postal es 1885 podria identificarlo fácilmente

Para aplicar *k-anonymity*, podríamos generalizar los datos:

Edad	**Código Postal**
30-40	188X
30-40	188X
40-50	188X

Tabla 3 Ejemplo k-anonymity.

Ahora, cada combinación de edad y código postal aparece en al menos *k* (por ejemplo, 2 o más) registros, dificultando la identificación de una persona específica.

Differential privacy:

La **differential privacy** es una técnica de anonimización que protege la identidad de los individuos al introducir **ruido estadístico** en los datos, lo que impide la identificación exacta de registros individuales sin afectar significativamente los resultados globales del análisis. En lugar de modificar directamente los datos, este método altera respuestas o valores numéricos de manera controlada, asegurando que la información agregada siga siendo útil sin comprometer la privacidad de los usuarios.

Empresas como **Apple y Google** implementan esta técnica para recopilar información de uso sin exponer datos personales, aplicándola en funciones como la predicción de texto y el análisis de tendencias de usuarios. Su principal ventaja es que logra un equilibrio entre privacidad y utilidad, permitiendo extraer conocimientos valiosos sin comprometer la seguridad de la información individual. Sin embargo, su correcta implementación requiere un ajuste preciso del nivel de ruido para garantizar tanto la protección de datos como la precisión del análisis.

Ejemplo:

Supongamos que queremos saber la edad promedio de una empresa de alimentos sin saber la edad exacta de nadie

Para hacerlo con privacidad diferencial, en lugar de usar las edades reales, aplicamos **ruido aleatorio** (por ejemplo, sumando o restando un número aleatorio a cada edad).

Edad real	**Edad con ruido**
34	36
28	30
40	37
30	31

Tabla 4 *Ejemplo Differential privacy*

Ahora, si alguien intenta identificar a una persona basándose en la edad, no podrá hacerlo con certeza, pero el **promedio** de todas las edades aún será cercano al valor real.

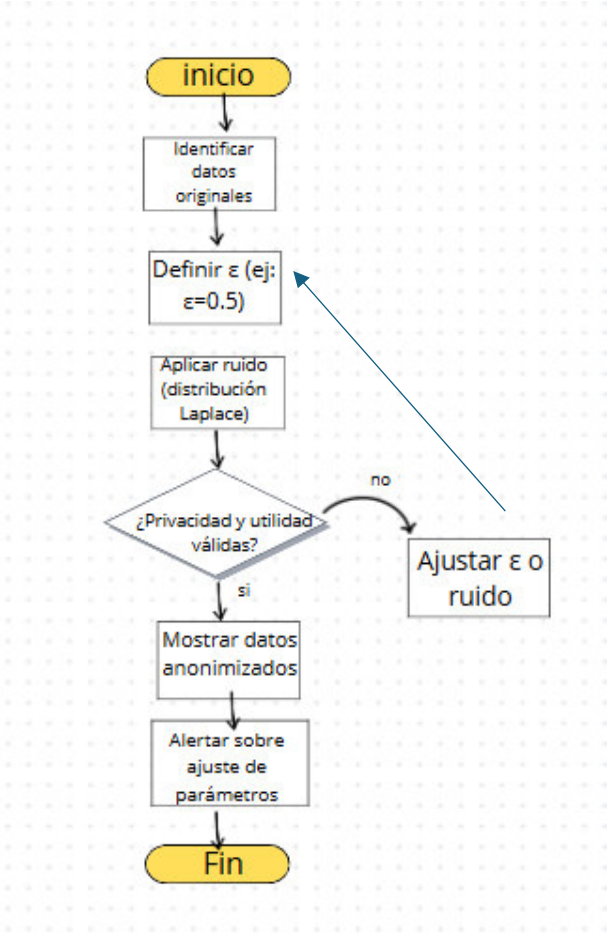

Figura 3 *Diagrama de flujo Differential privacy*

Tokenización:

La **tokenización** es una técnica de protección de datos que reemplaza información sensible con **tokens**, es decir, identificadores aleatorios que no tienen relación matemática con los datos originales y, por lo tanto, no pueden ser revertidos sin acceso a un sistema seguro de referencia. A diferencia del cifrado, donde los datos pueden ser desencriptados con una clave, en la tokenización los valores originales se almacenan en una base de datos segura y solo pueden ser recuperados por sistemas autorizados.

Un uso común de esta técnica se da en los **pagos en línea**, donde los números de tarjetas de crédito son reemplazados por tokens únicos, reduciendo el riesgo de fraude en caso de una filtración. Como el token en sí mismo no tiene valor fuera del sistema que lo generó, incluso si es interceptado, no puede ser utilizado sin acceso a la base de datos segura que lo asocia con los datos originales. Esta solución es ampliamente utilizada en la industria financiera y en plataformas de comercio electrónico para garantizar transacciones seguras sin comprometer la información real del usuario.

Un hospital quiere almacenar los datos de sus pacientes de manera segura. En lugar de guardar nombres y documentos de identidad, usa tokens.

Nombre	DNI	Token
Nicolas Ruiz	12345678	X9A5-BQ28
Micaela Pérez	91011121	L5U9-MI26

Tabla 5 Ejemplo toketizacion

En la base de datos pública, solo aparecen los tokens. Si un investigador necesita analizar datos de pacientes, verá solo los tokens y no podrá saber a quién pertenecen los registros sin acceso a la clave de descifrado.

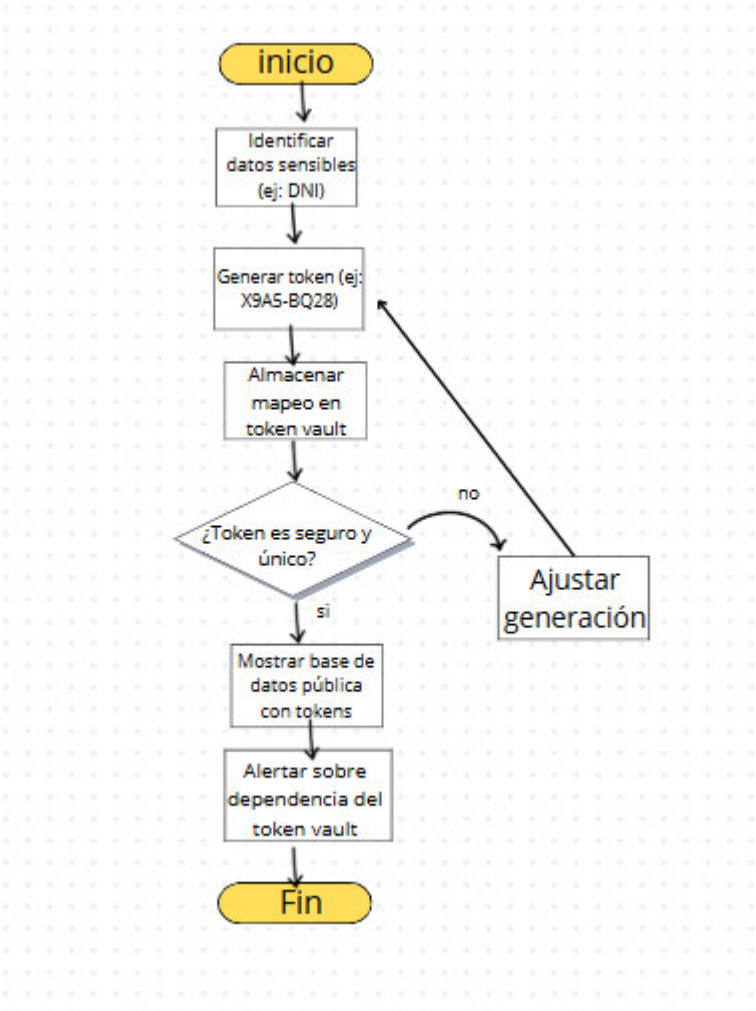

Figura 4 Diagrama de flujo toketizacion

Data Masking:

El *data masking* es una técnica de protección de datos que oculta parcial o totalmente la información sensible, permitiendo que se utilice sin revelar su contenido completo. Un ejemplo común es mostrar solo los últimos cuatro dígitos de un documento de identidad o una tarjeta de crédito, como "**--******-1234".

Esta técnica es ampliamente utilizada en entornos donde se necesita procesar o visualizar datos sin exponer información confidencial, como en bases de datos de pruebas, sistemas de atención al cliente o registros de transacciones. Al aplicar enmascaramiento, se minimiza el riesgo de filtraciones y accesos no autorizados, manteniendo la funcionalidad de los datos sin comprometer la seguridad del usuario.

Nombre real	Correo real	Nombre enmascarado	Correo enmascarado
Luis Días	Luis.dias@gmai.com	L*** D***	L***@gmail.com
Ana López	Ana.lopez@gmail.com	A*** L***	A***@gmail.com

Tabla 6 Ejemplo Data Masking

Ahora, los datos son inutilizables fuera del entorno real, pero aún pueden usarse para probar el sistema sin riesgos.

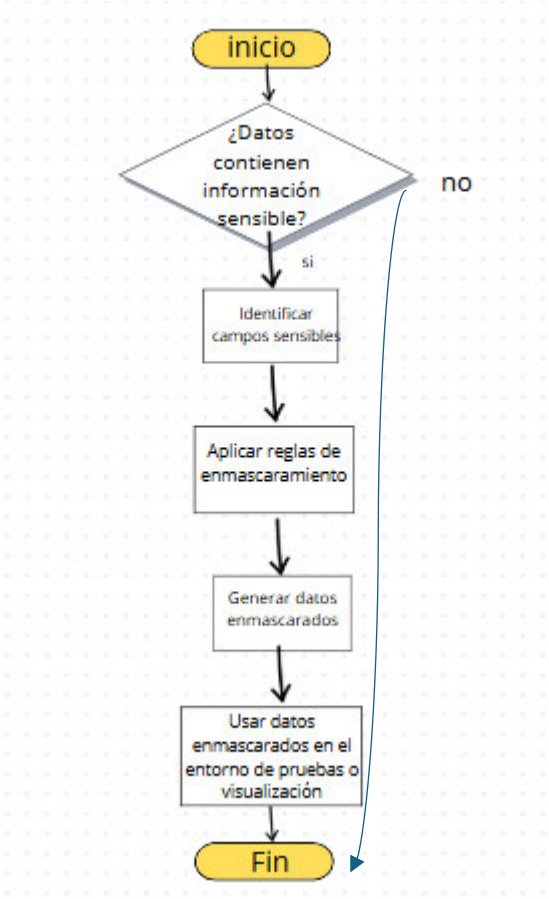

Figura 5 Diagrama de flujo Data Masking

Implementación de algoritmos de anonimizacion en escenarios del mundo real:

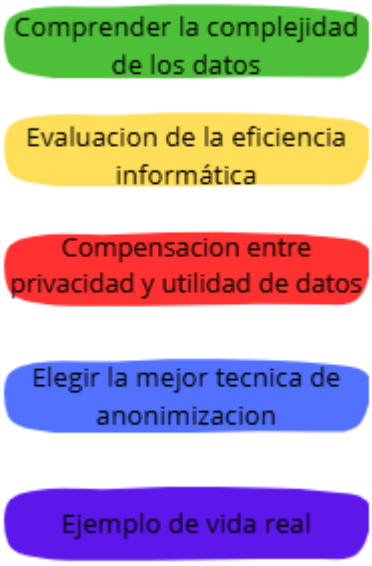

Figura 6 Implementación de algoritmos de anonimizacion

Para implementar algoritmos de anonimización de manera efectiva, primero es fundamental entender la complejidad de los datos que se van a proteger. Según su naturaleza —estructurados, no estructurados o semiestructurados— será necesario aplicar distintas técnicas: por ejemplo, en bases de datos tabulares se puede recurrir a la generalización o la supresión de atributos, mientras que en documentos de texto no estructurado se suelen usar métodos como la tokenización o la perturbación. Además, es importante evaluar la eficiencia informática de las técnicas elegidas, ya que trabajar con grandes volúmenes de datos puede ser muy costoso en términos de procesamiento. Técnicas como el k-anonimato son efectivas, pero pueden ser computacionalmente pesadas, por lo que alternativas como l-diversidad o t-cercanía a veces resultan más convenientes.

Otro aspecto clave es encontrar un equilibrio entre la privacidad y la utilidad de los datos: métodos como la privacidad diferencial permiten medir matemáticamente las garantías de protección sin sacrificar demasiado la calidad de la información. La elección de la técnica de anonimización dependerá siempre de los objetivos específicos; por ejemplo, si se necesita máxima protección frente a ataques de reidentificación, técnicas como el k-anonimato son apropiadas, pero si se prioriza la utilidad de los datos, la privacidad diferencial puede ser más adecuada. Para verlo en un caso práctico, imaginemos un hospital que quiere compartir registros médicos con investigadores. Podría aplicar k-anonimato agrupando atributos sensibles como edad, sexo y código postal, de modo que cada paciente no pueda ser identificado individualmente. Además, podría reforzar la protección aplicando privacidad diferencial a datos sensibles como diagnósticos o tratamientos, permitiendo así que los investigadores trabajen sobre información estadísticamente válida sin comprometer la identidad de los pacientes.

La protección de datos personales es un reto creciente, especialmente para pequeñas empresas que deben cumplir con normativas sin contar con grandes recursos. La anonimización se presenta como una estrategia clave para resguardar la privacidad, permitiendo usar datos sin comprometer la información sensible. En este capítulo, se analizan los principales desafíos y buenas prácticas para implementar técnicas de anonimización efectivas en entornos con limitaciones técnicas y legales.

Recursos limitados en pequeños proyectos

Los proyectos de pequeña escala, como startups en etapa temprana o iniciativas académicas, enfrentan grandes limitaciones tanto presupuestarias como técnicas. Esto impacta directamente en su capacidad para implementar medidas sólidas de protección de datos personales. Herramientas como Amnesia, Skyflow o Privitar ofrecen soluciones avanzadas de anonimización,

pero muchas de ellas implican licencias costosas, requerimientos de infraestructura tecnológica o personal especializado en privacidad de datos, algo que muchos de estos proyectos no pueden asumir.

En este contexto, es común que las startups opten por soluciones gratuitas, de código abierto o incluso desarrollen sus propios métodos rudimentarios de protección de datos. Aunque estas alternativas pueden ser útiles para comenzar, a menudo carecen de características clave como auditoría de seguridad, integración con normativas locales e internacionales o defensa frente a ataques de reidentificación. Esta situación puede exponer a las organizaciones a riesgos legales, sanciones regulatorias o pérdida de la confianza del usuario.

A esto se suma que muchos de estos proyectos no cuentan con un equipo legal o de compliance, lo que dificulta aún más la comprensión de las implicancias jurídicas del uso y tratamiento de datos personales. La Ley 25.326 en Argentina, por ejemplo, establece obligaciones claras sobre el consentimiento, la confidencialidad y los derechos del titular de los datos, pero su implementación práctica requiere una interpretación técnica y legal que no siempre está al alcance de las startups.

Por ello, es necesario pensar en soluciones específicas para este segmento. Las herramientas low-code o no-code orientadas a la protección de datos pueden ser una respuesta intermedia viable, ya que permiten a usuarios sin conocimientos técnicos profundos configurar mecanismos de anonimización mediante interfaces visuales. Sin embargo, aún queda camino por recorrer para asegurar que estas herramientas ofrezcan suficiente nivel de protección, sean interoperables con marcos legales locales, y cuenten con soporte técnico adecuado.

Desde una perspectiva de política pública, podrían impulsarse programas de acompañamiento técnico, subsidios para la adopción de tecnologías de protección de datos o incluso certificaciones gratuitas para herramientas de código abierto. Estas medidas contribuirían a nivelar el terreno para que los pequeños proyectos puedan cumplir con estándares adecuados de privacidad, incluso en contextos de recursos limitados.

Limitaciones de presupuesto

El costo de las soluciones avanzadas como Amnesia o Skyflow puede ser una barrera significativa para muchas startups argentinas, que enfrentan restricciones de presupuesto. Estas herramientas, aunque efectivas en la protección de datos personales y la anonimización, requieren una inversión considerable, tanto en licencias como en capacitación técnica. La falta de presupuesto en startups hace que muchas opten por herramientas gratuitas o de

bajo costo, que aunque pueden ofrecer funcionalidades básicas, suelen carecer de las capacidades avanzadas necesarias para asegurar la anonimización irreversible y protegerse contra los riesgos de reidentificación.

