MEMORIA DE LA CANDIDATURA

Premio de Privacidad y Protección de Datos en Iberoamérica – AEPD

1. Identificación de la candidatura

Título

Privacy AI Studio: gobernanza y privacidad por diseño para el uso seguro de inteligencia artificial generativa en Iberoamérica

Entidad responsable del diseño:
Laboratorio de Innovación e Inteligencia Artificial (IALAB)
Facultad de Derecho – Universidad de Buenos Aires

Entidad desarrolladora tecnológica:
Puzzle AI Agents

2. Descripción del proyecto

Privacy AI Studio es una plataforma tecnológica diseñada para permitir a organizaciones públicas y privadas utilizar inteligencia artificial generativa sobre documentos, audios y videos que contienen datos personales y sensibles, sin poner en riesgo la privacidad, la legalidad ni la rendición de cuentas institucional.

El proyecto surge como respuesta a uno de los principales desafíos contemporáneos en materia de protección de datos: el fenómeno del *Shadow AI*, es decir, el uso no gobernado de herramientas de IA por parte de empleados que, ante la falta de soluciones institucionales seguras, recurren a plataformas externas para procesar información sensible.

Lejos de prohibir la IA o tolerar su uso informal, Privacy AI Studio ofrece una alternativa estructural: una infraestructura que integra anonimización, validación humana significativa, trazabilidad y gobernanza embebida, permitiendo aprovechar los beneficios de la IA generativa sin comprometer los derechos de las personas.

3. Objetivos

El proyecto se estructura en torno a tres objetivos estratégicos:

1. **Proteger efectivamente los datos personales** mediante mecanismos de privacidad por diseño y por defecto que impidan la exposición indebida de información identificable.

2. **Habilitar el uso legítimo y útil de la inteligencia artificial** en sectores sensibles como justicia, salud, educación y administración pública.

3. **Transformar principios jurídicos y éticos en reglas técnicas ejecutables**, asegurando proporcionalidad, trazabilidad, supervisión humana y rendición de cuentas institucional.

4. Funcionamiento general

Privacy AI Studio organiza el tratamiento de la información en un flujo gobernado que incluye:

- Ingesta controlada de documentos, audios y videos.
- Transcripción local supervisada de audios y videos.
- Curado y anonimización de texto.
- Validación humana obligatoria.
- Análisis mediante inteligencia artificial únicamente sobre información previamente protegida.

Este diseño garantiza que ningún dato personal sea procesado por modelos de IA sin haber pasado antes por controles de protección de datos y verificación humana.

5. Marco normativo y ético

La plataforma se alinea con:

- Normativa de protección de datos personales como la Ley 25.326 de Argentina, asegurando finalidad, minimización, seguridad y control de transferencias.
- La Recomendación de la UNESCO sobre la Ética de la Inteligencia Artificial (2021), incorporando privacidad por diseño, proporcionalidad en el uso de la IA, supervisión humana significativa, responsabilidad, rendición de cuentas y gobernanza.

Estos principios éticos no se aplican como políticas abstractas, sino que se traducen en reglas técnicas embebidas, que se ejecutan automáticamente dentro del sistema.

6. Casos de uso y estado del proyecto

Privacy AI Studio se encuentra actualmente en fase de despliegue e implementación piloto en dos entornos reales:

- ZLT (empresa privada)
- Procuración General de la Ciudad Autónoma de Buenos Aires (sector público-justicia)

En ambos casos, la plataforma se utiliza para procesar documentos, audios y videos con datos sensibles dentro de un entorno institucional gobernado y trazable.

7. Impacto

El impacto de la solución es doble:

- **Operativo:** habilita el uso institucional de IA generativa sobre información en organizaciones que antes no podían utilizarla o lo hacían de forma informal y riesgosa.

- **Normativo y ético:** reduce estructuralmente el riesgo de exposición de datos personales, facilita el cumplimiento normativo y refuerza la rendición de cuentas institucional.

La plataforma no solo mejora la eficiencia, sino que cambia el modo en que las organizaciones se relacionan con la IA, sustituyendo el Shadow AI por gobernanza ética y tecnológica.

