Propuesta de innovación

PrevIA (v0.1.0)

Proyecto de innovación para uso responsable de IA en la Administración

PRUEBA DE EVALUACIÓN (PE)
INAP (ESTRATEGIAS DE APRENDIZAJE EN LAS ORGANIZACIONES PÚBLICAS)

Versión	**1.0 (Prototipo MVP)**
Fecha	7 de noviembre de 2025
Autor / Unidad	Sara Cubero García-Conde
Contacto	sara.cgc.94@gmail.com

Contenido

1. Resumen ejecutivo

El personal público y profesional comparte textos en herramientas de IA para apoyo a la redacción y análisis. Sin un paso previo de preparación segura, existe riesgo de exposición de datos personales (PII) y de información sensible. PrevIA es una extensión del navegador que analiza, seudonimiza, generaliza y anonimiza texto de forma 100 % local, con revisión asistida y personalización por tipo de dato. La propuesta incluye marco de calidad, indicadores y plan de implantación.

Problema	Riesgo de filtración de datos al usar IA generativa sin preparar el texto.
Solución	Extensión local con detección y transformación de PII antes de compartir.
Beneficios	Confianza digital, cumplimiento reforzado, hábitos seguros en el flujo de trabajo.
Evidencia	Prototipo funcional en Firefox/Edge, compatible manualmente en otros navegadores.

2. Contexto y alineamiento institucional

La organización aplica gestión por procesos (PDCA) para mejorar de forma continua y asegurar calidad orientada a resultados, con procesos mapeados, indicadores y revisiones periódicas.

El aprendizaje en el flujo de trabajo se impulsa con mentorización y comunidades de práctica, compartiendo casos y plantillas para acelerar la adopción y la transferencia de conocimiento.

Se refuerza el gobierno del dato y una cultura de IA responsable, con políticas de privacidad y ética, controles de acceso, anonimización cuando proceda, auditorías y formación específica.

2.1 Objetivos del proyecto

El proyecto busca reducir el riesgo de exposición de PII en interacciones con IA, estandarizar un proceso de preparación segura de textos aplicable en toda la organización y medir la adopción y la eficacia mediante indicadores claros, comparables y auditables que permitan la mejora continua.

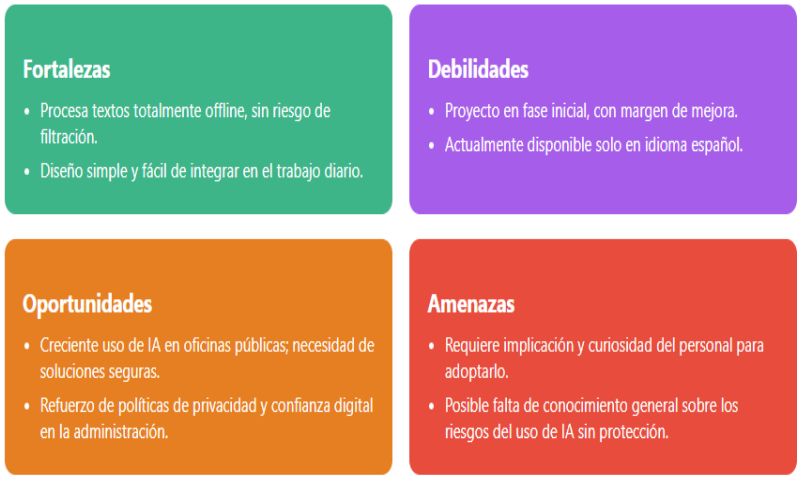
SMART

Objetivo general	Fomentar el uso responsable y seguro de la inteligencia artificial en entornos administrativos mediante la herramienta PrevIA.
Específico	Proporcionar una extensión local que permita analizar y anonimizar textos sensibles antes de compartirlos o procesarlos con IA.
Medible	Evaluar el nivel de adopción, satisfacción y percepción de seguridad del personal tras la implementación de la herramienta.
Alcanzable	Desplegar PrevIA de forma progresiva en las áreas con mayor exposición a datos sensibles, ofreciendo formación y soporte.
Relevante	Contribuir a la mejora de la cultura digital y al cumplimiento de las normativas de protección de datos en la administración pública.
A tiempo	Realizar un seguimiento periódico de su uso y resultados, integrando mejoras en las versiones futuras de la herramienta.