Faltante de expertise técnico

Además del condicionante económico, la escasez de conocimientos técnicos especializados representa una barrera significativa para muchas startups al momento de implementar mecanismos efectivos de protección de datos personales. Herramientas como Amnesia, Skyflow o incluso bibliotecas más accesibles como ARX, requieren cierto nivel de experiencia en privacidad diferencial, técnicas de anonimización y pseudonimización, así como en seguridad de la información.

Esta falta de expertise no se limita únicamente a la configuración técnica de las herramientas, sino que también se extiende al entendimiento de los marcos legales que rigen el tratamiento de datos personales. Por ejemplo, una correcta implementación de anonimización debe considerar los principios de la Ley 25.326 en Argentina o, en casos donde se apunten a mercados internacionales, cumplir con regulaciones como el GDPR. Sin un conocimiento adecuado de estas normativas, es común que las soluciones aplicadas sean insuficientes o incluso incumplan con requisitos legales clave, como el consentimiento informado, el derecho al olvido o la protección contra reidentificación.

Además, muchos equipos técnicos en startups están enfocados en el desarrollo de productos mínimos viables y en la escalabilidad del negocio, por lo que la privacidad suele quedar relegada como una prioridad secundaria. Esto agrava aún más el problema, ya que la adopción tardía de mecanismos de protección de datos puede ser más costosa y compleja, sobre todo si la arquitectura del sistema no fue diseñada desde el inicio con principios de privacidad por diseño.

Por ello, se vuelve crucial fomentar iniciativas de formación específica para equipos técnicos de startups, incluyendo talleres, cursos breves y documentación accesible sobre privacidad de datos. También sería valioso el impulso de políticas públicas que ofrezcan asistencia técnica gratuita o subsidiada, orientada a apoyar a emprendimientos tecnológicos en el cumplimiento de normativas de privacidad. De esta forma, se puede contribuir a cerrar la brecha entre el conocimiento técnico necesario y la capacidad real de las pequeñas empresas para proteger adecuadamente los datos personales de sus usuarios.

Prioridad por soluciones gratuitas

El hecho de que alrededor del 70% de las startups argentinas prioricen el uso de soluciones gratuitas para la protección de datos personales es un reflejo directo de las limitaciones presupuestarias que enfrentan muchas de estas iniciativas. En contextos de alta incertidumbre económica y acceso limitado a financiamiento, las herramientas gratuitas resultan atractivas por su bajo costo de adopción y su aparente facilidad de uso. Sin embargo, esta elección puede acarrear importantes riesgos.

Las herramientas gratuitas de anonimización o protección de datos suelen carecer de soporte técnico, actualizaciones constantes o certificaciones de seguridad. Además, muchas de ellas no incluyen algoritmos avanzados ni configuraciones personalizables que permitan una adaptación precisa a las necesidades de cada proyecto. Esta falta de robustez técnica puede facilitar ataques de reidentificación, especialmente cuando los datos anonimizados se combinan con bases externas o metadatos que no fueron adecuadamente protegidos.

El uso de estas herramientas, aunque comprensible, puede generar una falsa sensación de seguridad. Las startups pueden creer que están cumpliendo con las normas de protección de datos, cuando en realidad están dejando expuestos datos sensibles debido a deficiencias en las medidas implementadas. Esto no solo pone en riesgo la privacidad de los usuarios, sino que también puede derivar en consecuencias legales o pérdida de reputación en caso de incidentes de seguridad.

Para abordar esta problemática, sería recomendable que los organismos públicos o instituciones académicas desarrollen y ofrezcan herramientas gratuitas de anonimización con estándares más altos de seguridad, acompañadas de guías prácticas y capacitaciones. De esta manera, se podrían reducir las brechas entre la necesidad de proteger datos y la capacidad económica real de muchas startups. Asimismo, fomentar colaboraciones con comunidades de código abierto podría contribuir al desarrollo de soluciones más seguras y accesibles, sin comprometer la privacidad ni la integridad de los datos.

Consecuencias a largo plazo

La falta de inversión en soluciones robustas y bien implementadas puede tener consecuencias a largo plazo. Las startups que eligen herramientas gratuitas o de bajo costo pueden enfrentar incumplimientos normativos, lo que podría llevar a sanciones legales y pérdida de confianza de los usuarios. Además, la exposición de datos personales a través de vulnerabilidades técnicas puede dañar la reputación de la empresa y disminuir su competitividad en el mercado. Aunque las soluciones de bajo costo pueden ser atractivas a corto

plazo, no proporcionar una protección adecuada podría resultar en mayores costos y riesgos en el futuro.

En resumen, la falta de presupuesto y expertise técnico en pequeños proyectos afecta directamente la implementación de herramientas avanzadas de protección de datos, como Amnesia y Skyflow. Al optar por soluciones gratuitas o de bajo costo, muchas startups aumentan su exposición al riesgo de reidentificación, lo que puede comprometer la privacidad de los usuarios y dificultar el cumplimiento de las normativas vigentes.

Tensión entre privacidad y utilidad

Encontrar un equilibrio adecuado entre la privacidad de los datos y su utilidad es uno de los desafíos más complejos en la protección de datos personales. Las técnicas diseñadas para salvaguardar la identidad de los individuos, como **differential privacy**, funcionan introduciendo ruido o modificando los datos de manera que sea difícil identificar a una persona específica dentro del conjunto.

Aunque este enfoque mejora la privacidad, también afecta la calidad y precisión de los datos. Esta alteración puede reducir la eficacia de los análisis estadísticos y de los modelos de machine learning que dependen de datos precisos para generar resultados confiables. En muchas aplicaciones, especialmente en investigación o desarrollo de tecnologías avanzadas, la utilidad de los datos puede verse comprometida, lo que limita su valor para la toma de decisiones o la generación de conocimiento.

Este conflicto genera una tensión inherente: aumentar la privacidad puede reducir la utilidad y viceversa. Por ello, las organizaciones y los desarrolladores deben evaluar cuidadosamente sus objetivos y prioridades para decidir hasta qué punto están dispuestos a sacrificar precisión por protección de datos. En algunos casos, puede ser necesario aceptar una reducción en la calidad de los resultados para garantizar que la privacidad no se vea comprometida.

Para afrontar esta tensión, se están explorando soluciones que buscan optimizar el balance entre ambos aspectos, como el desarrollo de técnicas híbridas que reduzcan la cantidad de ruido o el uso de métodos adaptativos que ajusten la protección según el tipo de dato o contexto. Además, la transparencia en el proceso y la comunicación clara sobre las limitaciones de los datos modificados son cruciales para que los usuarios finales comprendan las posibles implicaciones.

Una síntesis orientativa sobre la selección de técnicas según el tipo de dato puede encontrarse en el Anexo II.

Impacto en la precisión

El uso de técnicas como differential privacy representa un avance significativo en la protección de datos personales, ya que introduce ruido aleatorio para evitar la reidentificación de individuos en conjuntos de datos. Sin embargo, esta protección tiene un costo en términos de precisión. La introducción de ruido puede alterar los valores originales, lo que afecta la exactitud de los análisis y modelos que dependen de datos precisos.

Este impacto es especialmente relevante en contextos donde la exactitud es fundamental, como en proyectos académicos, investigaciones científicas o modelos predictivos basados en machine learning. Por ejemplo, una reducción del 15% en la precisión de los resultados puede ser suficiente para invalidar conclusiones o hacer que ciertos enfoques analíticos sean inviables.

Esta situación plantea un dilema para los investigadores y desarrolladores: deben balancear la necesidad de proteger la privacidad con la necesidad de mantener la calidad y utilidad de los datos. En muchos casos, el compromiso puede resultar en limitaciones prácticas que frenan la adopción de técnicas avanzadas de privacidad.

Para mitigar estos efectos, es fundamental continuar investigando métodos que minimicen la pérdida de precisión, como ajustar la cantidad y tipo de ruido aplicado, o combinar differential privacy con otras técnicas complementarias que potencien tanto la privacidad como la utilidad. Además, la transparencia en cómo se aplican estas técnicas y sus limitaciones ayuda a que los usuarios y responsables de proyectos tomen decisiones informadas.

Compromisos entre privacidad y funcionalidad

Este compromiso también implica decisiones estratégicas sobre qué nivel de privacidad es aceptable según el contexto del proyecto y los riesgos involucrados. En algunos casos, puede ser preferible sacrificar cierta precisión para proteger la confidencialidad de los datos, especialmente cuando se manejan datos sensibles o personales. Sin embargo, en otros escenarios, la pérdida de funcionalidad puede hacer que los resultados sean menos útiles o incluso inválidos para la toma de decisiones.

Además, el diseño de algoritmos debe considerar estos trade-offs desde etapas tempranas, buscando métodos que minimicen la pérdida de utilidad mientras garantizan un nivel adecuado de privacidad. La investigación en técnicas de privacidad diferencial adaptativa y en algoritmos robustos que toleren el ruido es crucial para avanzar en esta dirección.

Por último, la transparencia con los usuarios y stakeholders acerca de las limitaciones y compromisos entre privacidad y funcionalidad es fundamental para

mantener la confianza y establecer expectativas realistas sobre el desempeño y la seguridad de los sistemas.

Desafíos en la adopción en proyectos académicos

Otro aspecto relevante es la necesidad de capacitación y formación en privacidad de datos dentro de las instituciones académicas. Muchos investigadores no están familiarizados con los principios y técnicas de anonimización, lo que dificulta su aplicación adecuada. Por ello, es fundamental promover programas educativos y talleres que integren conocimientos de protección de datos en los currículos de ciencias sociales, informática, y otras disciplinas que manejen información sensible.

Además, la colaboración interdisciplinaria entre expertos en privacidad, estadística y el área temática del estudio puede mejorar la implementación de técnicas que protejan la privacidad sin sacrificar la calidad de los resultados. Esta colaboración también puede ayudar a desarrollar nuevos métodos que respondan a las necesidades específicas de cada tipo de investigación.

Finalmente, los organismos financiadores y reguladores pueden jugar un papel crucial incentivando el uso de técnicas de privacidad mediante la creación de políticas claras y ofreciendo recursos para su adopción. Esto ayudaría a superar barreras económicas y técnicas, impulsando una cultura de responsabilidad y ética en el manejo de datos en el ámbito académico.

Soluciones posibles

Para abordar esta tensión, se están desarrollando técnicas híbridas que intentan optimizar tanto la privacidad como la utilidad. Por ejemplo, el uso de privacidad diferencial en conjunto con técnicas de optimización en machine learning puede ayudar a reducir la pérdida de precisión mientras se mantiene un nivel adecuado de protección. Sin embargo, esto sigue siendo un campo en desarrollo, y los proyectos deben evaluar cuidadosamente qué nivel de privacidad es necesario y qué impacto tendrá en la funcionalidad del análisis.

Herramientas subestimadas

A pesar de su bajo costo y su capacidad para mitigar riesgos legales, herramientas como Faker, que generan datos sintéticos, no son ampliamente adoptadas en muchos proyectos, lo que genera una aparente contradicción. Estas herramientas tienen el potencial de proporcionar una solución efectiva para proteger la privacidad de los usuarios sin comprometer la funcionalidad de

los datos, pero su uso sigue siendo limitado, especialmente en entornos más pequeños o proyectos que no cuentan con el conocimiento técnico adecuado.

Bajo costo y alta efectividad

La generación de datos sintéticos mediante herramientas como Faker ofrece una solución económica y eficiente para superar las limitaciones de acceso a datos reales en proyectos de análisis, desarrollo y pruebas. Al replicar patrones y distribuciones estadísticas sin revelar información sensible, estas herramientas facilitan la experimentación y el desarrollo ágil de aplicaciones, especialmente en entornos con restricciones estrictas de privacidad.

Además, los datos sintéticos permiten a los equipos técnicos entrenar modelos de inteligencia artificial, validar algoritmos y realizar simulaciones sin comprometer la confidencialidad de los usuarios. Esto resulta crucial en sectores donde la protección de datos es especialmente sensible, como el sector salud, financiero o educativo.

Otra ventaja importante es la escalabilidad: al no depender de datos reales, es posible generar grandes volúmenes de datos bajo demanda, ajustando la complejidad y variedad según las necesidades específicas del proyecto. Esto contribuye a acelerar los ciclos de desarrollo y mejora la calidad de las soluciones finales.

No obstante, es fundamental tener en cuenta que los datos sintéticos, aunque seguros, deben ser generados con metodologías rigurosas para asegurar que sean representativos y útiles. La calidad de estos datos influye directamente en la validez de los resultados obtenidos, por lo que es necesario diseñar procesos que reflejen fielmente las características del universo real que se pretende estudiar.

En resumen, el uso de datos sintéticos representa una estrategia innovadora y accesible para startups y pequeños proyectos, que equilibra la necesidad de proteger la privacidad con la exigencia de contar con datos fiables para el desarrollo tecnológico

Desconocimiento y falta de confianza

A pesar de sus ventajas, el uso de datos sintéticos sigue siendo limitado debido al desconocimiento sobre su aplicación y efectividad. Muchos proyectos, especialmente aquellos con recursos limitados o sin personal técnico especializado, no están familiarizados con cómo generar datos sintéticos de manera eficaz. A menudo, los equipos prefieren recurrir a soluciones más

convencionales, aunque estas impliquen mayores riesgos en términos de privacidad y cumplimiento normativo.

Además, existe una falta de confianza en la calidad de los datos sintéticos. Algunos desarrolladores y analistas pueden cuestionar si los datos sintéticos generados por herramientas como Faker son realmente representativos de los datos reales en términos de patrones y relaciones subyacentes, lo que puede generar reticencia a su adopción. El miedo a que estos datos no sean adecuados para entrenar modelos de machine learning o realizar análisis precisos también limita su implementación, a pesar de los avances en la generación de datos sintéticos de alta calidad.

Barreras culturales y organizativas

Otra razón por la cual herramientas como Faker no son ampliamente adoptadas es la resistencia al cambio dentro de las organizaciones. Muchas veces, las empresas y equipos de desarrollo están acostumbrados a trabajar con datos reales, y cambiar a datos sintéticos puede parecer un desafío adicional. Las culturales organizacionales y la falta de incentivos para explorar nuevas soluciones también juegan un papel importante en la subestimación de estas herramientas.

Legislación vs. innovación tecnológica

La relación entre legislación e innovación tecnológica suele ser compleja, ya que las leyes tienden a ser estáticas mientras que la tecnología evoluciona de forma rápida y disruptiva. Esto crea un desfase temporal que dificulta que los marcos regulatorios respondan con agilidad a los nuevos retos y escenarios que surgen con el desarrollo tecnológico. En el caso específico de la protección de datos personales, este desfase puede traducirse en incertidumbre jurídica para quienes deben cumplir con las normativas, así como en brechas que pueden ser explotadas de manera indebida.

La Ley 25.326, aprobada en un momento en que tecnologías como la IA y el IoT eran apenas incipientes o inexistentes en el ámbito cotidiano, no contempla aspectos críticos de estos avances, tales como el tratamiento automatizado de datos, la recolección masiva a través de sensores conectados o el uso de algoritmos para la toma de decisiones. Esta carencia normativa genera una tensión entre la necesidad de innovación y la obligación de proteger derechos fundamentales como la privacidad y la seguridad de la información.

Para las startups, este contexto puede representar tanto un riesgo como una oportunidad. Por un lado, la falta de claridad legal puede exponerlas a sanciones o a la obligación de realizar costosas adecuaciones posteriores. Por otro lado, la ausencia de regulaciones específicas abre espacio para la creación

de soluciones innovadoras que, bien diseñadas, pueden posicionarlas como líderes en un mercado emergente de tecnologías responsables y confiables.

Las empresas más grandes, por su parte, enfrentan el desafío de ajustar sus políticas internas y prácticas operativas a estándares que, aunque no estén explícitos en la ley local, se derivan de regulaciones internacionales o de las expectativas del mercado global. Esto implica una inversión constante en actualización tecnológica, capacitación y auditorías para garantizar la conformidad y evitar daños reputacionales.

En definitiva, es imprescindible promover un diálogo activo entre legisladores, sector privado, expertos en tecnología y sociedad civil, para construir marcos regulatorios flexibles y dinámicos que acompañen el ritmo de la innovación sin sacrificar la protección de los derechos de los usuarios. Solo así se podrá lograr un equilibrio sostenible que impulse el desarrollo tecnológico responsable y seguro.