8. Proyección iberoamericana

Privacy AI Studio fue concebida como una infraestructura replicable para Iberoamérica. Ha sido presentada en redes de innovación judicial en octubre de 2025, generando interés de distintos países. Su arquitectura permite su adopción por gobiernos, universidades y empresas en contextos regulatorios diversos, respetando los principios comunes de protección de datos y ética de la IA.

9. Valor para el Premio AEPD

La candidatura aporta una innovación singular: convierte la protección de datos y los principios éticos de la IA en reglas técnicas ejecutables que se aplican por defecto en el uso cotidiano de la inteligencia artificial.

Privacy AI Studio no es una herramienta aislada, sino una infraestructura de confianza para la adopción responsable de IA en Iberoamérica, alineada con los objetivos de la Agencia Española de Protección de Datos y con los estándares internacionales en la materia.

Privacy AI Studio by Puzzle AI Agents

Diseñado por IALAB

Gobernanza y privacidad por diseño para el uso seguro de IA generativa en Iberoamérica

Resumen ejecutivo

La expansión acelerada de la inteligencia artificial generativa ha abierto una nueva frontera de productividad para organizaciones públicas y privadas. Sin embargo, este avance también ha generado un riesgo silencioso y creciente: el uso no gobernado de estas tecnologías dentro de las propias instituciones.

Este fenómeno, conocido como *Shadow AI*, se produce cuando equipos operativos recurren a herramientas externas para procesar documentos, audios, bases de datos o información estratégica sin controles institucionales ni salvaguardas de protección de datos. Las consecuencias incluyen exposición de información de identificación personal, incumplimientos normativos, pérdida de trazabilidad y afectación de derechos de las personas.

En este contexto, prohibir la IA no es una solución viable, e ignorar su uso espontáneo resulta aún más peligroso. La única respuesta sostenible es gobernar su utilización desde el diseño, mediante infraestructuras que permitan innovar con inteligencia artificial sin sacrificar la privacidad, la legalidad ni la responsabilidad institucional.

Privacy AI Studio, diseñada por el Laboratorio de Innovación e Inteligencia Artificial de la Facultad de Derecho de la Universidad de Buenos Aires y desarrollada por Puzzle AI Agents, responde a este desafío mediante una plataforma que permite procesar documentos, audios y videos que contienen datos personales en entornos seguros, trazables y alineados con los más altos estándares jurídicos y éticos.

La solución fue concebida como una infraestructura modular que integra anonimización, validación humana significativa y análisis mediante inteligencia artificial gobernada, todo dentro de entornos controlados por las propias organizaciones. De este modo, la plataforma garantiza que la información sólo sea utilizada tras haber sido debidamente anonimizada.

Su impacto se manifiesta en dos planos complementarios. En el plano operativo, habilita el uso institucional y gobernado de IA generativa en sectores donde antes estaba prohibido o se realizaba de forma informal y riesgosa. En el plano normativo y

ético, materializa los principios de la protección de datos y de la Recomendación de la UNESCO sobre la Ética de la Inteligencia Artificial (2021), incluyendo la privacidad por diseño, la proporcionalidad, la supervisión humana significativa y la rendición de cuentas.

Lejos de ser una herramienta aislada, Privacy AI Studio constituye una estrategia tecnológica de largo plazo para gobiernos, sistemas judiciales, organizaciones públicas y empresas que necesitan incorporar inteligencia artificial sin erosionar la confianza institucional ni el derecho a la privacidad y protección de datos personales.

Gobernar el uso de la IA desde dentro de las organizaciones es posible. Privacy AI Studio lo hace realidad.

1. Introducción

Organizaciones públicas y privadas de todos los sectores (justicia, salud, educación, administración, banca, seguros) gestionan a diario grandes volúmenes de información que contiene datos personales, comerciales y sensibles que deben protegerse: expedientes, denuncias, contratos, grabaciones de audiencias, historiales clínicos, entre muchos otros.

Proteger esa información ya no es solo una obligación legal. Es también una condición indispensable para garantizar la confianza institucional, la continuidad operativa y la legitimidad ante la ciudadanía o los clientes. En ese escenario, la inteligencia artificial (especialmente los modelos de lenguaje) aparece como una oportunidad para ganar eficiencia, pero también como un riesgo si no se implementan medidas adecuadas de resguardo.

