Integración de modelos de IA para la identificación de Información de Identificación Personal (PII) en el marco de las Regulaciones vigentes de la República Argentina

Resumen. La protección de la Información de Identificación Personal (PII) se ha convertido en un aspecto crítico de la ciberseguridad a nivel global. En Argentina, la Ley 25.326 establece el marco regulatorio para el resguardo de estos datos, los cuales suelen encontrarse dispersos en documentos organizacionales, dificultando su detección y protección. Para abordar esta problemática, el presente estudio propone el desarrollo de un Framework de detección de PII que integra modelos de Inteligencia Artificial (IA) capaz de identificar posibles PII en documentos empresariales, asegurando el cumplimiento de la legislación vigente en materia de privacidad en la República Argentina. El principal desafío radica en garantizar la trazabilidad de los datos sensibles en el idioma Español y optimizar la capacidad de respuesta ante incidentes de filtración de información. Para ello, además del framework, se diseñará una herramienta de benchmarking que permitirá evaluar su precisión y eficiencia en escenarios reales. Este trabajo no solo fortalece la seguridad de la información, sino que también impulsa la transferencia de conocimiento hacia distintos sectores, promoviendo el uso de soluciones basadas en IA en el campo de la ciberseguridad.

Palabras Claves. Personally Identifiable Information (PII), Cybersecurity, Data Protection, Argentina, PII Detection Framework, Benchmarking Tool.

Title: Automation of Personally Identifiable Information (PII) Detection in Argentina Using Artificial Intelligence.

Abstract. The protection of Personally Identifiable Information (PII) has become a critical aspect of cybersecurity worldwide. In Argentina, Law 25,326 establishes the regulatory framework for safeguarding this data, which is often dispersed across organizational documents, making its detection and protection more difficult. To address this issue, this study proposes the development of a PII detection framework that integrates Artificial Intelligence (AI) models capable of identifying potential PII in corporate documents, ensuring compliance with current privacy legislation in the Argentine Republic. The main challenge lies in guaranteeing the traceability of sensitive data in the Spanish language and optimizing response capabilities in the event of information leakage incidents. To this end, in addition to the framework, a benchmarking tool will be designed to evaluate its accuracy and efficiency in real-world scenarios. This work not only strengthens information security but also promotes knowledge transfer across different sectors, encouraging the use of AI-based solutions in the field of cybersecurity.

Keywords. Automation of Personally Identifiable Information (PII) Detection in Argentina Using Artificial Intelligence.

1 Introducción

A nivel internacional, el tratamiento y la protección de PII han sido abordados por múltiples normativas sectoriales, como HIPAA (NIST, 2010) o el GDPR (General Data Protection Regulation) . Estas regulaciones han impulsado el desarrollo de herramientas tecnológicas —muchas de ellas basadas en técnicas de procesamiento de lenguaje natural y aprendizaje automático— capaces de detectar y clasificar automáticamente datos sensibles. Investigaciones recientes exploran modelos de detección aplicables a

contextos multilingües, aunque la mayoría de ellos se encuentran orientados a la lengua inglesa y a marcos normativos foráneos (Silva et al, 2020).

En el caso argentino, si bien la Ley 25.326 y su decreto reglamentario (Gobierno de Argentina, 2000) constituyen un marco jurídico vigente y claro, no se observa aún una adopción extendida de herramientas automatizadas de detección de PII adaptadas al idioma español y al contexto legal local. Según Cimpanu (Cimpanu, C, (2021) esto genera un vacío tecnológico que complica la trazabilidad de los datos personales y dificulta la respuesta ante incidentes de seguridad, en un entorno donde las filtraciones de datos se vuelven cada vez más frecuentes (Heiligenstein, M. X. (2024)).

La identificación de PII en documentos organizacionales enfrenta obstáculos adicionales vinculados a la variedad de formatos (textos planos, PDFs, imágenes escaneadas), al uso de jergas sectoriales, y a la presencia de datos indirectos que permiten inferencias sensibles (Splunk). En este escenario, resulta fundamental avanzar hacia soluciones locales que integren conocimiento normativo, lingüístico y organizacional.

Según el reporte 2023/2024 del Foro Internacional de Economía, a corto plazo (2 años), la inseguridad digital se coloca como el cuarto riesgo con impacto negativo en la economía y, a largo plazo (10 años, 2034), esta preocupación ocuparía el puesto 8 en su ranking de riesgos.

De acuerdo al mismo informe, la mayoría de los riesgos tecnológicos destacados a corto plazo, Desinformación e Inseguridad digital, bajan de ranking, pero permanecen en el top 10 a largo plazo, en el quinto y octavo lugar, respectivamente. El Foro Internacional de Economía define Inseguridad Digital como el uso de armas y herramientas cibernéticas para llevar a cabo guerra cibernética, espionaje y delitos cibernéticos para obtener control sobre una presencia digital y/o causar interrupción operativa.

Como ejemplos de fuga de datos se puede mencionar los 700 millones de datos filtrados de LinkedIn, los cuales incluían datos de ubicación de usuarios (ciudad, provincia, país), nombre de usuarios y correo electrónico (Cimpanu, C., 2021).