Desactualización de las leyes frente a nuevas tecnologías

La Ley 25.326, diseñada en un contexto tecnológico muy diferente, no aborda directamente las complejidades que surgen con la recolección y procesamiento masivo de datos generados por dispositivos conectados y sistemas automatizados. En la actualidad, la cantidad y variedad de datos personales que se recopilan diariamente han aumentado exponencialmente, lo que exige una revisión profunda de las normativas para garantizar que los derechos de los individuos estén adecuadamente protegidos.

Por ejemplo, los dispositivos IoT instalados en hogares, vehículos o ciudades inteligentes recogen datos en tiempo real sobre hábitos, ubicaciones y preferencias de los usuarios. Esta recolección continua y a gran escala no solo plantea riesgos en términos de privacidad, sino que también demanda nuevos mecanismos de control y transparencia que no están contemplados en la ley actual. Asimismo, la automatización mediante algoritmos de IA puede tomar decisiones que afectan directamente a los usuarios, desde recomendaciones hasta autorizaciones, sin que exista una regulación clara sobre la responsabilidad y la supervisión de estos procesos.

La ausencia de disposiciones específicas en la ley genera vacíos legales que dificultan la implementación de buenas prácticas por parte de startups y pequeñas empresas tecnológicas, que muchas veces carecen de recursos para asesorarse legalmente o desarrollar soluciones robustas de protección de datos. Esta situación puede llevar a incumplimientos involuntarios y, en casos más graves, a sanciones que impactan negativamente en su desarrollo y crecimiento.

Por ello, es fundamental impulsar una actualización normativa que contemple las particularidades de estas tecnologías emergentes. Dicha actualización debería incluir definiciones claras sobre conceptos como datos biométricos, perfilado automatizado y consentimiento informado en entornos digitales, así como establecer obligaciones concretas para la transparencia, seguridad y rendición de cuentas en el uso de IA e IoT.

Además, la integración de estándares internacionales, como los planteados en el Reglamento General de Protección de Datos (GDPR) de la Unión Europea, puede servir como referencia para fortalecer el marco legal local, asegurando una protección más robusta y alineada con las mejores prácticas globales.

Vacíos legales para startups

Esta incertidumbre legal dificulta que las startups y pequeños proyectos adopten plenamente tecnologías como la inteligencia artificial (IA) y el Internet de las cosas (IoT) sin temor a incurrir en incumplimientos involuntarios. La rápida evolución de estas tecnologías supera el ritmo de actualización de las normativas existentes, lo que genera una brecha entre innovación y regulación. Además, las implicancias éticas y de privacidad que estas tecnologías conllevan requieren un análisis profundo, ya que el tratamiento de datos personales puede incluir información sensible o contextual, lo que aumenta el riesgo de vulneraciones.

Por ejemplo, los chatbots que recopilan datos para personalizar la experiencia del usuario deben garantizar transparencia y consentimiento claro, aspectos que no siempre están definidos en la legislación vigente. De igual forma, los dispositivos IoT en hogares inteligentes recogen continuamente grandes volúmenes de datos, lo que plantea interrogantes sobre quién tiene acceso, cómo se almacenan y durante cuánto tiempo se conservan dichos datos.

Frente a este escenario, es imprescindible que los legisladores trabajen en actualizar y complementar las leyes de protección de datos, incorporando definiciones específicas y pautas claras para el manejo de tecnologías emergentes. Esto incluiría lineamientos para el consentimiento dinámico, mecanismos de auditoría para algoritmos de IA, y estándares mínimos de seguridad para dispositivos conectados.

Paralelamente, las startups pueden beneficiarse de adoptar prácticas proactivas de privacidad desde el diseño (privacy by design), implementando protocolos internos que superen los requisitos legales actuales y que garanticen la protección de los datos desde la recolección hasta el procesamiento y almacenamiento. Esta actitud no solo reduce riesgos legales, sino que también fortalece la confianza del usuario y mejora la reputación de la empresa en un mercado cada vez más sensible a la privacidad.

La brecha entre innovación y regulación

Por otro lado, la innovación tecnológica avanza a un ritmo mucho más rápido que la capacidad de los legisladores para adaptar las leyes existentes. Las tecnologías como IA, blockchain y IoT evolucionan constantemente, lo que hace que las leyes vigentes queden rápidamente desactualizadas. Los marcos regulatorios existentes no fueron diseñados para abordar de manera efectiva los nuevos retos que surgen con estas tecnologías, como la recolección masiva de datos en tiempo real, el análisis predictivo, o la automatización de decisiones importantes sin intervención humana. Esta desalineación crea una brecha en la protección legal de los individuos y pone en evidencia la necesidad urgente de una actualización normativa.

Desafíos en la implementación de nuevas leyes

La actualización de leyes en relación con tecnologías emergentes es un proceso complicado. A medida que las startups y las grandes corporaciones intentan innovar, los legisladores deben considerar no solo el impacto tecnológico, sino también las implicaciones sociales y éticas de estas nuevas herramientas. Mientras tanto, las empresas deben adaptarse a un marco legal que no siempre ofrece una orientación clara. Además, la creación de nuevas leyes debe equilibrar la protección de la privacidad con la facilitación de la innovación, lo que no siempre es sencillo de lograr.

Impacto en Argentina

La posible reforma de la Ley 25.326 en Argentina, especialmente con la inclusión de nuevos derechos y regulaciones, como el "derecho al olvido", podría tener un impacto significativo en los pequeños proyectos y startups. Este derecho, que permite a los individuos solicitar la eliminación de sus datos personales de sistemas de almacenamiento, exige que las empresas implementen sistemas de borrado irreversible para cumplir con la normativa. Para muchos proyectos pequeños, esta reforma puede representar tanto desafíos como oportunidades.

Desafíos para los pequeños proyectos

La implementación de un sistema de borrado irreversible para cumplir con el "derecho al olvido" tiene un costo asociado que puede ser difícil de asumir para los proyectos pequeños o startups con recursos limitados. La infraestructura técnica necesaria para garantizar que los datos sean eliminados de forma segura y definitiva requiere inversiones en sistemas de gestión de datos más complejos

y costosos, así como en entrenamiento de personal para asegurar que el proceso se lleve a cabo de manera eficiente.

Para muchos pequeños emprendedores y startups, estas inversiones adicionales podrían representar un obstáculo considerable, especialmente si ya están trabajando con presupuestos limitados y priorizando el desarrollo de productos o servicios. La adaptación a los nuevos requisitos legales podría implicar una reestructuración significativa de las operaciones tecnológicas, lo que podría ser una carga adicional para empresas que ya luchan por mantenerse competitivas en un mercado dinámico.

Oportunidades de adaptación

A pesar de los desafíos, la reforma de la Ley 25.326 también podría ofrecer oportunidades para los pequeños proyectos que se adapten con éxito a los nuevos requisitos. En primer lugar, el cumplimiento de la normativa puede proporcionar una ventaja competitiva en un mercado donde los usuarios valoran la privacidad de sus datos. Las empresas que implementen soluciones robustas para garantizar el derecho al olvido podrían ganar la confianza de los consumidores, lo que, a largo plazo, podría traducirse en un aumento de clientes y una mejor reputación.

Además, la implementación de tecnologías de privacidad como sistemas de borrado seguro o anonimización de datos puede hacer que las startups se alineen mejor con las normativas internacionales de protección de datos, como el GDPR en Europa, lo que facilitaría su expansión global y mejoraría sus oportunidades de negocio en mercados internacionales.

Repercusiones en el ecosistema local

A nivel local, la reforma de la Ley 25.326 podría generar un efecto dominó en el ecosistema de tecnologías emergentes en Argentina. Si bien los pequeños proyectos enfrentan retos para cumplir con las nuevas exigencias, la reforma podría incentivar la creación de nuevas soluciones tecnológicas para gestion de datos personales, abriendo espacio para startups dedicadas a la creación de herramientas de protección de datos o de borrado seguro. Esto podría fomentar la innovación en el sector de ciberseguridad y privacidad, promoviendo un entorno más seguro para el tratamiento de datos personales.

Implicaciones Globales y Locales: Lecciones para otros países

Argentina ha vivido un proceso continuo de adaptación en cuanto a la protección de datos personales, y las lecciones aprendidas de su experiencia

pueden ser valiosas para otros países, especialmente aquellos con normativas más laxas o en desarrollo. Los desafíos y soluciones implementadas en Argentina ofrecen lecciones clave que podrían ser aplicadas a contextos en América Latina o en países que aún no cuentan con un marco legal robusto como el GDPR en Europa.

Combina técnicas de privacidad accesibles y efectivas

Una de las principales lecciones es la combinación de tecnologías de protección de datos como la tokenización y el data masking (enmascarado de datos), que pueden implementarse de manera efectiva incluso en países sin regulaciones estrictas como el GDPR. Estas herramientas permiten proteger la privacidad de los datos personales sin la necesidad de una legislación estricta, y son relativamente fáciles de implementar para pequeñas y medianas empresas.

Por ejemplo, en Argentina, empresas tecnológicas que operan en sectores de bajo costo como startups han adoptado tokenización para garantizar que los datos sensibles sean reemplazados por valores irreversibles, y data masking para mostrar solo una parte de la información real en sus sistemas. Esta combinación no solo reduce los riesgos de exposición de datos, sino que también cumple con las expectativas de los usuarios en cuanto a privacidad. Estos enfoques pueden ser fácilmente replicados en países latinoamericanos que no cuenten con normas estrictas como el GDPR, ofreciendo un equilibrio entre la protección de datos y la viabilidad económica de las soluciones.

Enfoque en la educación y sensibilización

Otra lección clave para países con normativas laxas es la importancia de la educación y la sensibilización sobre la privacidad de datos, tanto a nivel empresarial como en la ciudadanía. En Argentina, la concientización sobre la protección de datos personales ha sido un desafío debido a la falta de regulaciones claras, lo que ha llevado a que muchas startups operen sin el conocimiento adecuado de cómo manejar la información sensible. Sin embargo, a medida que la información y la formación sobre ciberseguridad y privacidad se han fortalecido, las empresas han mejorado significativamente sus prácticas.

Para países sin marcos regulatorios estrictos, invertir en programas de educación y sensibilización en torno a la protección de datos puede ser un primer paso fundamental. Este enfoque no solo fomenta un cambio cultural en cuanto al manejo de la información, sino que también capacita a las empresas para que adopten prácticas responsables de forma proactiva, sin esperar a que las leyes impongan cambios.

Creación de normativas locales adaptadas

A pesar de la falta de una ley tan robusta como el GDPR, Argentina ha avanzado en la implementación de reformas dentro de su Ley 25.326 y otras normativas locales que abordan el tratamiento de datos personales. Estas modificaciones se han orientado a adaptarse a las necesidades del entorno local y a los nuevos desafíos impuestos por las tecnologías emergentes como la IA y el Internet de las Cosas. Para países latinoamericanos sin una regulación comparable, esto resalta la importancia de crear normativas locales que no necesariamente copien el modelo europeo, sino que se adapten a las realidades socioeconómicas y tecnológicas del contexto local.

Colaboración y enfoque multilateral

Finalmente, otro aprendizaje clave es la necesidad de colaboración entre gobiernos, empresas y organizaciones internacionales para desarrollar y aplicar políticas comunes en torno a la protección de datos. En el caso de Argentina, aunque se han hecho esfuerzos aislados para mejorar la protección de la privacidad, aún persisten desafíos relacionados con la implementación de medidas en todo el territorio. Para los países que aún no tienen normativas estrictas, la cooperación multilateral puede garantizar que las políticas de privacidad no solo sean más efectivas, sino que también se alineen con las mejores prácticas internacionales.

Limitaciones del Estudio: Falta de Datos Públicos

Una de las principales limitaciones de este estudio ha sido la falta de acceso a datos públicos o bases de datos reales, particularmente en el contexto de las startups. La naturaleza confidencial y sensible de los datos manejados por estas empresas ha dificultado la posibilidad de realizar pruebas o análisis detallados con datos reales. Muchas startups optan por no compartir su información o por mantenerla dentro de entornos altamente protegidos para evitar posibles filtraciones de datos personales o riesgos asociados a la seguridad cibernética.

Desafíos de obtener datos reales

Debido a la *confidencialidad* inherente al funcionamiento de muchas *startups*, se ha hecho complicado acceder a bases de datos reales que permitan realizar un análisis más profundo sobre cómo las herramientas de anonimización y protección de datos operan en escenarios del mundo real. La información confidencial que manejan las startups generalmente está sujeta a políticas de

privacidad y acuerdos de no divulgación, lo que hace que el acceso a datos de prueba sea limitado. Esta falta de acceso a datos públicos limita la posibilidad de validar ciertos modelos o herramientas en situaciones reales, y puede generar incertidumbre sobre la efectividad de las soluciones propuestas.

Dependencia de datos sintéticos

Para sortear esta limitación, se recurrió en muchos casos a la generación de datos sintéticos, lo cual, aunque útil, no siempre refleja con precisión los patrones o comportamientos de los datos reales. Si bien los datos sintéticos pueden ser una solución viable para realizar pruebas, su validez es limitada en cuanto a la representación de datos reales, lo que podría afectar la exactitud de los resultados y las conclusiones del estudio. El uso exclusivo de datos sintéticos también puede no tener en cuenta todos los matices que existen en los datos reales, lo que lleva a una representación imperfecta de los posibles riesgos o beneficios de las herramientas de anonimización en el contexto de startups.

Implicaciones para la investigación

La falta de acceso a datos públicos reales también restringe la capacidad de realizar estudios de gran escala que involucren múltiples industrias o sectores. La recopilación de datos en sectores como salud, finanzas o educación, que requieren estrictos niveles de seguridad y privacidad, representa una barrera significativa para obtener información que permita realizar un análisis más exhaustivo sobre el impacto y la efectividad de las técnicas de protección de datos en diversos entornos. Esto limita la generalización de los hallazgos y su aplicabilidad a diferentes contextos.

Riesgos Técnicos

Evaluación de la eficiencia de las técnicas de anonimización

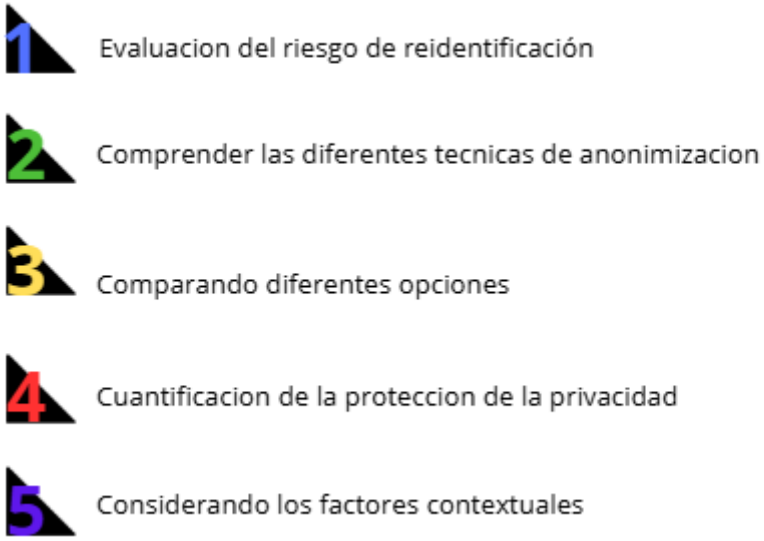

Figura 7 Evaluación de la eficiencia de las técnicas de anonimización

1 Evaluación del riesgo de reidentificación:

Un paso fundamental en la evaluación de técnicas de anonimización es analizar el riesgo de reidentificación, es decir, la posibilidad de volver a asociar datos anónimos con las personas reales a las que pertenecen. Este riesgo puede variar dependiendo de la sensibilidad y el tipo de información manejada. Por ejemplo, un conjunto de registros médicos puede tener un riesgo de reidentificación mayor que un listado anónimo de edades. Evaluar este riesgo permite elegir la técnica de anonimización más adecuada y aplicar medidas adicionales si fuera necesario, como aumentar el nivel de generalización o aplicar combinaciones de técnicas.