El uso no gobernado de plataformas de IA generativa por parte de los equipos de trabajo puede derivar en exposiciones accidentales de datos, filtraciones por reentrenamiento de modelos de terceros, o pérdida de control sobre los registros y flujos de información. Lo que debería acelerar procesos, puede terminar comprometiendo derechos, reputaciones, intereses de las empresas y generando responsabilidades legales.

Este desafío es aún más urgente ante el crecimiento del fenómeno conocido como *Shadow AI*, donde equipos dentro de las propias organizaciones comienzan a utilizar herramientas de inteligencia artificial por fuera de los circuitos autorizados, exponiendo sin saberlo información sensible y comprometiendo el cumplimiento normativo.

Frente a este desafío, el Laboratorio de Innovación e Inteligencia Artificial (IALAB) de la Facultad de Derecho de la Universidad de Buenos Aires diseñó una respuesta concreta: una plataforma que combina automatización, trazabilidad y control humano, integrando desde el inicio los principios de privacidad por diseño y por defecto, proporcionalidad en el uso de la IA, gobernanza humana significativa, responsabilidad y rendición de cuentas.

Este desarrollo no es una herramienta puntual, sino una infraestructura modular que propone otra forma de operar con datos: una lógica en la cual la privacidad se incorpora desde el diseño de los procesos, las arquitecturas técnicas y la gobernanza organizacional. En lugar de añadir capas de protección *a posteriori*, se construye desde el principio un entorno donde los datos se tratan de forma segura, auditable y conforme al marco legal.

El objetivo de este documento es compartir esa experiencia: cómo se diseñó, cómo funciona y qué principios la sostienen. A lo largo de las próximas secciones, se describe una solución que busca resolver un problema urgente y creciente, pero sin resignar valores fundamentales. Una propuesta para demostrar que es posible innovar con inteligencia artificial sin poner en riesgo la privacidad de las personas.

2. El riesgo invisible: IA en las sombras (Shadow AI)

Actualmente, uno de los mayores desafíos de privacidad y seguridad no proviene de amenazas externas, sino de adentro de las propias organizaciones. El fenómeno conocido como Shadow AI se refiere al uso no autorizado o no supervisado de herramientas de inteligencia artificial (particularmente IA generativa) por parte de equipos que, ante la falta de alternativas institucionales, recurren a soluciones públicas, gratuitas o comerciales sin control.

Casos típicos incluyen:

- Un abogado junior que carga un escrito judicial en un asistente gratuito para “mejorar la redacción”.
- Un médico residente que transcribe una historia clínica en un traductor automático para generar un informe en otro idioma.
- Un analista de RR. HH. que sube legajos de empleados a un generador de descripciones de puestos.
- Un contador que copia balances contables en una plataforma abierta para obtener un resumen financiero.

Este uso aparentemente inofensivo representa, en realidad, un escenario crítico:

- Fuga de información hacia proveedores que pueden retener, indexar, reentrenar modelos con esos datos, reservarse otros posibles usos comerciales y compartirlos con proveedores, afiliados o autoridades.
- Incumplimientos regulatorios con consecuencias legales severas (Ley 25.326, GDPR, violación del secreto profesional, fiscal y financiero).
- Pérdida de propiedad intelectual, cuando diseños, algoritmos o estrategias se diluyen en modelos ajenos.
- Outputs inseguros y no auditables, que comprometen la calidad y la responsabilidad de las decisiones.

La paradoja es clara: la productividad que ofrecen estas herramientas es inmediata, pero también lo es el riesgo cuando no existen políticas claras ni soluciones internas confiables. En todos estos casos, los datos de terceros (pacientes, denunciantes, trabajadores, clientes o ciudadanos) quedan expuestos sin su conocimiento ni consentimiento.

Prohibir la IA no es una solución viable. Ignorar su uso espontáneo es aún más riesgoso. La única alternativa sustentable es gobernar su uso con inteligencia, ofreciendo entornos seguros, auditables y éticamente alineados donde los equipos puedan operar con IA sin poner en riesgo la información.

La Plataforma de Privacidad fue diseñada precisamente para responder a este desafío: brindar una infraestructura que permita usar inteligencia artificial de forma segura, trazable y conforme a derecho, evitando así que las organizaciones queden expuestas a los riesgos del Shadow AI.