El proceso de digitalización en el que se encuentran inmersas las organizaciones ha traído consigo beneficios significativos en eficiencia, productividad y accesibilidad a la información. Sin embargo, este avance también ha generado nuevos riesgos en materia de seguridad de la información, siendo uno de los más críticos la exposición de Información de Identificación Personal (PII).

Toda organización que maneja información de identificación personal (PII) tiene el deber de cuidar a sus usuarios y clientes para asegurarse de que la PII no quede expuesta al mundo exterior y, al mismo tiempo, cumplir con los requisitos normativos, de seguridad y de cumplimiento para mantenerla interna e intacta. El desafío que se presenta es que la PII aparece en lugares inesperados en los archivos de registro de una organización e identificarla puede ser complicado. La PII directa, como direcciones de correo electrónico o números de tarjetas de crédito, puede ser bastante sencilla de identificar observando la forma en que está formateada. Pero la PII también puede ser indirecta y menos fácil de identificar a partir de datos sin procesar, como información médica, historial laboral o información descriptiva sobre un individuo (Splunk. (s.f.).

La PII se define como cualquier dato que pueda ser utilizado para identificar directa o indirectamente a una persona física. Esto incluye no solo datos obvios como el nombre, la dirección o el número de documento, sino también datos contextuales o combinados, como direcciones IP, registros médicos o datos biométricos [NIST SP 800-122. Guide to Protecting the Confidentiality of Personally Identifiable Information (PII). NIST, 2010.]. La exposición de estos datos puede tener consecuencias legales, financieras y sociales de gran alcance, como el robo de identidad, el fraude financiero o la pérdida de confianza en instituciones (ISO, 2013).

Segun Revista Innovación Seguridad (2024) y Segu-Info (2023) región ha experimentado incidentes de fuga de datos de gran escala en los últimos años, que han afectado tanto a organismos públicos como a empresas privadas, evidenciando la urgencia de fortalecer las capacidades técnicas para proteger la PII; dentro de estas fuga de datos se puede mencionar la de RENAPER en el 2023 y 2024

En este contexto, el World Economic Forum (2024) y CISCO señala que desarrollar soluciones que integren Inteligencia Artificial (IA) aplicadas a la ciberseguridad se vuelve una necesidad estratégica para resguardar la privacidad de los ciudadanos y el cumplimiento normativo de las organizaciones.

Esta aseveración es parte del objetivo propuesto en el proyecto “Desarrollo de un modelo de IA para detección de Personal Identificable Information (PII)”; con código COSIEC611, en el cual esté inserto el presente trabajo. Parte de las pruebas de concepto de este proyecto incluye probar/catalogar herramientas actuales de detección de patrones de datos que sean de código abierto. Siendo la primera Presidio (Microsoft Presidio), y del cual se exhiben a continuación los avances sobre el mismo utilizando el idioma español.

Cabe aclarar que el presente trabajo no expone un framework finalizado, sino que presenta resultados preliminares correspondientes a los avances de esta prueba de concepto, enmarcada en la fase inicial del proyecto COSIEC611.

Este trabajo propone un enfoque híbrido basado en Procesamiento de Lenguaje Natural (NLP) en español, contenedorización con Docker y reconocedores personalizados en Microsoft Presidio, adaptado al contexto argentino. El Procesamiento del Lenguaje Natural (PLN) es una rama de la Inteligencia Artificial y la Lingüística, dedicada a hacer que las computadoras comprendan las declaraciones o palabras escritas en lenguajes humanos según Khurana et al.

Se busca responder a una brecha tecnológica clara: la falta de soluciones que integren requerimientos normativos locales, características lingüísticas del español rioplatense y escalabilidad técnica para entornos organizacionales.

2 Grado de avance

Para lograr una integración que satisfaga la necesidad del idioma español fue necesario tener un panorama de los modelos/frameworks disponibles que identifiquen palabras en español en diferentes contextos y soporte al idioma español, identificar su tipo de licencia y costo, dicho panorama muestra a continuación (Tabla 1).

Para lograr entender esta comparativa es necesario describirlos, se puede agrupar por tipo

- Encoder / BERT (Skim AI, 2020) y Encoder / RoBERTa: son modelos de lenguaje que "leen" texto en ambas direcciones simultáneamente (izquierda y derecha) para entender el contexto de cada palabra. No generan texto nuevo, sino que producen representaciones numéricas (embeddings) del texto. RoBERTa (Liu et al, 2019) es una versión mejorada de BERT con un proceso de entrenamiento más robusto. Para usarlos en detección de PII, se requiere un paso extra de ajuste (fine-tuning): entrenarlos con datos etiquetados para la tarea específica de NER o Named Entity Recognition (Seow, 2025).
- Encoder multilingüe: mismo tipo de arquitectura encoder, pero entrenado con texto en decenas o cientos de idiomas al mismo tiempo. La ventaja es que un solo modelo sirve para múltiples idiomas. La desventaja es que suele rendir un poco menos que un modelo monolingüe dedicado, porque "reparte" su capacidad entre idiomas.
- Encoder NER Zero-shot: también es un encoder, pero diseñado específicamente para que se pueda indicarle en tiempo de ejecución qué entidades querés detectar, sin necesidad de fine-tuning previo. GLiNER entra en esta categoría: se provee el texto y una lista de etiquetas ("CUIL", "nombre", "teléfono") y detecta sin haber sido entrenado específicamente para eso. El modelo GLiNER (gliner_multi_pii-v1), según Chacón (2025), es un modelo multilingüe adaptado para entidades de información de identificación personal (PII), compartido por los creadores de GLINER. Utiliza la variante base del codificador bidireccional multilingüe DEBERTaV3.
- Framework PII: no es un modelo en sí, sino una plataforma o biblioteca que orquesta múltiples modelos y reglas. Presidio es el mejor ejemplo: combina un motor NER, expresiones regulares, listas de contexto y lógica de puntuación.
- LLM generativo (Mora-Delgado et al, 2026): modelos de lenguaje de gran escala que generan texto nuevo como respuesta a una instrucción. A diferencia de los encoders, no requieren fine-tuning para nuevas tareas: se les brinda un prompt en lenguaje natural ("identificá todos los datos personales en este texto"). GPT-4 o Rajgarhia, H. et al. (2025), Claude (2026) y OpenPipe entran acá. Son más flexibles pero más costosos y menos determinísticos.