2. Comprender las diferentes técnicas de anonimización:

Existen diversas técnicas de anonimización, cada una con características y limitaciones propias. La **generalización** reemplaza datos específicos por rangos más amplios (por ejemplo, convertir una edad de 27 años en "25-30 años"). La **supresión** elimina atributos sensibles del conjunto de datos, mientras que la **perturbación** agrega ruido o modifica los datos ligeramente para dificultar la identificación. Otra técnica común es el **k-anonimato**, que asegura que cada registro sea indistinguible de al menos k-1 registros más. Además, técnicas más avanzadas como la **l-diversidad** y la **t-cercanía** refinan el k-anonimato para protegerse mejor contra ataques de inferencia.

3. Comparando diferentes opciones:

La comparación entre técnicas de anonimización es crucial para evaluar cuál resulta más efectiva en un escenario concreto. Por ejemplo, frente a un conjunto de datos con información de edad e ingresos, podría compararse el impacto de la generalización (reemplazar edades exactas por rangos) frente a la supresión (eliminar directamente el atributo de ingresos). Se debe analizar cómo cada técnica afecta la utilidad de los datos y el nivel de protección alcanzado, considerando tanto el riesgo de reidentificación como la capacidad de seguir extrayendo valor analítico de los datos anonimizados.

4. Cuantificación de la protección de la privacidad:

Para medir la eficacia de la anonimización, es necesario usar métricas objetivas que cuantifiquen tanto la privacidad como la utilidad de los datos. Entre las métricas más utilizadas están:

- **Pérdida de información**, que mide cuánto se deteriora el valor analítico de los datos.
- **Entropía**, que evalúa el grado de incertidumbre que presenta el conjunto de datos (mayor entropía significa menor riesgo de identificación).
- **Unicidad**, que indica cuántos registros pueden distinguirse fácilmente de los demás (mayor unicidad implica mayor riesgo de reidentificación). Medir estos factores ayuda a comparar de manera más precisa distintas técnicas de anonimización.

5. Considerando los factores contextuales:

Finalmente, la elección de una técnica de anonimización no puede hacerse sin considerar el contexto específico. Factores como el propósito de la recopilación de datos, las regulaciones legales (por ejemplo, GDPR en Europa o HIPAA en Estados Unidos), las expectativas de privacidad de las personas involucradas, y los riesgos específicos del sector, influyen directamente en qué métodos son más apropiados. Por ejemplo, en el ámbito de la salud o en servicios financieros, las exigencias legales y éticas son muy estrictas, por lo que se requieren técnicas más robustas. Tener en cuenta estos factores ayuda a garantizar que los datos sigan siendo útiles sin comprometer la privacidad.

La **reidentificación** es el proceso mediante el cual datos previamente anonimizados pueden ser vinculados nuevamente a individuos al cruzarlos con

información de fuentes externas. Este tipo de ataque expone una vulnerabilidad en los métodos de anonimización cuando los datos, aunque despojados de identificadores directos, conservan patrones que pueden correlacionarse con otros conjuntos de datos públicos o filtrados.

Un caso emblemático ocurrió en **2006**, cuando Netflix publicó un conjunto de datos anonimizados con información de usuarios para un concurso de mejora de su sistema de recomendación. Investigadores lograron cruzar estos datos con perfiles públicos de IMDb y reidentificar a varias personas basándose en sus patrones de calificación. Este incidente demostró que incluso una anonimización parcial puede ser insuficiente si los datos pueden compararse con otras bases de datos accesibles.

Sin embargo, la **pérdida de utilidad** es un problema recurrente en la anonimización. Si se eliminan demasiados detalles o se agregan niveles excesivos de ruido, los datos pueden volverse inservibles para análisis estadísticos, machine learning o toma de decisiones. El desafío radica en encontrar un equilibrio entre proteger la privacidad y mantener el valor analítico de la información.

2.2 Herramientas para Implementación en Pequeños Proyectos

Herramienta	Técnica	Costo	Facilidad de Uso
ARX	k-anonymity, l-diversity	Gratuito (open-source)	Moderada (interfaz gráfica)
Microsoft Presidio	Data masking, detección de datos sensibles	Gratuito	Alta (integración API)
Amnesia	Differential Privacy	Gratuito	Baja (requiere código)
Faker	Generación de datos sintéticos	Gratuito	Alta (biblioteca Python)

Tabla 7 comparativa entre herramientas de anonimización

*Para una comparación práctica entre herramientas disponibles según su nivel de seguridad, facilidad y costo, puede consultarse el **Anexo I** al final de este trabajo*

¿Qué es ARX?

ARX es una herramienta de código abierto desarrollada por el *Fraunhofer Institute for Algorithms and Scientific Computing (SCAI)* en Alemania. Su propósito central es permitir la anonimización segura de conjuntos de datos estructurados, especialmente aquellos que contienen información sensible o identificable, como nombres, fechas de nacimiento, direcciones o datos médicos.

Gracias a su robustez y precisión, se ha convertido en una de las herramientas más reconocidas en el ámbito de la protección de datos.

Uno de los aspectos que distingue a ARX es su capacidad para ayudar a las organizaciones a cumplir con marcos legales de privacidad exigentes, como el Reglamento General de Protección de Datos (GDPR) en Europa, la Ley de Portabilidad y Responsabilidad del Seguro de Salud (HIPAA) en Estados Unidos, y la Ley 25.326 de Protección de Datos Personales en Argentina. A diferencia de soluciones comerciales que suelen tener un alto costo, ARX ofrece una alternativa gratuita, confiable y con soporte para múltiples modelos de privacidad reconocidos internacionalmente.

El objetivo principal de ARX es proteger la privacidad individual, pero sin sacrificar por completo la utilidad de los datos. Esto es especialmente importante en contextos como la investigación científica, el análisis de datos y el entrenamiento de modelos de aprendizaje automático, donde se necesita conservar cierta calidad y representatividad en los datos anonimizados. La herramienta permite a los usuarios encontrar un equilibrio adecuado entre riesgo de reidentificación y precisión de los resultados, a través de métodos matemáticos y estadísticos avanzados.

Además de sus capacidades técnicas, ARX destaca por su interfaz visual intuitiva que facilita su uso incluso a usuarios no expertos en programación. También ofrece funciones como análisis de riesgo de reidentificación, soporte para jerarquías de generalización personalizadas, y aplicación combinada de métodos como *k-anonymity*, *l-diversity* y *t-closeness*. Estas características convierten a ARX en una solución flexible, potente y adaptable tanto para grandes organizaciones como para proyectos académicos o startups que manejan datos personales.

Técnicas soportadas:

ARX implementa varios modelos de privacidad reconocidos en el ámbito académico y profesional, lo que permite a los usuarios adaptar la anonimización a diferentes niveles de riesgo y sensibilidad. Entre los modelos más destacados se encuentra k-anonymity, que garantiza que cada registro en el conjunto de datos sea indistinguible de al menos $k - 1$ otros. Esto significa que un atacante no puede asociar un registro individual con una persona específica con una probabilidad mayor que 1/k, reduciendo significativamente el riesgo de reidentificación.

Otro modelo avanzado incluido en ARX es l-diversity, que exige que dentro de cada grupo anonimizado haya al menos *l* valores distintos para un atributo sensible. Esto evita que, incluso si un grupo cumple con k-anonymity, se puedan hacer inferencias sobre datos sensibles si todos los registros dentro del grupo

comparten el mismo valor. Por ejemplo, si todos los registros tienen el mismo diagnóstico médico, la privacidad se compromete, aun si son anónimos.

Para escenarios aún más exigentes, ARX permite aplicar t-closeness, un modelo que no solo exige diversidad, sino también que la distribución de los valores sensibles dentro de cada grupo sea estadísticamente cercana a la distribución general del conjunto de datos. Esto previene ataques basados en inferencias estadísticas, como suposiciones sobre la probabilidad de que un paciente tenga una enfermedad en función de su grupo.

En cuanto a su compatibilidad técnica, ARX trabaja con formatos de datos comunes como archivos CSV y Excel, y puede conectarse a bases de datos relacionales como MySQL y PostgreSQL, lo que la hace fácil de integrar en flujos de trabajo existentes. Una de sus ventajas más apreciadas es su interfaz gráfica de usuario (GUI), que permite configurar el proceso de anonimización paso a paso sin necesidad de escribir código, lo cual resulta especialmente útil para usuarios no técnicos o equipos interdisciplinarios.

Finalmente, ARX es una herramienta de código abierto y totalmente gratuita, disponible en GitHub. Esta apertura permite a los usuarios auditar el código, personalizar funcionalidades y participar en su desarrollo. Su enfoque en la transparencia y flexibilidad la convierte en una excelente opción tanto para instituciones académicas como para empresas que buscan cumplir con normativas de protección de datos sin incurrir en altos costos de licencias.

Instalación y Requisitos

- **Sistema operativo:** Windows, macOS, Linux.
- **Requisitos mínimos:**
    - 4 GB de RAM.
    - Java 8 o superior.
- **Descarga:**
    1. Ve a ARX Data Anonymization Tool.
    2. Descarga la versión estable (.jar o instalador).
    3. Ejecuta el archivo (requiere Java instalado).

¿Qué es Microsoft Presidio?

Microsoft Presidio es una herramienta de código abierto desarrollada por Microsoft, diseñada para proteger datos sensibles mediante técnicas de detección, anonimización y enmascaramiento. Su objetivo principal es facilitar el cumplimiento de normativas de privacidad como el Reglamento General de Protección de Datos (GDPR) en Europa o la Ley 25.326 en Argentina, ofreciendo una solución accesible incluso para startups o pequeños proyectos que no cuentan con grandes recursos técnicos o económicos. A diferencia de otras herramientas comerciales, Presidio puede instalarse localmente o desplegarse en la nube, lo que permite una mayor flexibilidad en términos de infraestructura y control sobre los datos.

Esta herramienta está especialmente orientada a trabajar con texto no estructurado, como correos electrónicos, comentarios o formularios, y puede identificar automáticamente información personal identificable (PII), como nombres, direcciones, números de documento o correos electrónicos. Una vez detectados, estos datos pueden ser anonimizados mediante técnicas como redacción, enmascaramiento parcial o reemplazo con datos ficticios. Presidio es altamente configurable y permite adaptar los modelos de detección al lenguaje y dominio específico del proyecto, lo que lo hace útil en una amplia variedad de contextos, desde análisis de sentimiento hasta sistemas de atención al cliente basados en chatbots.

Una de sus principales ventajas es que, al ser de código abierto, no implica costos de licencia, lo cual es especialmente valioso para equipos con presupuestos limitados. Sin embargo, su implementación puede requerir ciertos conocimientos técnicos, ya que su uso implica configurar detectores, definir patrones y eventualmente entrenar modelos personalizados. Además, si bien Presidio mejora significativamente la protección de los datos, no garantiza por sí sola el cumplimiento legal completo, por lo que se recomienda complementar su uso con buenas prácticas de gestión de datos y asesoramiento legal adecuado. Aun con estas limitaciones, Presidio representa una alternativa poderosa y flexible para aplicar técnicas de anonimización en contextos donde se necesita proteger la privacidad de los usuarios sin sacrificar el acceso a herramientas avanzadas.

Funcionamiento y Componentes Clave

Microsoft Presidio funciona a través de una arquitectura modular dividida en dos etapas principales: la detección y la anonimización de datos sensibles. En la primera etapa, el Analyzer se encarga de identificar información personal utilizando una combinación de expresiones regulares, listas predefinidas y modelos de aprendizaje automático. Esta etapa permite detectar diversos tipos

de datos personales como correos electrónicos, números de teléfono, direcciones físicas, identificadores gubernamentales (como el CUIL en Argentina) y números de tarjetas de crédito, entre otros. Presidio permite personalizar los analizadores para adaptarlos a contextos específicos, incluyendo distintos idiomas o estructuras propias de datos locales.

Una vez identificada la información sensible, entra en juego la segunda etapa, gestionada por el módulo Anonymizer. Este componente permite aplicar diferentes técnicas de anonimización según las necesidades del proyecto. Entre las más utilizadas se encuentra el enmascaramiento, que consiste en ocultar parcialmente los datos sensibles (por ejemplo, transformando "juan@dominio.com" en "***@dominio.com"); la tokenización, que reemplaza los datos reales por tokens irreversibles que no permiten reconstruir la información original (por ejemplo, "DNI 12.345.678" se convierte en "TK-9X2Y7Z"); y la redacción, que elimina completamente la información detectada. Estas técnicas pueden aplicarse de manera individual o combinada, dependiendo del nivel de protección deseado.

El diseño modular de Presidio también facilita su integración con otros sistemas y su personalización mediante APIs, lo cual lo convierte en una herramienta escalable y adaptable tanto para desarrolladores como para equipos multidisciplinarios que trabajen en la protección de datos. Además, al funcionar en entornos locales o en la nube, Presidio permite a las organizaciones mantener control total sobre sus datos, algo esencial en contextos donde la seguridad es prioritaria o las regulaciones impiden el uso de servicios externos.

Características Principales

Microsoft Presidio se destaca por su versatilidad y facilidad de integración en diversos entornos tecnológicos. Como herramienta multiplataforma, está desarrollada en Python, lo que permite su instalación sencilla a través de pip e integración mediante APIs REST o SDKs. Esta compatibilidad la convierte en una solución adaptable tanto para desarrolladores individuales como para equipos que trabajen con flujos automatizados de tratamiento de datos. Además, puede incorporarse sin inconvenientes en arquitecturas basadas en la nube, como Azure o AWS, así como en entornos con bases de datos tradicionales como MySQL y PostgreSQL.

Uno de los atributos más valiosos de Presidio es su personalización. La herramienta permite agregar patrones de detección personalizados, lo que facilita su aplicación en contextos locales. Por ejemplo, se pueden configurar reglas específicas para reconocer formatos de identificación típicos de Argentina, como el DNI, el CUIL o ciertas estructuras de documentos administrativos. Esta flexibilidad también permite adaptar la herramienta a regulaciones particulares

de sectores sensibles como la salud o las finanzas, cumpliendo con estándares normativos más exigentes.

En cuanto a su escalabilidad, Presidio ha sido diseñado para funcionar de manera eficiente incluso en proyectos con recursos limitados, lo que lo convierte en una alternativa viable para startups, universidades o pequeños desarrollos que necesitan cumplir con normativas de privacidad sin incurrir en altos costos de infraestructura. Además, soporta procesamiento en tiempo real, lo que permite aplicar anonimización en flujos activos, como en APIs de atención al cliente o procesamiento de formularios web, fortaleciendo la protección de datos desde el momento en que se recogen.

Ventajas

- Costo cero: Gratuito y open-source (disponible en GitHub).
- Flexibilidad: Se adapta a necesidades específicas con mínima codificación.
- Documentación robusta: Tutoriales, ejemplos de código y soporte de la comunidad.
- Compatibilidad con estándares: Alineado con GDPR, CCPA y otras normativas.

Limitaciones

A pesar de sus múltiples ventajas, Microsoft Presidio presenta algunas limitaciones importantes que deben ser consideradas al momento de su implementación. En primer lugar, requiere conocimientos técnicos básicos, ya que su configuración y personalización —especialmente la creación de nuevos patrones de detección o el ajuste de modelos— demanda familiaridad con Python y cierto dominio en procesamiento de datos. Esto puede representar una barrera para proyectos que no cuentan con personal técnico o cuya prioridad es la facilidad de uso.

Además, no es una solución completamente autónoma. Si bien Presidio es eficaz en la detección y anonimización de datos estructurados y texto plano, no está diseñado para manejar por sí mismo formatos más complejos como imágenes, audio o video. Para abordar estos casos, es necesario integrarlo con otras herramientas de análisis o reconocimiento, lo que añade una capa de complejidad técnica y operativa.

Por último, su sistema de detección, aunque robusto, no es infalible. Presidio puede omitir ciertos datos sensibles cuando estos se presentan en formatos no convencionales o muy específicos del contexto local. Esto implica que la herramienta puede requerir ajustes frecuentes para garantizar una cobertura adecuada, especialmente en entornos donde los datos no siguen estándares previsibles o están redactados en lenguaje no estructurado.