Prohibir la IA no es una solución viable. Ignorar su uso espontáneo es aún más riesgoso. La única alternativa sustentable es gobernar su uso con inteligencia, ofreciendo entornos seguros, auditables y éticamente alineados donde los equipos puedan operar con IA sin poner en riesgo los derechos de las personas ni la integridad de la información.

La Plataforma de Privacidad fue diseñada precisamente para responder a este desafío: brindar una infraestructura que permita usar inteligencia artificial de forma segura, trazable y conforme a derecho, transformando los principios de protección de datos en garantías técnicas efectivas frente al fenómeno del Shadow AI.

3. Privacy AI Studio arquitectura, módulos y principios de diseño

3.1 Origen y objetivos del proyecto

La Plataforma de Privacidad surge como una respuesta técnico-organizativa a los desafíos estructurales que enfrentan las organizaciones al procesar documentos, audios y videos que contienen datos personales y sensibles.

Para todas las organizaciones, el crecimiento del uso de inteligencia artificial generativa ha ampliado de forma significativa las capacidades de análisis y automatización, pero también ha incrementado los riesgos de exposición de datos, pérdida de trazabilidad y cumplimiento normativo insuficiente.

Frente a este escenario, el equipo interdisciplinario del Laboratorio de Innovación e Inteligencia Artificial (IALAB) de la Facultad de Derecho de la Universidad de Buenos Aires diseñó, y el equipo de desarrollo de la firma argentina Puzzle AI Agents desarrolló, una solución integral orientada a garantizar el respeto efectivo de la normativa de protección de datos personales y la alineación con los principios éticos establecidos por la Recomendación sobre la Ética de la Inteligencia Artificial de la UNESCO (2021).

El proyecto se estructura en torno a tres objetivos principales:

- **Operacionalizar la privacidad por diseño y por defecto**, incorporando salvaguardas técnicas y organizativas para que el tratamiento de datos contenidos en documentos, audios y videos esté regido por reglas embebidas en la arquitectura del sistema.

- **Permitir el uso útil de IA sin comprometer información de identificación personal**, facilitando el análisis, la búsqueda y el procesamiento automatizado de contenidos previamente anonimizados con supervisión humana significativa.

- **Ofrecer una infraestructura para gobernanza del uso de la IA**, que combine procesamiento de información en entornos controlados, mecanismos de

anonimización, validación humana y análisis con inteligencia artificial bajo reglas de gobernanza institucional definidas por la organización.

Este enfoque permite que las organizaciones innoven con inteligencia artificial sin trasladar los riesgos a los titulares de los datos ni depender del comportamiento individual de los usuarios.

3.2 Arquitectura funcional y flujo de tratamiento de datos

La arquitectura de la Plataforma de Privacidad fue concebida para garantizar un tratamiento seguro, trazable y gobernado de documentos, audios y videos que puedan contener información personal o sensible.

En términos funcionales, la plataforma organiza el procesamiento de la información en una secuencia de capas que aseguran que ningún contenido sea utilizado con inteligencia artificial sin haber pasado previamente por controles de protección de datos y validación humana:

1. **Ingesta y normalización de contenidos.** La plataforma admite documentos y archivos audiovisuales.

2. **Transcripción supervisada.** En el caso de audios o videos, el contenido es convertido a texto en un entorno controlado y presentado al operador con asistencia para su revisión y validación.

3. **Curado y anonimización.** Todo contenido textual es sometido a procesos de detección y neutralización de información identificable, bajo reglas configurables por área de trabajo y con revisión humana obligatoria.

4. **Validación humana significativa.** Ningún contenido se considera apto para análisis posterior sin una validación expresa por parte de operadores autorizados.

5. **Análisis con inteligencia artificial gobernada.** Sólo los textos previamente curados y anonimizados pueden ser utilizados para análisis, extracción o generación mediante herramientas de IA, bajo reglas de uso definidas institucionalmente.

Este diseño por capas asegura que la privacidad y la protección de datos no dependan de decisiones individuales, sino de reglas técnicas que se ejecutan por defecto.

3.3 Núcleo de anonimización y curado de información

El núcleo de anonimización constituye el componente central de la Plataforma de Privacidad. Su función es eliminar o neutralizar información de identificación personal en documentos y transcripciones, preservando al mismo tiempo la utilidad del contenido para su análisis posterior.