- Framework NLP clínico: Similar a un framework PII pero especializado en el dominio de la salud. John Snow Labs combina modelos entrenados con texto médico, reglas específicas para PHI (información de salud protegida) y pipelines optimizados para documentos clínicos como historias clínicas o informes de laboratorio. El framework de NLP clínico de John Snow Labs (Kocaman et al, 2021) es una plataforma avanzada diseñada para procesar texto médico utilizando modelos de Procesamiento del Lenguaje Natural (NLP) especializados en el dominio de la salud. Su producto principal es Spark NLP y su extensión clínica Spark NLP for Healthcare, ampliamente usados en hospitales, investigación biomédica y sistemas de historia clínica electrónica (EHR).
- NLP library / NER: Biblioteca de propósito general para procesamiento de lenguaje natural. spaCy no es solo un modelo sino un ecosistema completo: tokenización, análisis morfológico, NER, entre otros. En el contexto de PII, actúa principalmente como motor NER subyacente de otros sistemas, como el Presidio.

Modelo / Framework	**Tipo**	**Open Source**	**Soporte Español**	**Costo Aprox.**
BETO (BERT español)	Encoder / BERT	Sí (Apache 2.0)	Nativo español	Gratuito
MarIA (RoBERTa-BNE)	Encoder / RoBERTa	Sí (Apache 2.0)	Nativo español	Gratuito
RigoBERTa	Encoder / RoBERTa	Sí (Apache 2.0)	Nativo español	Gratuito
XLM-RoBERTa (Schelb)	Encoder multilingüe	Sí (MIT)	104 idiomas incluyendo. ES	Gratuito
GLiNER (gliner_multi_pii-v1) [https://urchade.github.io/GLiNER/]	Encoder NER Zero-shot	Sí (Apache 2.0)	Multilingüe	Gratuito
Microsoft Presidio	Framework PII	Sí (MIT)	Multi (es_core_news_md)	Gratuito (self-hosted)
OpenPipe PII-Redact	LLM generativo	No	Inglés principalmente	API pago
GPT-4o (OpenAI)	LLM generativo	No	Multilingüe	USD 2.5-10/1M tokens
Claude 3.7 Sonnet	LLM generativo	No	Multilingüe	USD 3-15/1M tokens
John Snow Labs NLP	Framework NLP clínico	Parcial (community)	Inglés, extensible	Enterprise (pago)
spaCy es_core_news	NLP library / NER	Sí (MIT)	Nativo español	Gratuito

Tabla 1. Modelos/Frameworks actuales

Por otro lado, y teniendo en cuenta que uno de los criterios que se usa es que lo desarrollado sea de código abierto, se llevó a cabo una investigación exploratoria del estado del arte actual, sobre distintas herramientas de código abierto que logren identificar datos que puedan calificarse, directa e indirectamente como PII, cuyo resultado se muestra en la muestra la Tabla 2.

Herramienta	Precisión	Velocidad	Dificultad	Uso típico
Tesseract	M	M	M	Escaneos, documentos simples
EasyOCR	A	M	B	Texto en fotos, múltiples idiomas
YOLOv8 Ultralytics, 2023)	A	A	B	Cámaras en tiempo real, apps móviles
OpenCV	M	A	A	Preprocesamiento, filtros, visión clásica
Detectron2	AA	B/M	A	Investigación, datasets complejos

Tabla 2. Comparativa de herramientas específicas para determinados propósitos. Donde A (Alta); M (Media); B (Baja); AA (Muy Alta)

La elección de Microsoft Presidio por sobre las demás alternativas relevadas en la Tabla 1 se fundamenta en que es una solución de código abierto, autohospedable y extensible. Estas características se alinean con la premisa del proyecto de generar desarrollos de código abierto e impactan, de manera indirecta, en una reducción de costos —al no depender de servicios pagos por uso ni de licencias propietarias— y en un mayor control sobre los datos procesados. A diferencia de los modelos generativos (GPT-4o, Claude) o de las soluciones propietarias, su extensibilidad permite incorporar reconocedores personalizados adaptados al contexto argentino. En cuanto al motor lingüístico, se optó por el modelo es_core_news_md de spaCy debido a su soporte nativo del idioma español, requisito central para la detección de PII en documentos locales.