Implementación Paso a Paso

Ejemplo: Enmascarar DNI en una base de datos

Instalación:

```
pip install presidio-analyzer presidio-anonymizer
```

Código en Python:

```
from presidio_analyzer import AnalyzerEngine
from presidio_anonymizer import AnonymizerEngine

# Configurar el analizador para detectar DNIs argentinos
analyzer = AnalyzerEngine()
anonymizer = AnonymizerEngine()

texto = "El usuario con DNI 12.345.678 realizó una compra."
resultados = analyzer.analyze(text=texto, language="es")

# Anonimizar
texto_anonimo = anonymizer.anonymize(
    text=texto,
    analyzer_results=resultados,
    operators={"DEFAULT": {"type": "mask", "masking_char": "*", "chars_to_mask": 8}}
)
```

```
print(texto_anonimo.text)
# Output: "El usuario con DNI *** realizó una compra."
```

Este código detecta un número de DNI en un texto en español y lo reemplaza con asteriscos para proteger la privacidad del usuario. Es un ejemplo simple y eficaz del uso de Microsoft Presidio en un entorno Python

Integración con APIs:

Usar presidio-api para crear un servicio REST que procese datos en tiempo real.

Consideraciones Éticas y Sociales

La **equidad** en la anonimización de datos es fundamental para evitar sesgos que puedan afectar desproporcionadamente a ciertos grupos vulnerables. Si los datos anonimizados carecen de representatividad o ciertos patrones se eliminan de manera desigual, los modelos de análisis o machine learning pueden generar resultados discriminatorios. Es crucial garantizar que la anonimización no comprometa la diversidad de los datos y que las decisiones basadas en ellos no perpetúen desigualdades.

La **transparencia** es otro pilar clave en la protección de datos. Los usuarios deben ser informados sobre cómo se gestionan y protegen sus datos, qué técnicas de anonimización se utilizan y con qué propósito se procesará la información. Esto fortalece la confianza en las instituciones y empresas que manejan datos personales.

La **responsabilidad** implica la realización de auditorías periódicas para detectar posibles brechas de seguridad y garantizar que las medidas de protección sean efectivas. La revisión continua de los procesos de anonimización ayuda a prevenir riesgos de reidentificación y asegura el cumplimiento de normativas de privacidad, reforzando la protección de los datos en un entorno en constante evolución.

Herramienta Amnesia:

Amnesia es una herramienta de anonimización desarrollada por el proyecto europeo H2020 Privacy Flag, pensada para aplicar técnicas como k-anonymity y privacy diferencial sobre datasets tabulares. Se puede usar en línea mediante su interfaz web, o descargar para ejecución local.

Escenario

Se desea anonimizar un conjunto de datos que contiene información de pacientes con las siguientes columnas:

Edad	Código Postal	Diagnóstico
34	1001	Diabetes
36	1001	Hipertensión
38	1002	Gripe
35	1001	Diabetes
39	1002	Hipertensión

Tabla 8 Ejemplo para aplicar Amnesia

La combinación de 'Edad' y 'Código Postal' podría permitir identificar a personas si se accede a bases externas.

Pasos realizados en Amnesia

1. Se importa el archivo CSV en la interfaz web de Amnesia.

2. Se marca 'Edad' y 'Código Postal' como quasi-identificadores.

3. Se establece el modelo de anonimización: k-anonymity con k=3.

4. Se definen jerarquías de generalización:

- Edad: agrupar en rangos de 5 años.

- Código Postal: generalizar a los 2 primeros dígitos.

5. Amnesia aplica automáticamente la transformación que cumple con k=3 y reporta el nivel de información preservada.

Resultados obtenidos

Edad	Código Postal	Diagnóstico
30–39	10**	Diabetes
30–39	10**	Hipertensión
30–39	10**	Gripe
30–39	10**	Diabetes
30–39	10**	Hipertensión

Tabla 9 Datos anonimizado

Ventajas observadas

- Gratuita y accesible: no requiere instalación compleja.

- Interfaz visual: fácil de usar para quienes no tienen experiencia en programación.

- Informes claros: muestra gráficos de pérdida de información y riesgo de reidentificación.

- Exportable: permite descargar los datos anonimizados en CSV para usarlos en otros sistemas.

Faker: Generación de Datos Sintéticos para Protección de Datos Personales.

Faker es una biblioteca de código abierto escrita originalmente en Python, aunque también se encuentra disponible en otros lenguajes como Java, JavaScript, PHP y Ruby. Su propósito principal es la generación automatizada de datos sintéticos que imitan información personal realista, sin comprometer la privacidad de individuos reales.

Entre los tipos de datos que Faker permite generar se encuentran nombres completos, documentos de identidad, correos electrónicos, direcciones postales, números telefónicos, fechas de nacimiento, perfiles de usuario, datos financieros ficticios (como números de tarjeta de crédito), entre otros. Esta información, aunque simulada, sigue patrones realistas de acuerdo con el país o región configurados, lo que permite crear escenarios creíbles para el desarrollo y prueba de sistemas.

El uso de datos sintéticos es especialmente útil en contextos donde se requiere manipular información similar a la real, pero sin infringir normativas de privacidad como el Reglamento General de Protección de Datos (GDPR) en Europa, la Ley de Protección de Datos Personales en Argentina (Ley 25.326) o la Ley de Privacidad del Consumidor de California (CCPA). Faker ofrece una solución práctica para evitar el uso de datos personales reales en procesos de desarrollo, testing y entrenamiento de modelos de inteligencia artificial.

¿Por qué usar datos sintéticos?

En entornos donde no se cuenta con acceso a datos reales debido a restricciones legales, políticas de confidencialidad o riesgos éticos, los datos sintéticos generados con herramientas como Faker ofrecen una alternativa segura y práctica. Si bien no reflejan información auténtica de individuos reales, mantienen patrones estadísticos o formatos coherentes que permiten realizar pruebas, simulaciones y validaciones funcionales en sistemas que manipulan datos personales.

Generación de datos sintéticos con Faker (Python)

Escenario: Se necesita simular datos de usuarios para probar un sistema sin comprometer datos reales.

```
from faker import Faker
import pandas as pd
import matplotlib.pyplot as plt
import matplotlib.table as tbl


fake = Faker('es_ES')  # Podés cambiarlo por 'en_US' si lo preferís


data = {
    "Nombre completo": [fake.name() for _ in range(5)],
    "DNI": [fake.random_number(digits=8, fix_len=True) for _ in range(5)],
    "Dirección": [fake.address().replace('\n', ', ') for _ in range(5)],
    "Email": [fake.email() for _ in range(5)]
}


df = pd.DataFrame(data)
df.index += 1
```

Salida posible:

```
fig, ax = plt.subplots(figsize=(12, 2.5))
ax.axis('off')
tabla = tbl.table(ax, cellText=df.values, colLabels=df.columns, cellLoc='center', loc='center')
tabla.auto_set_font_size(False)
tabla.set_fontsize(10)
tabla.scale(1.2, 1.5)
plt.savefig("ejemplo_faker_datos_sinteticos.png", bbox_inches='tight')
```

Ventajas de Faker

- Privacidad garantizada**:** evita el uso de datos reales y, por tanto, reduce a cero el riesgo de filtraciones o reidentificación.
- Gratuito y de código abierto: no requiere licencias ni infraestructura adicional.
- Flexible y configurable: permite definir estructuras de datos complejas o personalizados (por ejemplo, listas de nombres de pacientes, usuarios, transacciones).
- Compatible con machine learning y pruebas de sistemas: útil para entrenar algoritmos o testear funcionalidades sin exponer datos sensibles.

Consideraciones

Si bien Faker genera datos plausibles, no debe confundirse con datos reales ni utilizarse en contextos donde se requiere fidelidad estadística o validez científica de las muestras**.** Para esos casos, existen herramientas específicas de generación de datos sintéticos basados en modelos probabilísticos.

Aplicación en pequeños proyectos

Para startups, estudiantes o proyectos de bajo presupuesto, Faker es una herramienta valiosa para simular bases de datos sin exponerse a sanciones por el uso indebido de datos personales. También resulta útil en procesos de anonimización, donde se desea validar visualmente la efectividad de técnicas como masking o tokenización sin trabajar con datos verdaderos.

Capítulo III

Mejores Prácticas de Protección de Datos

El capítulo presenta prácticas clave para proteger la privacidad digital, como limitar permisos, usar contraseñas seguras, educar sobre riesgos y proteger dispositivos. También resalta el rol de organizaciones y gobiernos en establecer políticas, mejorar la legislación y fomentar la colaboración internacional para una protección de datos más efectiva.

Revocar permisos innecesarios es fundamental para proteger la privacidad, ya que muchas aplicaciones mantienen acceso a nuestros datos incluso cuando ya no las utilizamos. Es importante revisar periódicamente las aplicaciones conectadas a nuestras cuentas y dispositivos, eliminando aquellas que ya no usamos o en las que no confiamos. También es recomendable desactivar permisos innecesarios, como el acceso a la ubicación, cámara o contactos, especialmente en aplicaciones que no los requieren para su funcionamiento.

Leer los términos y condiciones antes de usar una aplicación es crucial para entender cómo se recopilan, usan y almacenan nuestros datos. Es necesario prestar atención a qué información se solicita, con qué propósito y si se comparte con terceros. También es útil verificar las políticas de almacenamiento y eliminación de datos para asegurarse de que se respeten los derechos del usuario. Herramientas como *TOS;DR* pueden ayudar a simplificar esta tarea y facilitar una comprensión más rápida de los puntos clave.

La protección de dispositivos es esencial para garantizar la seguridad digital. Utilizar contraseñas seguras, combinando letras, números y símbolos, es una medida básica pero efectiva. También es recomendable utilizar generadores de contraseña. Además, activar la autenticación en dos pasos añade una capa extra de protección, dificultando el acceso no autorizado. Mantener los dispositivos y aplicaciones actualizados es igualmente importante, ya que las actualizaciones suelen corregir vulnerabilidades que podrían ser explotadas por atacantes. Por último, el uso de una red privada virtual (VPN) es clave para proteger la conexión a internet, especialmente al usar redes públicas, ya que cifra los datos y reduce el riesgo de interceptaciones.

Conciencia y Educación

La conciencia y la educación digital son herramientas clave para reducir los riesgos asociados al uso de internet. Reconocer intentos de phishing y estafas en línea es esencial, ya que este tipo de tácticas buscan engañar a los usuarios para obtener información personal o financiera. Esto implica identificar correos o mensajes sospechosos que suelen incluir enlaces maliciosos, archivos adjuntos peligrosos o solicitudes de datos urgentes provenientes de remitentes desconocidos. Comprender estas señales y actuar con cautela ayuda a evitar caer en estas trampas.

Por otro lado, la supervisión de las actividades de los menores en internet es una responsabilidad fundamental para garantizar su seguridad. Los menores son particularmente vulnerables a riesgos como el ciberacoso, el grooming o la exposición a contenido inapropiado. Es importante enseñarles a utilizar la tecnología de manera responsable, estableciendo límites en el tiempo de pantalla, promoviendo el respeto en las interacciones en línea y fomentando un

diálogo abierto sobre sus experiencias digitales. De esta forma, se puede contribuir a un entorno digital más seguro tanto para los jóvenes como para sus familias.

Políticas de Manejo de Datos

Las políticas de manejo de datos desempeñan un papel central en la protección de la privacidad de los usuarios, especialmente en un mundo donde la recopilación de información personal es constante. Implementar medidas transparentes permite a los usuarios comprender cómo se maneja su información, generando confianza y promoviendo un uso más consciente de las plataformas digitales. Estas políticas deben especificar de manera clara qué datos se recopilan, con qué propósito se utilizan y durante cuánto tiempo se almacenan, ofreciendo además opciones para que los usuarios gestionen o eliminen su información.

Un aspecto importante es minimizar la recopilación de datos, limitándola estrictamente a lo necesario para las funciones esenciales de un servicio o aplicación. Por ejemplo, muchas plataformas piden permisos para acceder a información adicional como la ubicación o la lista de contactos, incluso cuando no es relevante para su propósito principal. Reducir esta práctica no solo disminuye los riesgos de filtraciones o uso indebido, sino que también refuerza el compromiso ético de respetar la privacidad del usuario. Este enfoque equilibrado entre funcionalidad y privacidad es clave para construir un entorno digital más seguro y confiable.

Cifrado y Seguridad

El cifrado y la seguridad representan elementos clave para garantizar la integridad y confidencialidad de la información en el ámbito digital. El cifrado es una de las herramientas más efectivas para proteger datos sensibles, como información financiera, contraseñas o comunicaciones privadas. Mediante este proceso, los datos se transforman en un formato ilegible para cualquier persona que no cuente con la clave de descifrado, lo que dificulta significativamente el acceso no autorizado, incluso en casos de intercepción. El uso de cifrado es fundamental tanto en comunicaciones en línea, como correos electrónicos o aplicaciones de mensajería, como en el almacenamiento de datos en dispositivos o servicios en la nube.

Junto con el cifrado, las auditorías regulares son esenciales para garantizar la seguridad a largo plazo. Estas auditorías permiten evaluar de manera continua los sistemas y detectar posibles vulnerabilidades que puedan ser explotadas por ciberatacantes. Las revisiones deben incluir pruebas de

penetración, análisis de configuraciones de seguridad y verificación de actualizaciones de software y hardware. Además, estas auditorías son una oportunidad para revisar y actualizar las políticas internas de seguridad, adaptándolas a nuevas amenazas y tecnologías emergentes.

En un entorno digital donde las amenazas son cada vez más sofisticadas, adoptar estas medidas no solo protege la información sensible, sino que también fortalece la confianza de los usuarios y las organizaciones en la gestión de datos. Este enfoque combinado entre cifrado y auditorías regulares es indispensable para construir un ecosistema digital más seguro y resiliente frente a los desafíos actuales.

Concientización Interna

La concientización interna es un aspecto clave en la protección de la privacidad y los datos personales dentro de las organizaciones. Capacitar al personal sobre la importancia de la privacidad permite que cada miembro comprenda su responsabilidad en la gestión y protección de la información sensible. Además, es fundamental educarlos sobre las mejores prácticas para proteger los datos personales, como el manejo adecuado de contraseñas, la identificación de correos o enlaces sospechosos, y la implementación de medidas de seguridad en sus dispositivos y cuentas. Esta formación no solo reduce el riesgo de filtraciones de información, sino que también fomenta una cultura organizacional que valora la seguridad y la privacidad en todas sus operaciones.

Fortalecer la Legislación

Fortalecer la legislación en cuanto a la protección de datos personales es una necesidad urgente en un mundo cada vez más digitalizado. A medida que las tecnologías emergentes, como el internet de las cosas (IoT) y la inteligencia artificial (IA), continúan desarrollándose y siendo adoptadas en una variedad de sectores, las leyes deben actualizarse para abordar los nuevos retos que estas tecnologías presentan en términos de recopilación, almacenamiento y uso de datos. El IoT, por ejemplo, implica una gran cantidad de dispositivos conectados que recogen datos constantemente, mientras que la IA puede procesar grandes volúmenes de información personal con fines predictivos o de automatización, lo que puede generar riesgos para la privacidad si no se regula adecuadamente.

En Argentina, la protección de datos personales está regulada principalmente por la Ley 25.326, conocida como Ley de Protección de Datos Personales, sancionada en el año 2000. Esta ley establece principios fundamentales, como el consentimiento informado del titular para la recolección

de datos, la finalidad específica de su uso y la obligación de garantizar su seguridad. La normativa fue complementada con decretos y reglamentaciones, y su cumplimiento es supervisado por la Agencia de Acceso a la Información Pública (AAIP), que actúa como autoridad de control. Sin embargo, dado que la ley fue creada en un contexto tecnológico muy distinto al actual, surge la necesidad de actualizarla para enfrentar los desafíos que plantean las tecnologías modernas.

Además, las leyes deben garantizar que las empresas cumplan con estándares estrictos para la gestión de los datos personales. Esto incluye la obligación de implementar prácticas de transparencia, donde los usuarios estén completamente informados sobre qué datos se recopilan, cómo se utilizan y por cuánto tiempo se almacenan. También es necesario que las organizaciones adopten medidas adecuadas de seguridad para proteger esta información frente a ciberataques o filtraciones, como el cifrado de datos y la autenticación en múltiples pasos. En este sentido, la Argentina se encuentra en proceso de adaptar su legislación para alinearla con estándares internacionales, como el Reglamento General de Protección de Datos (GDPR) de la Unión Europea.

La imposición de sanciones y la realización de auditorías regulares son instrumentos efectivos para asegurar el cumplimiento de estas normativas. Estas medidas no solo fomentan la responsabilidad dentro de las empresas, sino que también sirven para crear un entorno de confianza, donde los consumidores puedan sentirse seguros de que su información personal está siendo gestionada de manera adecuada. En última instancia, una legislación robusta y bien adaptada, que contemple tanto los derechos de los usuarios como la evolución tecnológica, garantiza que los derechos de los individuos sobre sus datos sean protegidos, promoviendo un equilibrio entre la innovación tecnológica y la privacidad.