Este componente opera exclusivamente en entornos controlados por la organización y se organiza por proyectos, lo que permite adaptar las reglas de anonimización a distintos contextos jurídicos, administrativos o clínicos.

Entre sus funciones principales se incluyen:

- **Detección automática de información de identificación personal** mediante motores lingüísticos ejecutados localmente.

- **Aplicación de reglas configurables por área de trabajo** que permiten ajustar el nivel y tipo de anonimización requerido.

- **Generación de versiones curadas del contenido**, que mantienen la estructura y el significado del texto sin exponer identidades.

- **Registro de todas las operaciones**, de modo que cada proceso pueda ser auditado y reconstruido.

- **Validación humana obligatoria**, que permite revisar, corregir y aprobar los resultados antes de cualquier uso posterior.

Este núcleo permite reducir drásticamente el riesgo de exposición de datos personales, al tiempo que mantiene la utilidad de la información para fines legítimos de análisis y gestión.

3.4 Transcripción de contenidos audiovisuales bajo supervisión humana

La Plataforma de Privacidad extiende sus garantías de protección de datos al tratamiento de audios y videos mediante un módulo de transcripción que opera en entornos controlados y bajo supervisión humana significativa.

Los contenidos audiovisuales son convertidos a texto dentro de la infraestructura de la organización y presentados al operador con herramientas de apoyo visual que facilitan la detección de posibles errores o ambigüedades. El operador revisa, corrige y valida la transcripción antes de que esta ingrese al flujo de curado y anonimización.

Este enfoque asegura que ningún texto derivado de audio o video sea utilizado sin control humano ni sin pasar por los mismos mecanismos de protección de datos que los documentos escritos.

3.5 Análisis de texto curado mediante funciones gobernadas y modelos de lenguaje

Una vez que el contenido ha sido curado y anonimizado, la plataforma habilita su análisis mediante dos mecanismos complementarios:

- **Funciones preconfiguradas de análisis** (prompts-clic), que permiten realizar tareas como resúmenes, extracción de información, validaciones normativas o detección de inconsistencias de manera accesible para usuarios no técnicos.

- **Conexiones controladas a modelos de lenguaje**, que pueden ser habilitadas por área de trabajo cuando se requiere un procesamiento más avanzado,

siempre sobre texto previamente anonimizado y bajo reglas de privacidad, trazabilidad y gobernanza institucional.

Todas las interacciones quedan registradas para fines de auditoría, supervisión y rendición de cuentas, garantizando que el uso de inteligencia artificial sea transparente y verificable.

3.6 Principios de arquitectura y gobernanza por diseño

El diseño de la Plataforma de Privacidad se rige por un conjunto de principios que orientan su arquitectura y su operación:

A. Control institucional de la infraestructura

La plataforma opera en entornos controlados por la organización, lo que permite evitar dependencias de nubes públicas y reducir riesgos de transferencia internacional de datos personales sin garantías adecuadas.

B. Trazabilidad y auditabilidad

Cada operación sobre los datos queda registrada, permitiendo reconstruir quién hizo qué, cuándo y bajo qué condiciones.

C. Modularidad y escalabilidad

Los distintos componentes pueden activarse o desactivarse según el área de trabajo, lo que facilita la adaptación a distintos sectores y capacidades institucionales.

D. Usabilidad con supervisión humana significativa

La plataforma está diseñada para usuarios no técnicos, pero exige revisión humana en todos los puntos críticos del proceso.

E. Seguridad y privacidad por diseño

Las medidas de protección están integradas en la arquitectura misma del sistema, de modo que la privacidad no depende del comportamiento individual.

F. Gobernanza por roles y permisos

El acceso y las funciones están diferenciados según responsabilidades, garantizando separación de funciones, control institucional y rendición de cuentas.

4. Compliance architecture by design: fundamentos éticos y normativos de la Plataforma de Privacidad

La arquitectura de la Plataforma de Privacidad fue concebida para materializar, en reglas técnicas y organizativas ejecutables, los principios éticos y jurídicos que rigen el tratamiento de datos personales y el uso responsable de la inteligencia artificial.

En particular, integra los estándares de la Recomendación de la UNESCO sobre la Ética de la Inteligencia Artificial (2021) y los requisitos de la Ley 25.326 de Protección de Datos Personales de la República Argentina, bajo un enfoque de cumplimiento desde el diseño (*compliance by design*).