Explicado lo anterior, se puede pasar a la solución propuesta, la cual se basa en la integración de Microsoft Presidio, una plataforma open-source, autohosteable, de detección y anonimización de PII, con un entorno contenerizado de servicios mediante Docker. La arquitectura de esta prueba de concepto consta de tres componentes principales:

- Motor NLP: encargado del preprocesamiento y análisis lingüístico de textos, utilizando utilizando modelos de Procesamiento de Lenguaje Natural (NPL). El modelo seleccionado para la prueba de concepto es es_core_news_md ya que es uno de los modelos ya entrenados y disponibilizados por spaCy [spaCy. Industrial-Strength NLP in Python. https://spacy.io/] para aplicarlo sobre el idioma español.
- Servicio de detección (Presidio Analyzer y Anonymizer): realiza la identificación de entidades sensibles.
- API REST: interfaz que expone los servicios a aplicaciones externas mediante endpoints simples y escalables [FastAPI. Modern Web Framework for APIs. https://fastapi.tiangolo.com/].

Para entender mejor el flujo, es necesario mirar la arquitectura de los componentes de Microsoft Presidio [42], el Analyzer y Anonymizer. En la Figura 1 observamos el Analyzer, encargado de identificar datos considerados parte de PII en un texto.

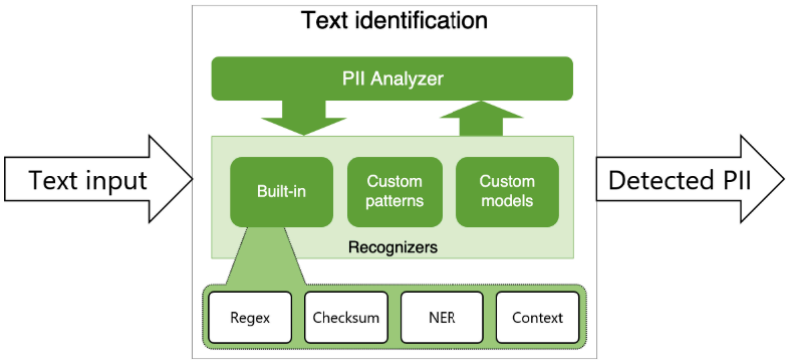

Figura 1. Componente Analyzer de Microsoft Presidio.

Por otro lado, se describe el componente Anonymizer (Figura 2), el cual toma los datos identificados del componente Analyzer y les aplica reglas de anonimización para que el texto no sea legible totalmente, protegiendo el valor real de los posibles datos PII identificados previamente. Este componente también permite anexar reglas de anonimización personalizadas.

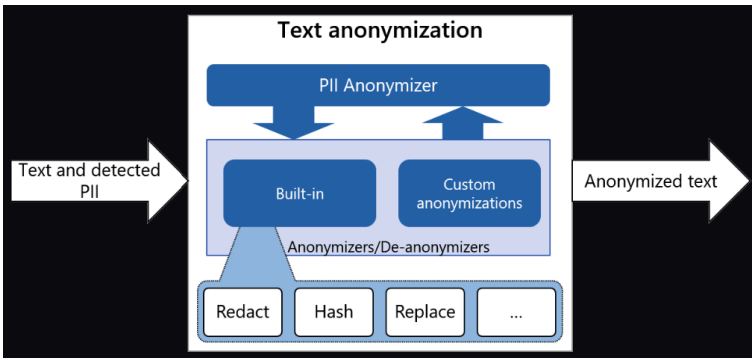

Figura 2. Componente Anonymizer de Microsoft Presidio.

Es importante destacar que las contribuciones técnicas presentadas en el presente son anexadas a este componente antes de su ejecución para su correcto procesamiento de los patrones buscando, teniendo en cuenta su aplicabilidad sea dentro del territorio Argentino. Mientras que Microsoft Presidio cuenta con reconocedores genéricos (emails, números de teléfono internacionales, tarjetas de crédito), fue necesario ampliar la base de patrones para incluir reconocedores con patrones adaptados de:

- DNI
- CUIL/CUIT
- Obra Social

Con esto como base podemos mostrar la arquitectura del framework, en conjunto con la herramienta de benchmarking presentados a continuación.