Campañas de Educación Pública

Las campañas de educación pública son esenciales para aumentar la conciencia sobre la protección de datos personales y los riesgos asociados con la exposición en línea. Promover la alfabetización digital es clave para que los ciudadanos comprendan no solo cómo interactuar de manera segura en el entorno digital, sino también los peligros potenciales que surgen cuando se comparten datos personales sin precauciones. Estas campañas pueden incluir desde la divulgación de información sobre cómo identificar amenazas como el phishing, hasta la educación sobre el manejo adecuado de la privacidad en redes sociales y otras plataformas.

Asimismo, es fundamental desarrollar herramientas accesibles que permitan a los usuarios controlar sus datos en línea de manera efectiva. Esto

implica ofrecer opciones claras y sencillas para gestionar la privacidad en diversas plataformas, como la posibilidad de ajustar configuraciones de seguridad, revisar qué información se comparte y cómo se utiliza, y tener control sobre el acceso a sus datos. Al proporcionar estas herramientas, se empodera a los usuarios para que tomen decisiones informadas sobre su privacidad, reduciendo así el riesgo de exposición indebida de su información personal.

4.2 Colaboración Internacional

La colaboración internacional desempeña un papel fundamental en la protección de los datos personales, ya que los desafíos relacionados con la privacidad no se limitan a una sola nación, sino que son problemas globales que requieren soluciones coordinadas. Participar en foros y organizaciones internacionales permite a los países trabajar juntos en el desarrollo de estándares unificados que aborden cuestiones como el manejo ético de la información, la regulación de tecnologías emergentes y la protección contra ciberamenazas.

Estas alianzas internacionales también fomentan el intercambio de conocimientos y experiencias sobre estrategias efectivas para mitigar riesgos, así como el establecimiento de protocolos comunes para la transferencia de datos entre jurisdicciones. Esto es especialmente relevante en un mundo donde las empresas operan a nivel global y los datos de los usuarios cruzan fronteras constantemente.

Además, la colaboración facilita la adopción de marcos legales y tecnológicos que sean compatibles entre diferentes países, promoviendo un enfoque armonizado para la seguridad digital. Este tipo de cooperación no solo refuerza la protección de los datos individuales, sino que también crea un entorno de confianza para las empresas y los usuarios, fortaleciendo la economía digital global mientras se respetan los derechos fundamentales a la privacidad.

4.3 Herramientas disponibles para pequeños proyectos

Las herramientas de anonimización, como **ARX** y **Microsoft Presidio**, ofrecen soluciones técnicas que pueden ayudar a las empresas a proteger los datos personales, pero por sí solas no garantizan el cumplimiento completo de las normativas. **ARX** permite aplicar técnicas como **k-anonymity**, **l-diversity** y **t-closeness** en datos estructurados, pero estas medidas no son suficientes para proteger contra ciertos tipos de ataques de reidentificación, como los de homogeneidad o vinculación con bases de datos externas. Aunque la **k-anonymity** es una técnica útil, no cumple completamente con los estrictos requisitos del **GDPR** para anonimizar los datos de manera irreversible.

Por su parte, **Microsoft Presidio** se enfoca en la detección y ofuscación de datos sensibles en textos no estructurados, pero también enfrenta limitaciones en cuanto a la exactitud de la detección de datos sensibles, lo que podría generar **falsos positivos o negativos**. Además, estas herramientas no proporcionan mecanismos suficientes para cumplir con otras exigencias del **GDPR**, como la **gestión del consentimiento** o la **auditoría continua**.

5.1 Implementación de estándares de privacidad

La implementación de los estándares de privacidad requeridos por el **GDPR** y la **Ley 25.326** implica más que solo aplicar técnicas de anonimización. Para cumplir completamente con estas normativas, las pequeñas empresas deben adoptar un enfoque integral que incluya mecanismos de consentimiento explícito, mantenimiento de registros de auditoría, y procesos de gestión de datos personales que no son completamente abordados por las herramientas de anonimización existentes. La falta de infraestructura adecuada en proyectos pequeños para implementar todos estos elementos puede hacer que el cumplimiento sea muy difícil sin un soporte técnico y legal adicional.

En conclusión, aunque las herramientas de anonimización disponibles, como **ARX** y **Microsoft Presidio**, pueden ser útiles, no son suficientes para garantizar el cumplimiento total de las normativas como el **GDPR** o la **Ley 25.326** en Argentina. Las empresas deben contar con una infraestructura más robusta y contar con asesoramiento legal para navegar las complejidades del cumplimiento normativo.

Capitulo IV.

Estudios de Caso y Lecciones Aprendidas

Este capítulo analiza casos emblemáticos de fallas y avances en la protección de datos. El escándalo de **Cambridge Analytica y Facebook** mostró cómo la falta de control sobre aplicaciones de terceros puede facilitar el uso indebido de datos personales con impacto político. El caso del **Netflix Prize** evidenció los límites de la anonimización tradicional al demostrarse que los datos supuestamente anónimos podían reidentificarse. Por otro lado, **Apple** implementó **differential privacy**, mostrando que es posible recolectar datos útiles sin comprometer la privacidad. En el sector público, una **filtración en Argentina** resaltó la necesidad de medidas robustas de seguridad estatal. Finalmente, empresas como **Auth0** destacaron por su uso de técnicas como **tokenización** y **autenticación multifactor**, marcando un estándar en la protección de identidades digitales.

Analizar casos reales de manejo y filtración de datos personales es fundamental para entender los retos y las mejores prácticas en la protección de la privacidad. Estos ejemplos permiten identificar fallos comunes, evaluar la efectividad de distintas técnicas de anonimización y seguridad, y aprender cómo las organizaciones pueden mejorar sus políticas y tecnologías para proteger mejor la información sensible. Además, conocer las lecciones aprendidas de estos incidentes contribuye a fortalecer la regulación, aumentar la transparencia y fomentar una cultura de responsabilidad en el manejo de datos en el ámbito público y privado.

Cambridge Analytica y Facebook:

El caso de *Cambridge Analytica y Facebook* fue una de las mayores filtraciones de datos personales en la historia reciente, revelando graves fallos en la protección de la privacidad. En 2014, el investigador Aleksandr Kogan desarrolló una aplicación llamada *"This Is Your Digital Life"*, que ofrecía un test de personalidad a los usuarios de Facebook. Sin embargo, la app no solo recopilaba datos de quienes la usaban, sino también de sus amigos en la plataforma, aprovechando una vulnerabilidad en la API de Facebook. Como resultado, Cambridge Analytica obtuvo información detallada de aproximadamente 87 millones de usuarios sin su consentimiento explícito. Estos datos fueron utilizados para crear perfiles psicológicos y estrategias de microsegmentación dirigidas a influir en campañas políticas, como la elección presidencial de EE.UU. en 2016 y el referéndum del Brexit en el Reino Unido. A pesar de que Facebook aseguraba haber restringido este tipo de acceso en 2015, Cambridge Analytica mantuvo los datos y los usó con fines comerciales y políticos.

Este caso demuestra que la pseudonimización de datos no es suficiente si existe la posibilidad de cruzarlos con otras fuentes externas, lo que permite reidentificar a los usuarios. Además, pone en evidencia los riesgos de otorgar acceso indiscriminado a datos personales a aplicaciones de terceros. Como consecuencia, se implementaron regulaciones más estrictas, como el Reglamento General de Protección de Datos (GDPR) en Europa, y Facebook recibió una multa de 5.000 millones de dólares por parte de la Comisión Federal de Comercio de EE.UU. Este incidente subraya la importancia de fortalecer los mecanismos de anonimización, limitar la recopilación de datos y garantizar que los usuarios tengan un control real sobre su información personal.

Netflix Prize (2006):

En 2006, Netflix lanzó el concurso Netflix Prize con el objetivo de mejorar su algoritmo de recomendación en al menos un 10%, ofreciendo un premio de 1

millón de dólares al equipo que lograra este avance. Para facilitar el trabajo de los participantes, la empresa publicó un conjunto de datos que contenía 100 millones de calificaciones de películas realizadas por 500.000 usuarios. Netflix aseguró que los datos habían sido anonimizados, eliminando nombres y otros identificadores directos, y confiando en técnicas de privacidad como k-anonymity para evitar la reidentificación de los usuarios. Sin embargo, investigadores de la Universidad de Texas demostraron que era posible desanonimizar la información al cruzarla con bases de datos públicas, como las reseñas de películas en IMDb. A través de técnicas de correlación, lograron inferir la identidad de algunas personas, lo que permitió revelar detalles sobre sus hábitos de visualización, incluyendo potencialmente preferencias sensibles o comprometedoras. Este descubrimiento generó preocupaciones sobre la seguridad de los datos anonimizados y llevó a una demanda colectiva contra Netflix, que finalmente canceló un segundo concurso similar en 2010.

Lecciones aprendidas

El caso de Netflix Prize puso en evidencia los límites de la anonimización basada en k-anonymity, que, si bien oculta identificadores directos, no impide la reidentificación si los datos pueden cruzarse con otras fuentes. La filtración dejó claro que, en muchos casos, las técnicas tradicionales de anonimización no son suficientes para proteger la privacidad de los usuarios y que es necesario aplicar métodos más avanzados, como la privacidad diferencial, que introduce ruido en los datos para evitar que individuos específicos puedan ser identificados. Además, este incidente resaltó la importancia de evaluar cuidadosamente los riesgos antes de compartir datos masivos, especialmente cuando contienen información sobre comportamientos personales. A raíz de este caso, la comunidad científica y tecnológica comenzó a prestar mayor atención a la robustez de los métodos de anonimización, lo que llevó a mejoras en las prácticas de privacidad en el manejo de grandes volúmenes de datos.

Apple y Differential Privacy:

Apple ha sido una de las empresas pioneras en la implementación de Differential Privacy, una técnica avanzada de anonimización que permite recopilar datos de los usuarios sin exponer sus identidades individuales. A diferencia de los enfoques tradicionales de anonimización, que pueden ser vulnerables a ataques de reidentificación al cruzar datos con otras fuentes, Differential Privacy introduce ruido matemático en la información recopilada, lo que dificulta la identificación de un usuario específico dentro de un conjunto de datos. Apple comenzó a aplicar esta técnica en 2016 con el objetivo de mejorar sus servicios sin comprometer la privacidad de sus clientes. Se ha utilizado en

funciones como la corrección automática del teclado, sugerencias de emojis, búsquedas en Spotlight y patrones de uso de aplicaciones, permitiendo a la compañía obtener información útil a nivel agregado sin almacenar datos individuales de forma reconocible.

Lecciones aprendidas

El enfoque de Apple demuestra que es posible recopilar datos para mejorar productos y servicios sin sacrificar la privacidad de los usuarios. **Differential Privacy** ha marcado un cambio significativo en la forma en que las grandes empresas manejan la información, estableciendo un nuevo estándar para la protección de datos. Sin embargo, su implementación no está exenta de desafíos, ya que el equilibrio entre agregar suficiente ruido para garantizar privacidad y mantener la utilidad de los datos es complejo. Apple ha sido criticada en algunas ocasiones por la falta de transparencia en sus algoritmos y por las limitaciones en la capacidad de los investigadores para auditar su implementación. Aun así, su adopción ha impulsado una mayor conciencia sobre la importancia de técnicas avanzadas de anonimización, alentando a otras empresas y gobiernos a explorar métodos similares para garantizar la seguridad de los datos personales en un mundo cada vez más digitalizado.

Filtraciones de datos en el sector público:

En 2021, se reportó en Argentina una **filtración de datos personales** de beneficiarios de programas sociales debido a una falta de anonimización adecuada. Esta exposición comprometió información sensible de ciudadanos que reciben asistencia estatal, generando preocupaciones sobre la seguridad y privacidad de los datos manejados por entidades gubernamentales. Aunque no se dispone de detalles específicos en las fuentes proporcionadas, este incidente resalta la importancia de implementar medidas robustas de protección de datos en el sector público.

Lecciones aprendidas

- **Importancia de la anonimización**: Es fundamental aplicar técnicas efectivas de anonimización y privacidad diferencial al manejar datos personales, especialmente en el ámbito gubernamental, para prevenir posibles reidentificaciones y proteger la privacidad de los ciudadanos.
- **Transparencia y responsabilidad**: Las instituciones públicas deben ser transparentes en sus procesos de manejo de datos y rendir cuentas en caso de incidentes, fortaleciendo la confianza pública.

- **Actualización de protocolos de seguridad**: Es esencial revisar y actualizar continuamente las políticas y protocolos de seguridad de la información para adaptarse a nuevas amenazas y garantizar la protección de los datos personales.

Este caso subraya la necesidad de que los organismos estatales en Argentina y América Latina refuercen sus prácticas de gestión de datos para salvaguardar la información de los ciudadanos y mantener la integridad de los programas sociales.

Startups tecnológicas:

Auth0, una startup tecnológica fundada en Argentina y posteriormente adquirida por Okta, se ha posicionado como un referente en seguridad digital mediante la implementación de **tokenización** para proteger las identidades de los usuarios. La **tokenización** es una técnica de seguridad que reemplaza datos sensibles, como credenciales de acceso, con identificadores únicos llamados **tokens**, que no tienen valor fuera del sistema en el que fueron generados. Esto minimiza el riesgo de exposición de información personal en caso de una filtración o ataque cibernético. En el ámbito de la autenticación, Auth0 utiliza **JSON Web Tokens (JWT)** para validar sesiones de usuario sin necesidad de almacenar o transmitir contraseñas repetidamente. Además, la empresa ofrece soluciones avanzadas como autenticación multifactor (MFA), detección de fraudes y Single Sign-On (SSO), permitiendo a desarrolladores integrar seguridad de forma sencilla en sus aplicaciones. A través de su blog técnico y webinars, Auth0 proporciona recursos educativos sobre mejores prácticas en autenticación y protección de identidades, ayudando a empresas de todos los tamaños a mejorar su ciberseguridad.

Lecciones aprendidas

- La tokenización mejora la seguridad sin comprometer la experiencia del usuario: Al eliminar la necesidad de transmitir datos sensibles, se reducen los riesgos sin afectar el rendimiento de las aplicaciones.
- Las credenciales tradicionales no son suficientes: Las startups tecnológicas deben complementar la autenticación con medidas como MFA y detección de amenazas en tiempo real para evitar accesos no autorizados.
- La educación en ciberseguridad es clave: Empresas como Auth0 demuestran que compartir conocimientos a través de blogs y webinars ayuda a mejorar la seguridad en el ecosistema digital.

- La seguridad debe ser un estándar, no una opción: La adopción temprana de estrategias como la tokenización permite a startups y grandes empresas proteger la identidad de sus usuarios desde el diseño del producto.

El enfoque de Auth0 refuerza la importancia de adoptar tecnologías avanzadas para garantizar la seguridad en el desarrollo de aplicaciones, estableciendo un modelo a seguir en la protección de datos personales en el mundo digital.

Casos:

Se presenta una discusión crítica sobre los desafíos y consideraciones en la anonimización y protección de datos. Es fundamental analizar no solo las técnicas y herramientas disponibles, sino también las limitaciones prácticas y normativas que enfrentan en distintos contextos. Reflexionar sobre estos aspectos ayuda a encontrar un equilibrio efectivo entre la protección de la privacidad y el uso útil de los datos, lo que resulta clave para orientar decisiones y políticas en la gestión segura de información personal.

Anonimización de Datos: Desafíos, Buenas Prácticas y Aplicación Práctica

Las herramientas de anonimización y protección de datos han avanzado significativamente, pero su efectividad depende tanto de la técnica utilizada como del contexto en el que se aplican. **ARX**, por ejemplo, permite implementar **k-anonymity, l-diversity y t-closeness** en datos estructurados. En la práctica, **k-anonymity en ARX** es útil para ocultar identificadores directos al agrupar registros con características similares, pero, como lo señala la teoría, sigue siendo vulnerable a ataques de homogeneidad y vinculación con bases de datos externas. Un ejemplo de esto se da en conjuntos de datos médicos, donde, si todos los pacientes de un grupo anonimizado comparten la misma enfermedad, su diagnóstico puede deducirse fácilmente, comprometiendo la privacidad.