2.2 Alcance

El alcance abarca entornos de redacción de alto volumen (oficios, resoluciones, informes y comunicaciones externas), el uso en navegadores compatibles actualmente Firefox y Edge, con empaquetado ZIP para otros**,** y la entrada de texto tanto escribiéndolo directamente como copiándolo desde cualquier origen (correo, gestor de expedientes, intranet, etc.), incluyendo DOCX, PDF y TXT. Está pensado para cualquier usuario, desde instituciones públicas hasta usuarios estándar que quieran proteger su información en sus interacciones con IA.

DAFO



CAME

Corregir	Afrontar	Mantener	Explotar
Reducir falsos positivos y negativos mediante pruebas reales.	Apoyar la revisión manual con ayudas visuales y avisos.	Procesamiento completamente local y privado.	Arquitectura ligera y portable entre navegadores.
Optimizar la interfaz para reducir fricción en el uso.	Gestionar expectativas sobre los límites técnicos del sistema.	Diseño ético, modular y transparente.	Difundir la herramienta en comunidades de innovación pública.

3. Descripción funcional

La solución detecta PII por patrones (nombres, DNI/NIE, teléfonos, emails, IBAN, direcciones, fechas, números de expediente, identificadores, URLs…) y ofrece modos de transformación configurables seudonimización (p. ej., [NOMBRE_#]), generalización (p. ej., “una persona) y anonimización manual, con un semáforo de riesgo e inserción de pautas de exportar. Incluye panel de personalización por tipo de dato y esquema de buenas prácticas, y garantiza procesamiento 100 % local, sin llamadas externas y con permisos mínimos conforme a Manifest V3.

Principio de uso responsable: PrevIA no sustituye la revisión manual. Aporta una barrera de protección y un estándar de preparación segura previo a compartir texto con herramientas de IA.

4. Arquitectura y seguridad

La arquitectura se basa en una extensión web (Manifest V3) con permisos mínimos y análisis local en el navegador, evitando exfiltraciones. No se almacena contenido, permite guardar el

resultado en .txt con formato con fecha y hora. El diseño aplica privacy by default y security by design, sin generar tráfico de red durante el análisis.

5. Proceso y roles (PDCA)

Proceso clave: Analizar → Transformar → Revisar → Exportar.

Etapa	Descripción	Control
Plan (P)	Definir reglas y plantillas por unidad	Aprobación de propietario del proceso
Do (D)	Análisis y transformación local del texto	Checklist de revisión en interfaz
Check (C)	Medir precisión y adopción	Indicadores y muestreo periódico de documentos
Act (A)	Mejora de patrones y UX	Acciones correctivas/preventivas

5.1 Roles

Rol	Personal	Responsabilidades clave
Propietario del proceso	Unidad promotora de aprendizaje digital / calidad	Define el proceso, aprueba cambios, prioriza mejoras y vela por los indicadores.
Usuarios finales	Personal que elabora textos con posible PII	Usan la herramienta, aplican la checklist y reportan incidencias y mejoras.
Soporte técnico	Equipo que mantiene la extensión y las reglas	Empaqueta, mantiene y distribuye la extensión, asegura compatibilidad entre navegadores y actualiza las **reglas de detección** (pruebas, versiones, publicación).
Comité de calidad	Dirección / calidad / seguridad de la info	Revisa indicadores, evalúa riesgos y decide acciones de mejora y escalado.