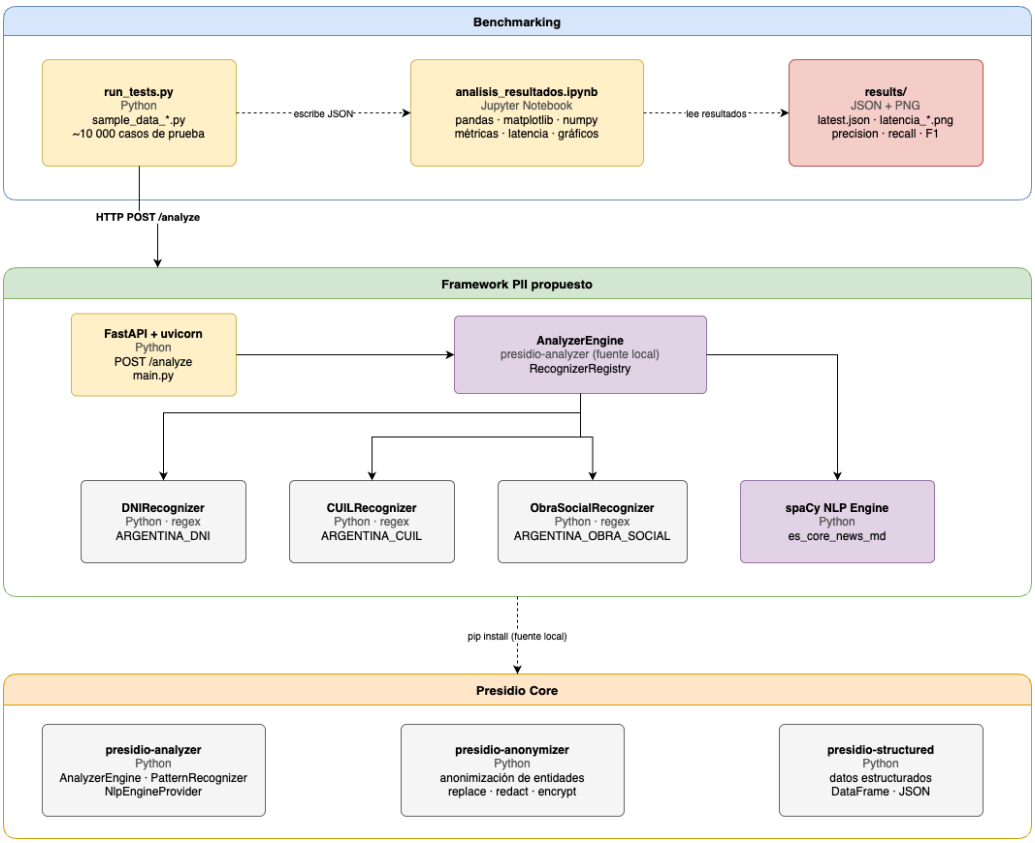

Figura 3. Arquitectura del framework presentado y herramienta de benchmarking.

Los reconocedores propuestos fueron desarrollados en Python. Para la herramienta de benchmarking también se usó Python integrado con Júpiter Notebook. Para las pruebas de los reconocedores, generamos muestras aleatorias de textos con posibles PII, listadas a continuación, siendo n la cantidad de muestras generadas:

- n1 = 250
- n2 = 1000
- n3 = 10000

Las muestras empleadas son datos sintéticos generados programáticamente —no datos reales de personas— a partir de plantillas de texto representativas de contextos administrativos, médicos y financieros, combinadas con identificadores construidos sobre rangos numéricos controlados (DNI de 7 y 8 dígitos, y CUIL/CUIT con sus prefijos válidos). La generación es reproducible mediante una semilla fija e incluye de manera deliberada tanto casos positivos (textos con PII esperado) como casos negativos o adversariales (números y cadenas similares en contextos que no constituyen PII, como precios, números de lote o identificadores de transacción), lo que permite evaluar también los falsos positivos. El empleo de datos sintéticos evita exponer información personal real y proporciona una referencia exacta (ground truth) de las entidades esperadas en cada muestra; como contrapartida, la validez externa de los resultados deberá corroborarse en etapas posteriores sobre documentos reales.

Antes de definir las métricas, es útil conocer los cuatro resultados posibles de la clasificación aplicada (Tabla 3) y las métricas a utilizar:

	Detectado como PII	**No detectado como PII**

Es PII	Verdadero Positivo (VP)	Falso Negativo (FN)
No es PII	Falso Positivo (FP)	Verdadero Negativo (VN)

Tabla 3. Tabla de clasificación posible de resultados.

A continuación las métricas utilizadas para cada posible detección de PII, denominándose entidad al string identificado como PII

- **Precisión = VP/(VP+FP).** Si detecta 10 entidades y 8 son realmente PII → Precisión = 0.80 (80%).
  - Alta precisión → pocos falsos positivos, la integración no marca cosas incorrectas.
  - Baja precisión → la integración "dispara de más", marcando texto que no es PII.
- **Recall (Exhaustividad) = VP/(VP+FN)**: Si había 10 entidades PII reales y encontró 7 → Recall = 0.70 (70%).
  - Alto recall → la integración encuentra la mayoría del PII presente.
  - Bajo recall → la integración "se pierde" entidades reales (falsos negativos).
- **Puntaje F1 = 2 x ((Precisión x Recall)/(Precisión + Recall)):** Si Precisión = 0.80, Recall = 0.70 → F1 = 2 × (0.80 × 0.70) / (0.80 + 0.70) ≈ 0.747. Es la media armónica entre Precisión y Recall. Útil cuando se quiere un balance entre ambas métricas. Valor entre 0 y 1 (o 0% y 100%).
  - Si F1 = 1 → precisión y recall son perfectos.
  - Si F1 = 0 → no detectó nada correctamente.

El puntaje F1 de cada reconocedor actúa como umbral de confianza: a mayor score, más preciso pero potencialmente menor recall. En detección de PII, generalmente se prioriza el recall (no queremos dejar datos sensibles sin detectar), aceptando ciertos falsos positivos.

Explicada las métricas, se procede a mostrar los resultados por reconocedor de la muestra con n=10000 (Figura 4).