Por otro lado, **Microsoft Presidio** se enfoca en la detección y ofuscación de información sensible en textos no estructurados mediante procesamiento de lenguaje natural (NLP). A diferencia de ARX, no aplica k-anonymity, pero permite sustituir o eliminar identificadores personales en documentos y registros. Sin embargo, en la práctica, su precisión depende de la configuración y de los modelos de aprendizaje utilizados, lo que puede generar falsos positivos o negativos si los datos no están bien etiquetados.

Además, el enfoque de **differential privacy** ha surgido como una alternativa más robusta en comparación con los métodos clásicos de anonimización. Implementado en sistemas como los de **Apple y Google**, este

método introduce ruido matemático en los datos para proteger la identidad de los usuarios. No obstante, en la práctica, su implementación es compleja y requiere un equilibrio entre privacidad y utilidad de los datos. Si se introduce demasiado ruido, los datos pueden perder valor estadístico, mientras que si se introduce poco, podrían permitir reidentificaciones.

En síntesis, el análisis de estas herramientas confirma que no existe una solución única para la protección de datos. Mientras que **k-anonymity** sigue siendo útil para datos estructurados, sufre limitaciones que pueden mitigarse con técnicas complementarias como l-diversity o differential privacy. Al mismo tiempo, herramientas como Microsoft Presidio muestran la importancia del procesamiento de datos no estructurados, aunque dependen de la calidad de los modelos utilizados. En última instancia, la mejor estrategia de privacidad combina múltiples enfoques para lograr un equilibrio entre anonimización efectiva y conservación de la utilidad de los datos.

Cumplimiento normativo vs. realidad técnica:

El cumplimiento normativo en términos de protección de datos personales es una preocupación constante para las empresas, especialmente bajo regulaciones como la Ley 25.326 en Argentina y el Reglamento General de Protección de Datos (GDPR) en Europa. Estas normativas imponen altos estándares de protección, exigiendo, por ejemplo, el consentimiento explícito del usuario, la implementación de medidas de seguridad técnicas y organizativas, y la garantía de derechos como el acceso, rectificación o supresión de datos.

Sin embargo, en la práctica, los pequeños proyectos y startups se enfrentan a importantes obstáculos para cumplir plenamente con estos requisitos. Las herramientas de anonimización o gestión de privacidad más robustas, como Amnesia, ARX o Skyflow, requieren conocimientos técnicos especializados, infraestructura adecuada y, en muchos casos, licencias costosas. Para equipos reducidos sin un departamento legal o de IT dedicado, estas exigencias resultan difíciles de abordar.

A esta brecha se suma el hecho de que muchas soluciones tecnológicas gratuitas o de bajo costo carecen de certificaciones, validaciones oficiales o documentación suficiente que asegure su conformidad con las normativas. Por ejemplo, una herramienta puede aplicar una técnica de seudonimización sin cumplir con los criterios exigidos para considerar que los datos ya no son personales según el GDPR. Esto expone a las organizaciones a posibles sanciones, incluso cuando actúan de buena fe.

Además, los marcos regulatorios muchas veces asumen una capacidad técnica que no se condice con la realidad de los pequeños actores del ecosistema digital. No existen mecanismos que faciliten el cumplimiento

progresivo o adaptado al tamaño y recursos de la organización. Esta desconexión genera un contexto en el que muchas startups deben elegir entre avanzar con sus proyectos sin cumplir del todo con la normativa o frenar el desarrollo por temor a incumplimientos legales.

En este escenario, sería deseable que los entes reguladores desarrollen guías adaptadas a las realidades de pequeñas organizaciones, promuevan herramientas validadas open-source, e incluso generen espacios de acompañamiento técnico o legal para proyectos en etapa temprana. La armonización entre exigencias legales y capacidades técnicas reales resulta clave para fomentar una cultura de protección de datos sin desalentar la innovación.

Desafíos con la Ley 25.326 y el GDPR

La Ley 25.326 de Argentina establece principios generales que se alinean con marcos internacionales como el Reglamento General de Protección de Datos (GDPR) de la Unión Europea. Sin embargo, cuando se trata de la aplicación práctica de conceptos técnicos como la anonimización, la normativa argentina presenta vacíos importantes. Mientras que el GDPR define claramente la anonimización como un proceso irreversible, en el que los datos ya no pueden vincularse a una persona identificable por ningún medio razonablemente disponible, la legislación argentina no especifica con el mismo nivel de detalle los requisitos técnicos o metodológicos que debe cumplir un proceso para ser considerado verdaderamente irreversible.

Esta ambigüedad genera un entorno de incertidumbre para las empresas, especialmente las startups o proyectos pequeños que no cuentan con asesoría legal especializada. No está claro si técnicas como la pseudonimización, que aún permiten cierta reversibilidad bajo condiciones controladas, serían aceptadas como suficientes en caso de una auditoría o requerimiento de la autoridad de control. Además, tampoco existen lineamientos oficiales sobre qué algoritmos o herramientas son considerados adecuados para llevar a cabo una anonimización conforme a derecho.

En la práctica, muchas organizaciones deben interpretar la normativa por su cuenta o basarse en buenas prácticas internacionales, lo cual no siempre garantiza cumplimiento local. Esta falta de claridad puede derivar en decisiones técnicas mal fundamentadas que aumenten los riesgos de reidentificación de datos o infracciones legales, incluso si la intención de cumplir con la ley está presente.

Además, la Ley 25.326 no ha sido actualizada en profundidad desde su sanción en el año 2000, lo que deja fuera del marco regulatorio a muchas tecnologías emergentes que utilizan o generan datos personales de forma

automatizada, como ocurre en contextos de big data, inteligencia artificial, dispositivos IoT o aplicaciones móviles.

Por ello, se vuelve urgente una actualización normativa que incluya definiciones técnicas más precisas sobre anonimización, criterios de evaluación de riesgo de reidentificación, y guías prácticas que ayuden a las organizaciones a aplicar correctamente estos principios sin depender exclusivamente de interpretaciones jurídicas. También sería deseable que la Agencia de Acceso a la Información Pública publique recomendaciones específicas o criterios de evaluación técnica que otorguen mayor seguridad jurídica a quienes manejan datos personales en el país.

Capitulo V

Protocolo General de Protección de Datos Personales

Este capítulo define un conjunto de directrices técnicas y organizativas para proteger los datos personales en cualquier sistema o proyecto. El protocolo establece cómo identificar, clasificar y aplicar medidas de protección a los datos, desde la recolección hasta su eliminación, siguiendo principios de privacidad por diseño y por defecto.

1. Objetivo

Este protocolo establece las directrices técnicas y organizativas mínimas necesarias para garantizar la protección de los datos personales tratados en cualquier proyecto, sistema o flujo de información bajo responsabilidad de la organización. Se busca mitigar riesgos de identificación directa o indirecta de los titulares, en cumplimiento de principios de privacidad por diseño y por defecto.

2. Alcance

Aplica a todos los entornos, sistemas y procesos donde se recolecten, almacenen, procesen o compartan datos personales o seudonimizados, incluyendo entornos de desarrollo, testing, producción, almacenamiento en la nube, exportaciones de datos y procesos de análisis.

3. Principios Generales

- **Minimización de datos**: recolectar y procesar únicamente los datos estrictamente necesarios para los fines definidos.
- **Limitación de finalidad**: los datos solo pueden utilizarse para los propósitos para los cuales fueron recolectados.
- **Transparencia y trazabilidad**: todas las actividades sobre datos personales deben poder ser auditadas.
- **Seguridad**: garantizar la integridad, confidencialidad y disponibilidad de los datos en todo su ciclo de vida.

4. Clasificación de Datos

Todo conjunto de datos debe ser previamente clasificado según su sensibilidad. Se establecen las siguientes categorías:

- **Identificadores directos** (e.g., DNI, nombre completo, dirección exacta)
- **Identificadores indirectos** (e.g., combinación de edad, género, ubicación)
- **Datos sensibles** (e.g., salud, orientación sexual, religión)
- **Datos anonimizados o seudonimizados** (sin posibilidad o con baja probabilidad de reidentificación)

5. Técnicas de Protección

5.1 Seudonimización

Transformación de los datos personales de forma que no puedan atribuirse a un titular sin información adicional. Ejemplos: reemplazo de ID reales por códigos arbitrarios, uso de hashes con sal.

5.2 Tokenización

Reemplazo de elementos sensibles por tokens aleatorios o reversibles bajo custodia segura. Se utiliza en casos como tarjetas de crédito o identificadores únicos.

5.3 Enmascaramiento (Masking)

Ocultamiento parcial o total de datos en ambientes de no producción. Ejemplo: Juan Pérez → J*** P***.

5.4 Anonimización

Aplicación de técnicas irreversibles que impiden la identificación del titular. Incluye:

- **k-anonymity**: garantizar que cada registro no pueda distinguirse de al menos *k-1* registros.
- **l-diversity**: asegurar diversidad de valores sensibles dentro de los grupos de k-anonymity.
- **t-closeness**: evitar distribución de valores sensibles que se desvíen significativamente del total.

5.5 Differential Privacy

Agregación de ruido controlado a las salidas analíticas o consultas para preservar la privacidad individual sin sacrificar valor estadístico.

6. Proceso de Protección

1. **Identificación de datos personales** en las bases y flujos de información.
2. **Evaluación de riesgos de reidentificación** según el contexto, volumen, combinabilidad y exposición.
3. **Selección de técnicas de protección** adecuadas al nivel de riesgo y finalidad del tratamiento.

4. **Aplicación técnica** de mecanismos de anonimización o seudonimización.
5. **Documentación** del proceso, incluyendo algoritmos utilizados, parámetros aplicados, y evaluaciones de impacto.
6. **Revisión periódica** de la efectividad de las medidas implementadas.

7. Consideraciones de Implementación

- Los datos anonimizados no deben ser reversibles ni permitir inferencias.
- En ambientes de desarrollo o pruebas, está prohibido el uso de datos personales reales sin medidas de protección.
- Toda exportación de datos debe pasar por revisión de cumplimiento de este protocolo.

8. Control y Auditoría

- Deben establecerse mecanismos de revisión periódica del cumplimiento del protocolo.
- Toda excepción al presente documento debe estar documentada y justificada por el responsable del tratamiento.
- La organización deberá implementar un log de accesos, modificaciones y exportaciones sobre datos protegidos.

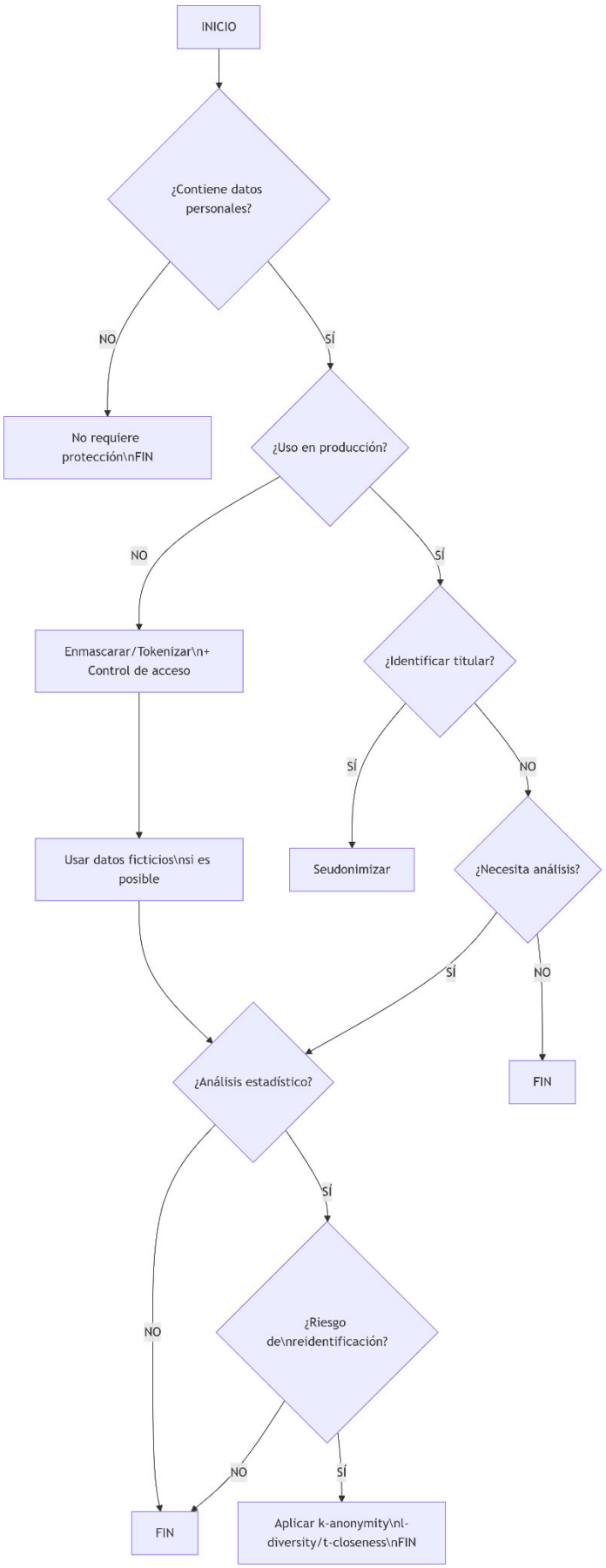

Figura 8. Diagrama de flujo Aplicación de Técnicas de Protección de Datos

Capitulo VII

Conclusiones y recomendaciones:

En este capítulo se resumen los principales desafíos que enfrentan las startups para proteger datos personales y se destacan oportunidades como el uso de herramientas low-code y la necesidad de actualizar las leyes frente al avance de la inteligencia artificial. También se recomienda fortalecer la confianza de los usuarios y aprender de normativas internacionales para mejorar la protección de datos en proyectos pequeños.

A lo largo de este trabajo, pude identificar y analizar los múltiples desafíos que enfrentan las startups y pequeños proyectos a la hora de implementar mecanismos de protección de datos personales, especialmente en contextos con recursos limitados. Me encontré con una dificultad concreta: la falta de acceso a datos reales, debido a políticas de confidencialidad y seguridad, lo que me impidió realizar pruebas más robustas. Para superar esta barrera, decidí utilizar datos sintéticos, lo cual me permitió avanzar, aunque reconozco que este enfoque tiene ciertas limitaciones en cuanto a representatividad y aplicabilidad general.

A partir de esta experiencia, me resultó interesante explorar nuevas líneas de investigación, entre ellas el desarrollo de herramientas low-code para anonimización de datos, pensadas para equipos sin conocimientos técnicos profundos. Estas herramientas, que integran técnicas como k-anonymity, differential privacy y data masking en interfaces visuales, me parecen una alternativa concreta para democratizar el acceso a la privacidad. Su uso representa ventajas claras como la accesibilidad, la reducción de riesgos legales y la posibilidad de escalar soluciones. Sin embargo, también advertí desafíos importantes, como mantener estándares de seguridad y evitar vulnerabilidades.

Durante el desarrollo del trabajo, reflexioné sobre la necesidad de revisar y actualizar la Ley 25.326, especialmente frente al avance de la inteligencia artificial aplicada a procesos de anonimización. La IA puede ser una gran aliada, pero también introduce riesgos como la reidentificación. Por eso, considero fundamental que existan regulaciones claras, que definan conceptos clave, exijan auditorías y garanticen la transparencia.

Además, me pareció relevante comparar cómo se abordan estos desafíos en otras regiones, como la Unión Europea a través del GDPR. Este enfoque comparativo me permitió tomar conciencia de buenas prácticas internacionales que podrían inspirar mejoras locales.

Desde una mirada social, me llamó la atención el impacto que tienen las filtraciones masivas de datos sobre la confianza de los usuarios. La desconfianza creciente puede afectar profundamente la participación digital y la vida social online. Por eso, creo que es importante desarrollar estrategias que restablezcan esa confianza, como mayor transparencia, educación digital y regulaciones más estrictas.

En definitiva, concluyo que si bien persisten obstáculos técnicos, legales y sociales, también existen oportunidades reales para construir soluciones más seguras e inclusivas. Aprendí que integrar tecnologías accesibles, actualizar marcos normativos y promover la conciencia digital son pasos esenciales para proteger la privacidad en un entorno digital cada vez más complejo.

También entiendo que es clave fomentar la adopción de herramientas low-code que faciliten el trabajo de emprendedores y pequeños equipos, y que resulta urgente capacitar a quienes lideran proyectos tecnológicos en buenas prácticas de protección de datos. Desde mi perspectiva, la privacidad no debe ser un elemento secundario, sino un valor que atraviese todo el desarrollo tecnológico.