4.1. Supervisión humana significativa

La plataforma garantiza que toda operación crítica sobre información personal esté sujeta a supervisión humana significativa. Ningún documento, audio o transcripción puede ser procesado con inteligencia artificial sin haber pasado previamente por procesos de curado, anonimización y validación por parte de operadores autorizados.

Esta supervisión no es un control formalista, sino un componente central de protección de derechos y de calidad del proceso, diseñado para integrarse de forma eficiente en el flujo de trabajo. La plataforma asiste activamente a los operadores mediante herramientas de apoyo visual y flujos guiados que facilitan la revisión, sin eliminar la responsabilidad humana.

De este modo, la automatización no sustituye el control humano, sino que lo refuerza y lo hace verificable.

4.2. Protección de la privacidad y minimización del riesgo

La protección de los datos personales se garantiza mediante una combinación de medidas técnicas y organizativas que permiten cumplir los principios de finalidad, licitud, minimización y seguridad establecidos en múltiples normativas de protección de datos.

Toda la información es sometida a procesos de anonimización automática y curado humano antes de cualquier análisis con inteligencia artificial, lo que reduce de manera estructural el riesgo de identificación, reidentificación o exposición indebida. La plataforma permite además adaptar estos controles al contexto específico de cada proyecto, en función de la sensibilidad de los datos y de la finalidad del tratamiento.

Este enfoque evita que la protección de la privacidad dependa de decisiones individuales y la integra como una propiedad inherente del sistema.

4.3. Gobernanza embebida

La Plataforma de Privacidad implementa un modelo de gobernanza embebida: las reglas de uso, los límites y las salvaguardas están incorporados en la arquitectura misma del sistema y no pueden ser omitidos por los usuarios.

Esto implica que:

- solo puede utilizarse inteligencia artificial sobre contenidos previamente anonimizados,
- todas las operaciones quedan registradas y son auditables,
- y los flujos de datos están controlados por reglas institucionales definidas por administradores responsables.

De este modo, la plataforma traslada la responsabilidad desde conductas individuales hacia un esquema institucional verificable, alineado con los principios de rendición de cuentas y control.

4.4. Proporcionalidad en el uso de la inteligencia artificial

La plataforma aplica el principio de proporcionalidad en el uso de la IA, eligiendo el tipo de tecnología adecuada para las distintas funciones del sistema.

Las funciones estructurales de protección de datos (como la detección y neutralización de información identificable) se realizan mediante herramientas diseñadas para ofrecer

resultados estables y verificables (no se utiliza IA generativa). Las capacidades generativas se utilizan únicamente cuando son necesarias, siempre bajo supervisión humana y sobre información previamente protegida.

Además, cada área de trabajo institucional define qué funciones y qué niveles de automatización están habilitados, asegurando que el uso de IA sea adecuado al contexto y al riesgo.

4.5. Responsabilidad y rendición de cuentas

La Plataforma de Privacidad garantiza que toda operación pueda ser atribuida, reconstruida y auditada. Para ello, registra de forma sistemática las acciones realizadas, los usuarios involucrados, los proyectos asociados y los resultados obtenidos.

El diseño incorpora separación de funciones, controles de acceso y mecanismos de auditoría que permiten verificar el cumplimiento normativo, investigar incidentes y ejercer supervisión institucional o externa cuando sea necesario.

Asimismo, la arquitectura impide técnicamente que datos no protegidos sean enviados a sistemas de inteligencia artificial, asegurando un cumplimiento automático y no delegable al comportamiento individual.

4.6. Enfoque de gestión de riesgos

La plataforma adopta un enfoque preventivo de gestión de riesgos, identificando y mitigando desde el diseño los principales peligros asociados al uso de inteligencia artificial con datos personales, tales como:

- errores o distorsiones en la transcripción,
- fallas en la anonimización,
- exposición indebida de información sensible,

- y falta de trazabilidad ante auditorías o incidentes.

Estos riesgos se mitigan mediante una combinación de controles técnicos, validación humana obligatoria y mecanismos de auditoría, lo que permite operar con inteligencia artificial de forma segura, transparente y jurídicamente responsable.