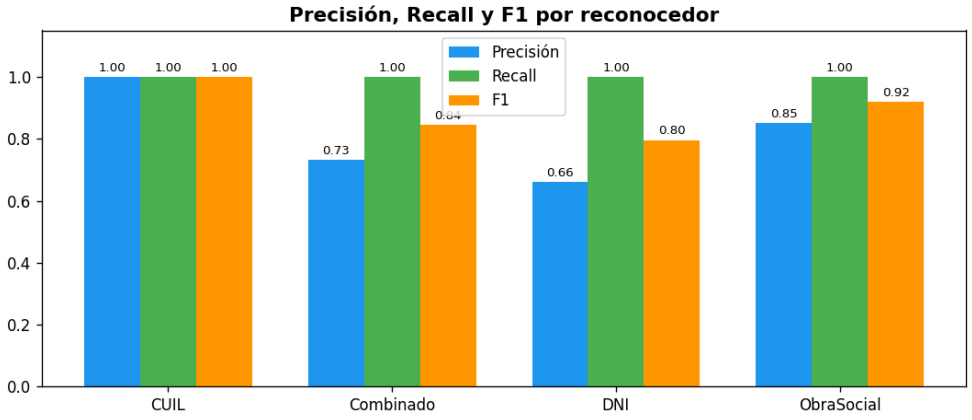

Figura 4. Métricas por reconocedor y combinadas.

Como se puede observar el reconocedor de CUIL/CUIT tiene un 100% de efectividad, mientras que el de DNI es el que tiene menos efectividad de los tres reconocedores, si bien su recall es 1, ya que se identificaron muchos falsos positivos (Figura 5.), seguido por el de obra social.

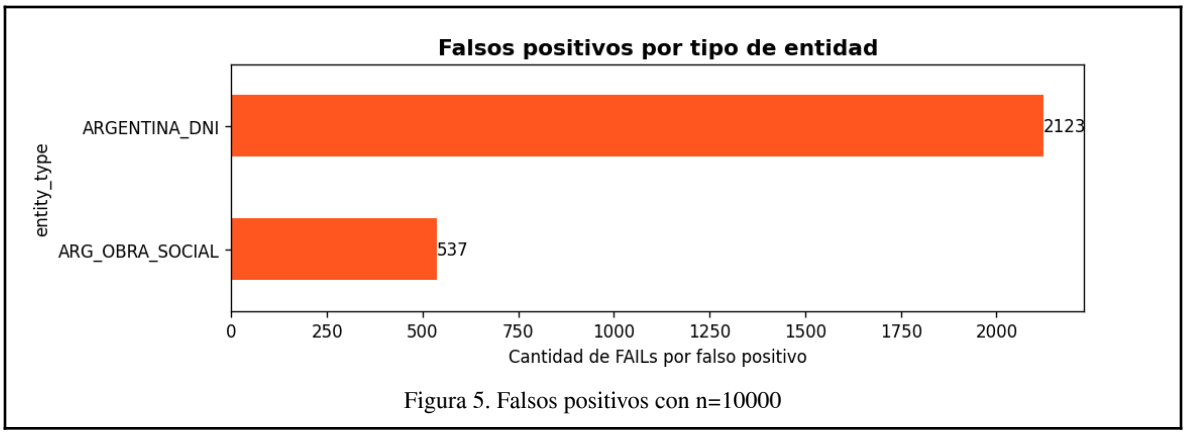

Figura 5. Falsos positivos con n=10000

El menor desempeño en precisión del reconocedor de DNI se explica por la naturaleza del propio identificador: el DNI argentino es una secuencia de 7 u 8 dígitos sin dígito verificador, por lo que —a diferencia del CUIL/CUIT o el CBU— no es posible validar su estructura interna. En consecuencia, el reconocedor, basado en una expresión regular que detecta cualquier número de 7 u 8 dígitos, marca como DNI a cifras que aparecen en contextos no personales (importes, números de lote, identificadores de transacción u orden de compra), lo que genera falsos positivos.

Se define la tasa de aprobación de un reconocedor como el porcentaje de casos de prueba aprobados (PASS) sobre el total de casos evaluados para ese reconocedor. Un caso se considera aprobado únicamente cuando se detectan todas las entidades esperadas y no se genera ningún falso positivo; se trata, por lo tanto, de una métrica binaria por caso (todo o nada) que no contabiliza coincidencias parciales. Los casos se agrupan según el reconocedor que evalúan (DNI, CUIL, Obra Social) y aquellos que contienen varios PII combinados se contabilizan en la categoría «Combinado».

En cuanto, a la identificación realmente de PII, podemos analizar la tasa de aprobación (Figura 6) de cada reconocedor. incluyendo muestras que tenían posibles PII combinados.