6. SIPOC del servicio PrevIA

Suppliers (S)	Inputs (I)	Process (P)	Outputs (O)	Customers (C)
Áreas que generan documentos; TI (paquetizado); Calidad/PDCA; DPO/Privacidad	Textos a tratar; **reglas PII versionadas** (JSON); **políticas** y plantillas de aviso; criterios de aceptación; catálogo de datos sensibles	1) **Cargar** texto → 2) **Analizar** (patrones PII) → 3) **Personalizar** (tipos/umbral por unidad) → 4) **Transformar** (seudonimizar/generalizar/anonimizar) → 5) **Revisar** (checklist + semáforo) → 6) **Exportar** (copiar/guardar) → 7) **Feedback** (incidencias/ideas)	**Texto anonimizado** con marca de transformación; **informe de riesgo** (semáforo + hallazgos); **registro local** (tiempo, nº reemplazos); **issues** para ajuste de reglas	Personal redactor; jefaturas/secretarías; **DPO/Calidad** (auditoría); **ciudadanía** (beneficio indirecto por privacidad)

7. Plan de implantación (12 semanas)

Fase	Semanas	Acciones principales	Entregables
Piloto	1–4	Seleccionar N=25 usuarios; formación 45’; línea base; soporte cercano	Guía rápida; baseline de PII y tiempos; feedback inicial
Mejora	5–8	Ajustar patrones; mejorar UX; checklist; comunidad de práctica	Versión v0.2; acta de lecciones; catálogo interno provisional
Decisión y escalado	9–12	Evaluación (satisfacción, precisión, tiempos); directriz de uso; paquetizado ZIP	Informe de evaluación; decisión; plan de despliegue por unidades

8. Riesgos y mitigación

El proyecto presenta algunos **riesgos clave**, como la posibilidad de falsos positivos o negativos en la detección de información sensible (PII), la necesidad de una revisión manual para

confirmar las sustituciones, las limitaciones de las expresiones regulares ante textos no estructurados y la posible exposición visual de datos durante dicha revisión. Para mitigarlos, **PrevIA** aplica una revisión asistida con control manual de resultados, actualiza de forma iterativa los patrones para mejorar la precisión, mantiene un procesamiento 100 % local sin llamadas externas y promueve la transparencia sobre sus límites técnicos, fomentando un uso responsable. En su análisis crítico, se constata la viabilidad del enfoque al demostrar que es posible anonimizar textos de forma privada y local; se reconocen limitaciones ligadas a la intervención humana necesaria para garantizar fiabilidad; se destaca la escalabilidad de su **arquitectura modular, adaptable a nuevos patrones e idiomas**; y se subraya su valor **ético**, al priorizar la privacidad y la transparencia sobre la automatización completa.

9. Indicadores (KPIs) y criterios de aceptación

Indicador	Definición	Meta	Periodicidad	Fuente
Precisión de detección	% aciertos sobre muestra de 10 textos tipo	≥ 70 %	Mensual	Muestreo/QA
Eficacia de anonimización	% elementos sensibles sustituidos	≥ 80 %	Mensual	Muestreo/QA
Tiempo análisis	Segundos por 1.000 palabras	≤ 2 s	Mensual	Métricas locales
Procesamiento local	% operaciones sin tráfico de red	100 %	Continuo	Verificación técnica
Adopción por unidad	% unidades que usan PrevIA en entregables con PII	≥ 60 % (3 meses)	Mensual	Encuesta + registros
Revisión humana	Tiempo medio de revisión por doc.	≤ 2 min	Mensual	Encuesta/observación
Reducción PII en auditoría	Comparativa antes/después sobre muestra	≥ 50 %	Trimestral	Auditoría interna
Satisfacción	Valoración media (1–4)	≥ 3,5	Trimestral	Encuesta

10. Gobierno del producto y mejora continua

Se establece un comité mensual de calidad (CAPA) que revisa indicadores e incidencias, acuerda acciones correctivas y preventivas con responsables y plazos y verifica su cierre. En paralelo, un canal interno de mejora permite compartir ejemplos, reglas y plantillas seguras, con curación y

repositorio común. Además, todo cambio en reglas o plantillas se gestiona con versionado y notas de lanzamiento, registrando autor, fecha e impacto para asegurar trazabilidad, auditoría y posibilidad de rollback.