5. Cierre

La investigación y el desarrollo plasmados en Privacy AI Studio, impulsado por el Laboratorio de Innovación e Inteligencia Artificial (IALAB) de la Universidad de Buenos Aires y desarrollado por Puzzle AI Agents, constituyen un aporte estratégico en la intersección entre derecho, ética y tecnología.

A lo largo de este documento se ha demostrado que el uso de inteligencia artificial generativa sobre información sensible plantea riesgos estructurales (exposición indebida de datos, pérdida de trazabilidad, transferencias internacionales sin garantías, entre otros) que no pueden abordarse únicamente mediante políticas o buenas prácticas individuales, sino que requieren infraestructuras institucionales de protección.

Este desafío es especialmente crítico frente al crecimiento del fenómeno conocido como *Shadow AI*: el uso informal y no gobernado de herramientas de IA dentro de las propias organizaciones, que expone datos personales y compromete el cumplimiento normativo sin que exista visibilidad ni control. Privacy AI Studio responde precisamente a esta problemática, ofreciendo una alternativa institucional que permite aprovechar el potencial de la IA sin sacrificar la privacidad ni la legalidad.

La plataforma constituye una respuesta concreta y operativa basada en anonimización, validación humana y gobernanza embebida, que permite tratar documentos, audios y videos de forma segura y conforme a los principios de privacidad por diseño y por defecto. Su arquitectura modular y su control institucional aseguran la soberanía de los datos y habilitan la integración responsable de inteligencia artificial en entornos sensibles.

Desde una perspectiva ética y jurídica, la solución se alinea plenamente con la Recomendación sobre la Ética de la Inteligencia Artificial de la UNESCO (2021), al garantizar:

- **Supervisión humana significativa**, evitando automatizaciones opacas o irresponsables.

- **Protección efectiva de la privacidad y de los datos personales.**

- **Proporcionalidad en el uso de la IA**, seleccionando las tecnologías adecuadas a cada finalidad.

- **Responsabilidad y rendición de cuentas institucional**, mediante trazabilidad y control verificables.

Para los responsables públicos y privados, el mensaje central es claro: innovar con inteligencia artificial en ámbitos sensibles (como la justicia, la salud, la educación o la administración pública) es viable y sostenible cuando se construye sobre una arquitectura de privacidad, gobernanza y ética incorporadas desde el diseño. No se trata de frenar la innovación, sino de hacerla digna de confianza.

Privacy AI Studio demuestra que es posible combinar eficiencia operativa, cumplimiento normativo y alineación con estándares internacionales en una única infraestructura. Su adopción por organizaciones públicas y privadas ofrece un modelo replicable para Iberoamérica de cómo enfrentar los desafíos de la IA generativa sin erosionar la privacidad.

En definitiva, la conclusión es doble: lo correcto puede automatizarse y escalarse, y la confianza social en la inteligencia artificial sólo se construye cuando las salvaguardas dejan de depender de la buena voluntad individual y pasan a formar parte de la política institucional y de la infraestructura tecnológica.

ANEXO — Impacto, despliegues y proyección regional

Privacy AI Studio – Premio de Privacidad y Protección de Datos en Iberoamérica (AEPD)

1. Casos de uso y despliegues

Privacy AI Studio se encuentra actualmente en fase de implementación piloto en entornos reales, tanto del sector privado como del sector público, con datos de alta sensibilidad y exigencias normativas estrictas.

Institución / Organización	País	Sector	Tipo de datos	Módulos utilizados	Estado
ZLT	Argentina	Empresa	Documentos, audios, videos	Anonimizador, Transcripción, Prompts-clic, Conectores LLM	Piloto
Procuración General de la Ciudad Autónoma de Buenos Aires	Argentina	Gobierno / Justicia	Documentos, audios, videos	Anonimizador, Transcripción, Prompts-clic, Conectores LLM	Piloto (despliegue inicial 2026)

En ambos casos, la plataforma se utiliza para procesar información que contiene datos personales, habilitando el uso de inteligencia artificial generativa dentro de un entorno gobernado, trazable y conforme a la normativa de protección de datos.

2. Impacto operativo y normativo

2.1 Cambio estructural frente al “Shadow AI”

Antes de la implementación de Privacy AI Studio, las organizaciones enfrentaban una de estas dos situaciones:

- **No podían usar IA generativa** sobre sus documentos y audios debido al riesgo legal y de privacidad, perdiendo oportunidades de eficiencia y análisis.