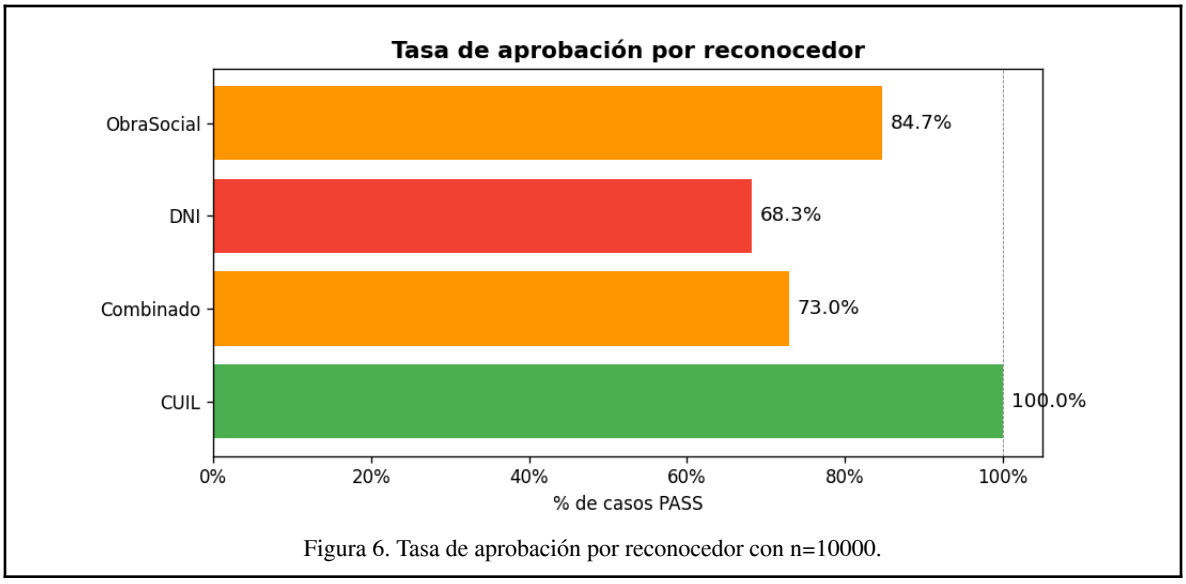

Figura 6. Tasa de aprobación por reconocedor con n=10000.

Es importante destacar que, si bien sólo se están mostrando los resultados con n=10000, los resultados con las otras muestras, muestran la misma tendencia, la cual se puede ver en el siguiente gráfico (Figura 7) de tasa de aprobación por ejecución de tests

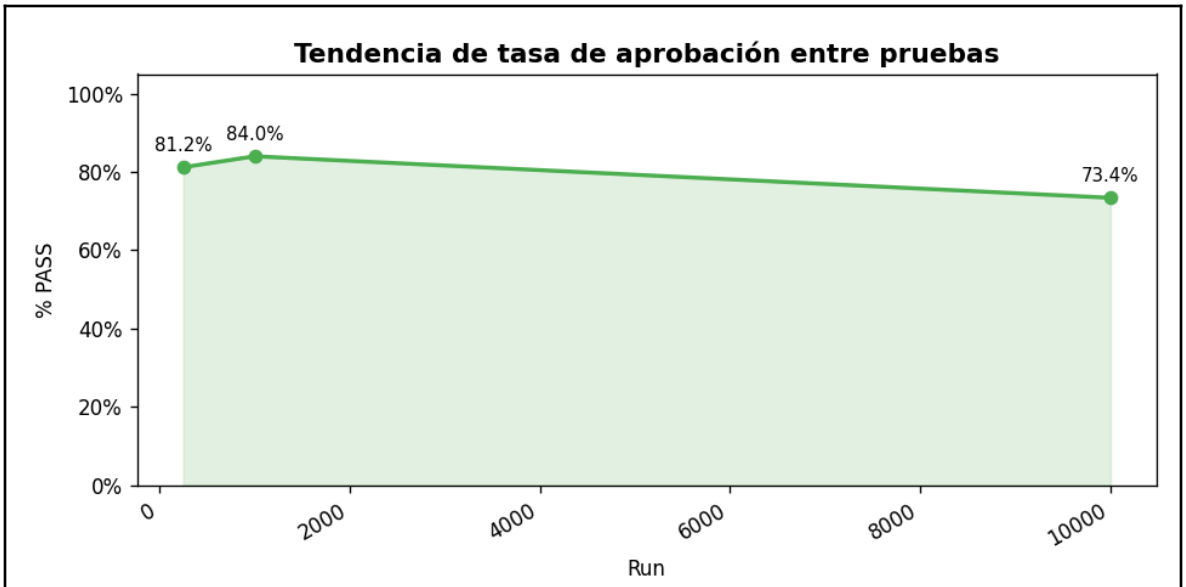

Figura 7. Tendencia de la tasa de aprobación entre pruebas ejecutadas.

Hasta ahora hemos solo analizado la efectividad de los reconocedores desarrollados y anexados, pero también hay que tener en cuenta los tiempos de ejecución y latencia ya que la criticidad de estos resultados depende del contexto y la situación. A continuación se visualiza la latencia para un n=10000, un percentil de latencia responde a la pregunta: "el X% de las llamadas se completaron en menos de N ms".