11. Formación, comunicación y cambio

La difusión de PrevIA se basará en una comunicación práctica y motivadora. Se elaborará un breve **vídeo divulgativo** que muestre su utilidad y funcionamiento básico, junto con una **guía visual o ficha rápida**. El objetivo es fomentar el uso voluntario y natural de la herramienta, destacando sus beneficios en la protección de datos.

12. Privacidad, ética y cumplimiento

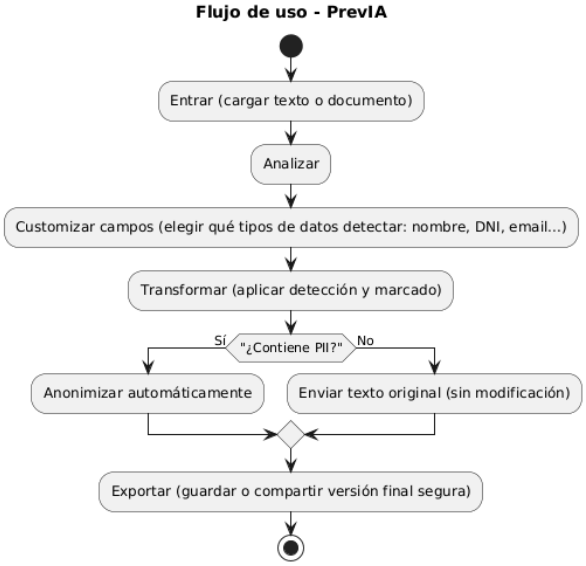
PrevIA se ha diseñado conforme a los principios de privacidad, ética y cumplimiento normativo. Aplica el principio de minimización, utilizando únicamente la información necesaria para la detección. Adopta un enfoque de privacy by design, con procesamiento totalmente local y permisos mínimos. Mantiene la transparencia mediante avisos claros sobre los límites técnicos y el uso responsable.

13. Roadmap

Hito	Contenido	Plazo
Q1	Piloto + ajuste de patrones; DOCX/PDF básico	0–3 meses
Q2	Soporte inglés; reglas por unidad; informes agregados (sin contenido)	4–6 meses
Q3–Q4	Integración con gestor documental (pre-export); plantillas seguras por defecto	7–12 meses

15. Anexos

1. Diagrama de Flujo



2. Mapa de procesos + Ciclo PDCA

3. Versiones

Versión 0.1

- Procesamiento totalmente offline y local.
- Detección automática de datos personales (PII).
- Seudonimización, generalización y anonimización manual o automática.
- Personalización de campos a analizar.
- Exportación segura en formato `.txt` o copia estándar.
- Guía de buenas prácticas y avisos éticos integrados.

Versión 0.2

- Soporte ampliado al idioma inglés.
- Personalización avanzada del análisis (inclusión o exclusión de tipos de datos).
- ...

Versión 0.3

- Mejoras en precisión y rendimiento del análisis contextual.
- Procesamiento de documentos básicos (TXT, DOCX, PDF).
- ...