- **O bien se utilizaba IA en forma informal y no autorizada** (*Shadow AI*), con empleados subiendo documentos con información de identificación personal a plataformas comerciales sin protección.

La plataforma produce un **cambio de régimen**: convierte un uso riesgoso, informal y opaco en un uso institucional, gobernado, trazable y conforme a derecho.

Este impacto es especialmente relevante para organizaciones públicas y empresas reguladas, que de otro modo no podrían adoptar IA generativa sin incurrir en riesgos legales.

2.2 Métricas de impacto (fase piloto)

Aunque los pilotos están en fase inicial, ya se observan impactos claros:

Dimensión	**Situación previa**	**Con Privacy AI Studio**
Uso de IA generativa	No permitido o uso informal (Shadow AI)	Uso institucional autorizado y gobernado
Riesgo de fuga de datos	Alto y no monitoreado	Mitigado por anonimización, control de endpoints y trazabilidad

Capacidad de procesar documentos sensibles con IA	Prácticamente nula	Habilitada previa anonimización, curado y validación humana
Proyectos con IA activa	Se veían limitados por la dificultad de anonimizar documentos de manera manual	Proyectos piloto en fase de diseño (ej. Trabajo con sumarios sancionatorios)
Usuarios activos	Uso oculto	Inicio controlado con expansión progresiva de usuarios
Incidentes conocidos de fuga de datos	No detectables por falta de trazabilidad	Detectables cuando se utilice la plataforma

Más que un simple ahorro de tiempo, el impacto principal es habilitar el uso seguro de IA generativa bajo una política de privacidad por diseño.

3. Proyección iberoamericana y cooperación

Privacy AI Studio fue concebida desde su diseño como una infraestructura replicable para Iberoamérica, tanto por su arquitectura técnica como por su marco normativo y ético.

- **Interés regional:** la solución fue presentada en octubre de 2025 en un evento de innovación en Poderes Judiciales de Iberoamérica, con muy alta aceptación y consultas de distintos países sobre su adopción.

- **Diseño multi-jurisdiccional:** La arquitectura no depende de normas locales particulares, sino que se apoya en principios comunes (protección de datos,

privacidad por diseño, UNESCO 2021), lo que facilita su despliegue en otros países.

- **Modelo exportable:** puede instalarse 100 % on-premise o en nube privada, lo que la hace compatible con los requisitos de soberanía de datos de distintos Estados y organizaciones.

- **Red académica e institucional:** IALAB integra redes de jueces, universidades y laboratorios de innovación pública en Iberoamérica, lo que facilita la transferencia del modelo en conjunto con Puzzle AI Studio.

4. Línea de tiempo del proyecto

Hito	**Fecha**
Inicio del diseño y desarrollo	**Julio 2025**
Finalización de arquitectura base	**Octubre 2025**
Primeros pilotos reales	**Enero 2026**
Estado actual	Piloto en empresa (ZLT) y despliegue inicial en sector público (Procuración General CABA)

Este encuadre temporal se encuentra plenamente dentro del período exigido por la convocatoria del Premio AEPD.

5. Valor diferencial

Privacy AI Studio no es una promesa ni un prototipo académico: es una infraestructura de software viva, que permite a organizaciones públicas y privadas:

- cumplir con la normativa de protección de datos,
- desincentivar el uso no autorizado de IA, y
- habilitar innovación con IA generativa sin sacrificar privacidad, trazabilidad ni responsabilidad institucional.

Lo que distingue a Privacy AI Studio es que traduce principios jurídicos y éticos abstractos (como la privacidad por diseño y por defecto, la protección de datos personales, la proporcionalidad en el uso de la IA, la supervisión humana significativa, rendición de cuentas y la trazabilidad, en reglas técnicas concretas que se ejecutan automáticamente dentro del software.

De este modo, estos principios dejan de depender del comportamiento individual de los usuarios y pasan a integrarse en la arquitectura misma del sistema, convirtiéndose en garantías operativas verificables.

Esto convierte a Privacy AI Studio en una solución especialmente valiosa para Iberoamérica: una región que necesita incorporar inteligencia artificial en sectores sensibles (como justicia, salud, educación y administración pública) sin erosionar la confianza institucional ni los derechos fundamentales de las personas.