Finalmente, me convenzo de que las políticas públicas deben evolucionar junto con la tecnología. La cooperación entre Estado, sector privado y sociedad civil puede generar un entorno digital más seguro, donde se respete la privacidad sin frenar la innovación.

6.1 Trabajos Futuros.

A partir del desarrollo del presente proyecto, se identifican diversas líneas de trabajo que podrían abordarse en investigaciones futuras para profundizar en la protección de datos personales en entornos digitales:

Estudio de técnicas de seudoanonimización y su aplicación en pequeños proyectos

La seudoanonimización es una técnica que permite ocultar la identidad de los titulares de datos mediante la sustitución de identificadores directos (como nombre, DNI o dirección de correo) por seudónimos o códigos. A diferencia de la anonimización completa, la seudoanonimización permite, bajo ciertas condiciones y con claves protegidas, revertir parcialmente el proceso. Esto resulta útil en escenarios donde es necesario mantener cierta trazabilidad, por ejemplo, en proyectos educativos, investigaciones longitudinales o aplicaciones donde se requiere personalización sin comprometer la identidad directa.

Futuras investigaciones podrían centrarse en el análisis comparativo entre técnicas de anonimización y seudoanonimización, evaluando su viabilidad en pequeños proyectos, así como los riesgos asociados a su implementación. Además, sería pertinente estudiar cómo gestionar de forma segura las claves de reversión, quién debe tener acceso a ellas y en qué circunstancias.

2. Evaluación de riesgos de reidentificación en entornos reales

Las técnicas de anonimización no garantizan una protección absoluta si los datos anonimizados pueden ser correlacionados con otras fuentes externas para reidentificar a los individuos. En este sentido, un campo de investigación necesario es el desarrollo de estudios empíricos que evalúen la vulnerabilidad a la reidentificación en contextos reales, como bases de datos de redes sociales, formularios web o encuestas abiertas.

Estos estudios permitirían medir la efectividad de distintas técnicas de anonimización según el tipo y volumen de datos, el entorno de publicación (por ejemplo, datos abiertos vs. internos), y los riesgos de correlación cruzada con otras bases públicas. El objetivo sería generar guías prácticas sobre qué nivel de anonimización es adecuado en cada contexto.

3. Desarrollo de herramientas accesibles para la gestión de privacidad

Un área de oportunidad clave consiste en diseñar e implementar herramientas que faciliten a usuarios no especializados la gestión de sus datos personales y la protección de la privacidad. Estas herramientas podrían incluir:

- Interfaces gráficas para aplicar anonimización o seudoanonimización sin conocimientos técnicos.

- Generadores de datos sintéticos para reemplazar bases de datos reales durante pruebas de sistemas.

- Detectores automáticos de datos sensibles en documentos o formularios.

- Asistentes interactivos que ayuden a cumplir con normativas como la Ley 25.326 o el GDPR.

Además, sería relevante investigar la usabilidad, portabilidad y costos de implementación de estas herramientas, especialmente en contextos con recursos limitados.

5 Desarrollar plantillas de arquitectura de software que integren los principios de Privacy by Design (PbD) desde las primeras etapas

Una línea concreta de trabajo consiste en el diseño de plantillas de arquitectura de software orientadas a incorporar los principios de privacidad desde el inicio del ciclo de vida del desarrollo. Estas plantillas podrían servir como base para proyectos nuevos y facilitar la integración de mecanismos como la minimización de datos, el cifrado por defecto, el registro de actividades (logs) con retención limitada, y la separación de datos sensibles por capas.

Además, podrían incluir ejemplos prácticos de flujos de autenticación respetuosos de la privacidad, patrones para el manejo de consentimiento informado y recomendaciones para la segmentación de acceso según perfiles de usuario. Este tipo de recursos serían especialmente valiosos para pequeñas organizaciones, startups o desarrolladores individuales que necesitan orientación clara y reutilizable.

Investigar cómo implementar auditorías automatizadas de privacidad en entornos de desarrollo ágil

Otra área prometedora es la incorporación de auditorías de privacidad automatizadas como parte de los ciclos de integración y despliegue continuo (CI/CD), cada vez más utilizados en desarrollo ágil. Este enfoque permitiría detectar posibles vulnerabilidades o malas prácticas desde etapas tempranas del desarrollo, como la exposición involuntaria de datos sensibles, el almacenamiento sin cifrado o la falta de consentimiento explícito.

Futuras investigaciones podrían abordar la creación de herramientas o plugins que integren estas auditorías con plataformas comunes de desarrollo (por ejemplo, GitHub Actions, Jenkins, GitLab CI). Además, se podrían definir métricas específicas para evaluar el cumplimiento de PbD, como la proporción de campos sensibles debidamente protegidos o la existencia de mecanismos de control de acceso en puntos críticos.

Explorar la formación en Privacy by Design como parte del currículo en carreras de informática, diseño o ingeniería

El concepto de PbD no solo debe implementarse a nivel técnico, sino que también requiere una transformación en la formación profesional. En este sentido, resulta fundamental incorporar contenidos específicos sobre privacidad y diseño ético en los planes de estudio de carreras tecnológicas.

Futuras líneas de trabajo pueden centrarse en el desarrollo de módulos formativos, materiales didácticos y proyectos de aula que permitan a los

estudiantes aplicar los principios de PbD en entornos simulados o reales. También sería importante fomentar la interdisciplinariedad, incluyendo aspectos legales, éticos y de comunicación junto con los aspectos técnicos. La formación en PbD desde etapas tempranas puede contribuir a una generación de profesionales más conscientes de su responsabilidad en la protección de datos.

Beneficios de aplicar Privacy by Design en proyectos pequeños

Aunque el enfoque de PbD suele relacionarse con grandes empresas o proyectos complejos, su aplicación en proyectos pequeños resulta igualmente crucial y viable. Incorporar desde el comienzo prácticas respetuosas de la privacidad puede evitar múltiples problemas futuros, como brechas de seguridad, sanciones regulatorias, reputación negativa y pérdida de usuarios.

Además, diseñar con privacidad en mente desde etapas tempranas generalmente resulta menos costoso que tener que rediseñar sistemas o añadir parches luego de que ocurra un incidente. Por eso, se vuelve estratégico promover la adopción de PbD como un valor añadido y una ventaja competitiva incluso en desarrollos de bajo presupuesto o alcance limitado.

La elaboración de este trabajo me permitió no solo profundizar mis conocimientos en técnicas de protección de datos, sino también enfrentar el desafío de diseñar soluciones adaptables y viables para contextos reales, especialmente en entornos con recursos limitados. A lo largo del proceso, pude aplicar conceptos teóricos abordados durante la carrera en un escenario práctico, lo cual me permitió comprender con mayor claridad las dificultades que enfrentan muchas organizaciones al implementar medidas de privacidad. Esta experiencia me impulsó a pensar de manera creativa y estratégica, priorizando la simplicidad, la eficiencia y el cumplimiento normativo.

Asimismo, el análisis de marcos legales, herramientas tecnológicas y enfoques de diseño ético me permitió desarrollar una mirada integral sobre la protección de datos, reconociendo que no se trata solo de una cuestión técnica, sino también social y humana. Considero que este enfoque me prepara para desempeñarme profesionalmente con responsabilidad ética y técnica frente al tratamiento de la información, siendo consciente del impacto que pueden tener las decisiones de diseño y gestión de datos en la vida de las personas. En definitiva, este proyecto reforzó mi compromiso con el desarrollo de soluciones inclusivas, seguras y respetuosas de los derechos fundamentales, y me brindó

herramientas valiosas para afrontar futuros desafíos profesionales en el campo de la privacidad digital

Bibliografía y Anexos

Leyes y Normativas

- Agencia de Acceso a la Información Pública (AAIP). (2000). *Ley 25.326 - Protección de los Datos Personales*. Recuperado de https://www.argentina.gob.ar/normativa/nacional/ley-25326-45178
- European Parliament and Council. (2016). *Regulation (EU) 2016/679 - General Data Protection Regulation (GDPR)*. Recuperado de https://eur-lex.europa.eu/eli/reg/2016/679/oj
- California Consumer Privacy Act (CCPA). (2018). *California Civil Code Title 1.81.5*. Recuperado de https://oag.ca.gov/privacy/ccpa

Libros

- Solove, D. J. (2008). *Understanding Privacy*. Harvard University Press.
- Zuboff, S. (2019). *The Age of Surveillance Capitalism*. Public Affairs.
- Tene, O., & Polonetsky, J. (2012). *Privacy and Big Data: Making Ends Meet*. Stanford Law Review.
- O'Flaherty, K. (2020). *Privacy Is Power: Why and How You Should Take Back Control of Your Data*. Bantam Press.

Artículos Académicos y Papers

- Narayanan, A., & Shmatikov, V. (2008). *Robust De-anonymization of Large Sparse Datasets*. Proceedings of the IEEE Symposium on Security and Privacy.
- Dwork, C. (2008). *Differential Privacy: A Survey of Results*. Lecture Notes in Computer Science.
- Sweeney, L. (2002). *k-Anonymity: A Model for Protecting Privacy*. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems.
- Ohm, P. (2010). *Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization*. UCLA Law Review.

Informes y Recursos Institucionales

- International Association of Privacy Professionals (IAPP). *Global Privacy and Data Protection Developments*.
- Comisión Europea. (2020). *Data Protection in the EU*.
- Organización de Estados Americanos (OEA). (2019). *Guía de Protección de Datos Personales para América Latina*.

- Pew Research Center. (2019). *Americans and Privacy: Concerned, Confused and Feeling Lack of Control Over Their Personal Information*.

Herramientas y Manuales Técnicos

- ARX Data Anonymization Tool. (s.f.). *Technical Documentation and User Guide*. Recuperado de https://arx.deidentifier.org/
- Microsoft Presidio. (2023). *Privacy Preserving Tools for Text Analysis*. Recuperado de https://microsoft.github.io/presidio/
- Faker Library. (s.f.). *Generating Fake Data for Testing Purposes*. Recuperado de https://faker.readthedocs.io/en/master/

Estudios de Caso y Lecciones Aprendidas

Cambridge Analytica y Facebook:

- The Guardian. (2018). *Cambridge Analytica scandal: Facebook fined for data breaches*. Recuperado de https://www.theguardian.com/technology/2018/jul/11/facebook-fined-500000-pounds-over-cambridge-analytica-scandal
- The New York Times. (2018). *How Trump Consultants Exploited the Facebook Data of Millions*. Recuperado de https://www.nytimes.com/2018/03/17/us/politics/cambridge-analytica-trump-campaign.html

Netflix Prize (2006):

- Narayanan, A., & Shmatikov, V. (2008). *Robust De-anonymization of Large Sparse Datasets (Netflix Prize example)*. Recuperado de https://www.cs.utexas.edu/~shmat/shmat_netflix-final.pdf
- Wired. (2009). *Netflix Spilled Your Brokeback Mountain Secret, Lawsuit Claims*. Recuperado de https://www.wired.com/2009/12/netflix-privacy-lawsuit/

Apple y Differential Privacy:

- Apple. (2017). *Differential Privacy Overview*. Recuperado de https://www.apple.com/privacy/docs/Differential_Privacy_Overview.pdf
- Wired. (2016). *Apple's 'Differential Privacy' Is About Collecting Your Data—But Not Your Data*. Recuperado de https://www.wired.com/2016/06/apples-differential-privacy-collecting-data/

Filtraciones de datos en el sector público (Argentina):

- Clarín. (2021). *Filtración de datos personales: investigan si se violó la privacidad de los beneficiarios de programas sociales*. Recuperado de https://www.clarin.com/politica/filtracion-datos-investigan-violo-privacidad-beneficiarios-planes-sociales_0_2sYXL1Utv.html
- Infobae. (2021). *Grave filtración de datos sensibles: alertan sobre la exposición de beneficiarios de programas sociales*. Recuperado de https://www.infobae.com/politica/2021/07/13/grave-filtracion-de-datos-sensibles-alertan-sobre-la-exposicion-de-beneficiarios-de-programas-sociales/

Startups tecnológicas – Auth0 y tokenización:

- Auth0 Official Blog. (2020). *What is Tokenization and Why Does it Matter?* Recuperado de https://auth0.com/blog/what-is-tokenization-and-why-does-it-matter/
- Okta Official Site. (s.f.). *Adquisición de Auth0*. Recuperado de https://www.okta.com/okta-auth0/

Recursos Adicionales de Privacidad y Anonimización

- CONICET. (s.f.). *Datos de investigación: el proceso de anonimizar*. Recuperado de https://datosdeinvestigacion.conicet.gov.ar/datos-de-investigacion-el-proceso-de-anonimizar/
- Agencia Española de Protección de Datos (AEPD). (2021). *Guía básica de anonimización*. Recuperado de https://www.aepd.es/documento/guia-basica-anonimizacion.pdf
- Brumen, B., Heričko, M., Rozman, I., & Podgorelec, V. (2021). *Survey of anonymization techniques for privacy-preserving data publishing*. Computers & Security, 110, 102403. https://doi.org/10.1016/j.cose.2021.102403
- FasterCapital. (s.f.). *Anonymity - Understanding its importance*. Recuperado de https://fastercapital.com/keyword/anonymity-anonymity.html
- OWASP. (s.f.). *Top 10 Privacy Risks*. Recuperado de https://owasp.org/www-project-top-10-privacy-risks/
- Privacy International. (s.f.). *Privacy Projects and Resources*. Recuperado de https://privacyinternational.org/
- Electronic Frontier Foundation (EFF). (s.f.). *Privacy and Security Resources*. Recuperado de https://www.eff.org/issues/privacy
- Future of Privacy Forum (FPF). (s.f.). *Research and Publications on Data Protection*. Recuperado de https://fpf.org/

Anexo I

La siguiente matriz permite comparar distintas herramientas de anonimización disponibles para pequeños proyectos o startups, evaluando tres dimensiones clave: costo, facilidad de uso y nivel de seguridad. Este tipo de análisis puede ayudar a elegir soluciones que se adapten al contexto específico del proyecto y sus recursos disponibles.

Herramienta	Técnica principal	Costo	Facilidad de uso	Nivel de seguridad
ARX	k-anonymity, l-diversity, t-closeness	Gratuito (open-source)	Media (requiere configuración)	Alta
Microsoft Presidio	Data masking, redacción, tokenización	Gratuito (open-source)	Alta (APIs y CLI)	Media-Alta
Amnesia	k-anonymity, privacy diferencial	Gratuito	Baja (requiere conocimientos técnicos)	Alta
Faker	Datos sintéticos (simulación)	Gratuito	Alta (muy simple de usar)	Media
Privitar	Privacidad diferencial, enmascaramiento	Pago (licencia comercial)	Media	Muy alta
Skyflow	Tokenización y almacenamiento seguro	Pago (SaaS)	Alta (interfaz y API)	Muy alta

Anexo II

Guía Práctica: Selección de Técnicas de Anonimización según el Tipo de Dato

Esta guía proporciona una orientación básica para seleccionar la técnica de anonimización más adecuada según el tipo de dato a tratar. El objetivo es ayudar a quienes desarrollan sistemas o gestionan información a aplicar métodos eficientes y proporcionales al nivel de sensibilidad y riesgo de cada categoría de dato.

Tipo de dato	Ejemplos	Técnicas recomendadas	Observaciones
Identificadores directos	Nombre completo, DNI, email, número de teléfono	Supresión, Tokenización, Pseudonimización	Son los datos más sensibles; deben eliminarse o transformarse en casi todos los casos.
Identificadores indirectos (quasi-identificadores)	Fecha de nacimiento, género, código postal	k-anonymity, generalización, supresión parcial	Permiten la reidentificación combinada; aplicar según el contexto y frecuencia.
Datos sensibles	Salud, religión, orientación sexual, opiniones políticas	Privacidad diferencial, enmascaramiento, agregación	Deben ser tratados con extremo cuidado; a veces se sugiere no recolectarlos.
Datos transaccionales	Montos, fechas de compra, productos adquiridos	Perturbación, binning, agregación temporal	Suelen ser útiles para análisis; proteger sin perder valor estadístico.
Datos sintéticos	Usuarios ficticios, datos de prueba	Generación aleatoria (Faker), simulación basada en reglas	No requieren anonimización adicional; ideales para desarrollo o testeo.