4. Enlaces

Tutorial

https://youtu.be/PTowVmzVVVE

Código fuente

https://github.com/SaraFullStack/PrevIA

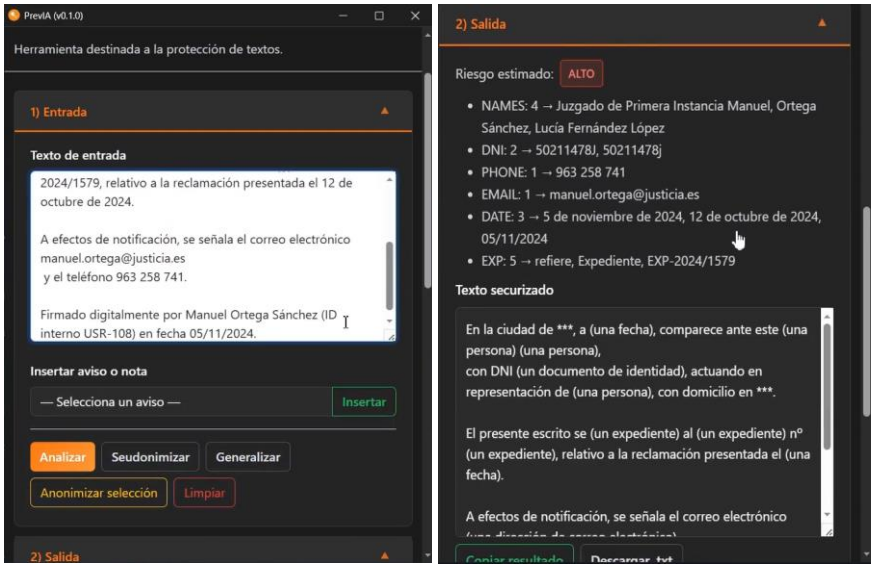
Disponible en Edge y Firefox

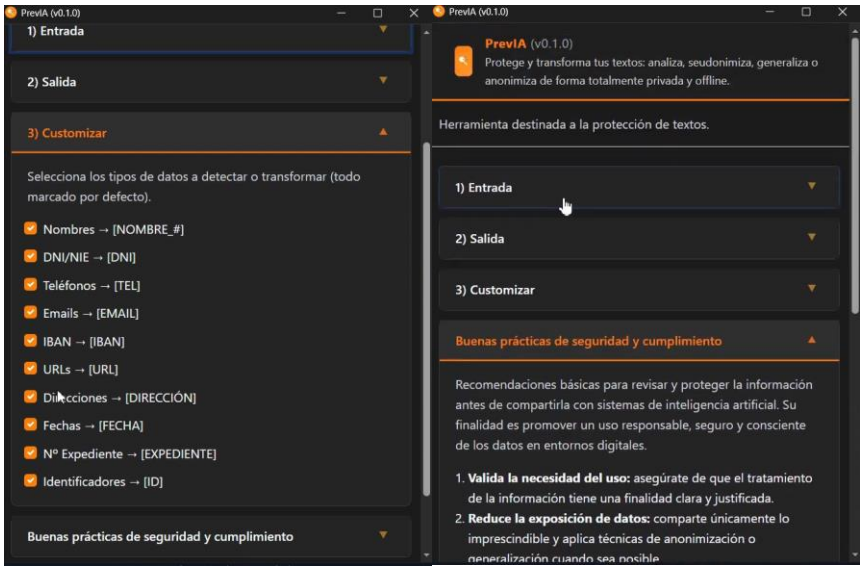
https://microsoftedge.microsoft.com/addons/detail/previa-%E2%80%93-protege-tus-text/kofcadpkohaomaabekmdkelbfpjcekpg

https://addons.mozilla.org/es-ES/firefox/addon/previa/

WebSide
https://sarafullstack.github.io/

5. Capturas





Este documento presenta un MVP (Producto Mínimo Viable) desarrollado en un periodo limitado y con recursos reducidos. Su objetivo es demostrar la viabilidad técnica y funcional de la idea, más que ofrecer un producto final. Es posible que contenga errores o limitaciones propias de una fase inicial, por lo que se considera una versión abierta a revisión, validación y mejora continua, especialmente a partir del uso real, la retroalimentación de los usuarios y la evaluación institucional.

6. Algunas Referencias

INAP – Gobernanza y uso ético de la IA:
https://laadministracionaldia.inap.es/noticia.asp?id=1256899
INAP – Estrategia de aprendizaje y seguridad TIC:

https://www.inap.es/es/aprendizaje/estrategias-de-aprendizaje
Centro Criptológico Nacional (CCN-CERT): https://www.ccn-cert.cni.es

European Commission – Ethics Guidelines for Trustworthy AI: https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai
OECD – Principles on Artificial Intelligence: https://oecd.ai/en/ai-principles
NIST – AI Risk Management Framework (RMF 1.0): https://www.nist.gov/itl/ai-risk-management-framework
Pega – AI Manifesto v3: https://www.scribd.com/document/820204908/Pega-Ai-Manifesto-v3
INAP – Colaboración con CCN para formación en seguridad: https://www.inap.es/es/noticias/acciones-formativas-en-materia-de-seguridad-tic-en-colaboracion-con-el-centro-criptologico-nacional