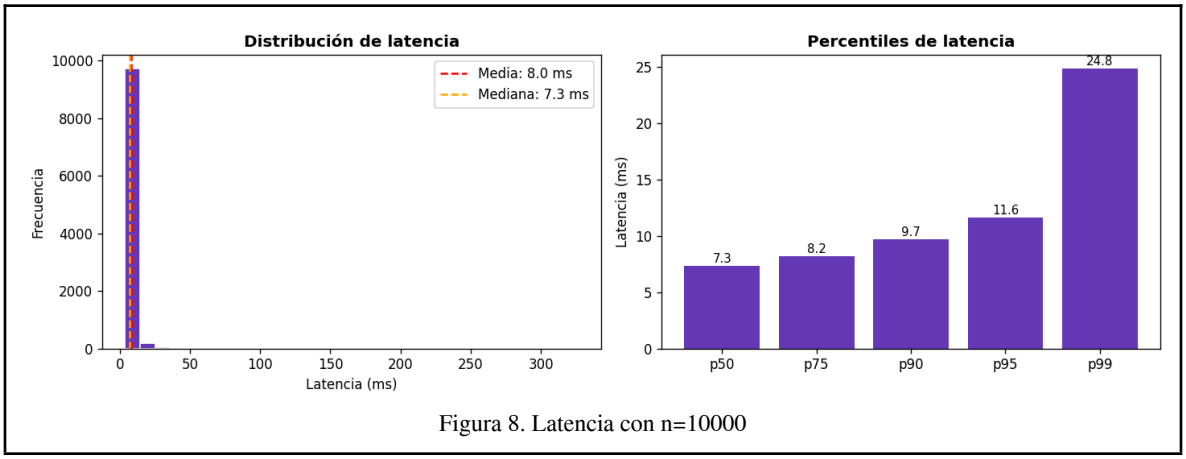

Figura 8. Latencia con n=10000

Acá se puede inferir que el 50% de las muestras tardan ≤ 7.3 ms (es la mediana) en ser analizadas; el 75% tardan ≤ 8.2 ms; el 90% tardan ≤ 9.7 ms; el 95% tardan ≤ 11.6 ms y el 99% tardan ≤ 24.8 ms.

El sistema está muy bien. La distribución (gráfico izquierdo) muestra que casi todas las 10.000 requests caen entre 0-10 ms, con muy pocos outliers (valores atípicos). La diferencia clave es entre p95 → p99: salta de 11.6 ms a 24.8 ms, lo que indica que hay un pequeño porcentaje (~1%) de muestras que por alguna razón tardan ~3x más que la mediana. En producción, el p99 es el más importante porque representa la peor experiencia que un usuario real va a experimentar frecuentemente. Con 24.8 ms de p99, el sistema responde en tiempo completamente aceptable para un servicio de detección de PII.

3 Conclusión

La automatización en la detección de Información de Identificación Personal representa una necesidad creciente en el contexto argentino, no solo por exigencias legales sino también por el aumento de las amenazas en el ecosistema digital. Este proyecto propone una solución innovadora, basada en inteligencia artificial, que busca abordar esa necesidad con un enfoque adaptado al idioma español y al marco normativo local Argentino.

El framework presentado permitirá no solo avanzar en la protección efectiva de los datos personales, sino también promover la adopción de tecnologías de vanguardia en sectores clave como la academia, la industria y el sector público. A largo plazo, este proyecto sienta las bases para una mejora continua en la gestión de datos sensibles y una cultura más robusta en torno a la ciberseguridad en Argentina.

En base a los resultados presentados, se puede enumerar algunas mejoras como

- Mejorar precisión de reconocedores DNI. Para mitigar este comportamiento se plantea: (i) reforzar la validación por contexto, exigiendo palabras clave cercanas («DNI», «documento», «D.N.I.») para elevar la confianza y penalizar o descartar las coincidencias aisladas; (ii) incorporar reglas de exclusión para patrones que no constituyen PII (montos con símbolo monetario, cantidades seguidas de unidades, prefijos de transacción o lote); (iii) apoyarse en el CUIL/CUIT asociado, que sí posee dígito verificador y embebe el número de DNI, para confirmar su validez; y (iv) complementar la detección por expresiones regulares con el análisis lingüístico del modelo de NLP, ponderando el contexto en que aparece la cifra. Estas mejoras buscan aumentar la precisión sin sacrificar el alto recall, prioritario en la detección de PII.
- Expandir la identificación del reconocedor de Obra Social mejorando su precisión
- Agregar más reconocedores como, Pasaporte, CBU, nros de AFIP, placas de auto, entre otros

De esta forma lograr una prueba integral, considerando latencia y con muestras más numerosas para estresar el framework .

Al utilizar python como lenguaje del framework y la herramienta de benchmarking se logra portabilidad de ejecución. Por último, al utilizar componentes de código abierto, se logra una transparencia del 100% en la funcionalidad, lo cual era una de las máximas del proyecto en el cual está inmerso este framework

Declaración sobre el uso de inteligencia artificial generativa

Para la realización de este trabajo se utilizó el modelo de lenguaje generativo Claude (Claude, 2026) como herramienta de asistencia en la generación de muestras de texto de prueba, en el desarrollo del código de la herramienta de benchmarking, y en la redacción y la mejora de la legibilidad del texto. El contenido generado fue revisado, editado y validado por los autores, quienes asumen la responsabilidad final sobre la totalidad del trabajo.

Referencias

Chacón, J. (2025). Implementing a Named Entity Recognition pipeline for public procurement contracts in the Municipality of Padova (Master's thesis, University of Padova).

Cimpanu, C. (2021). Hackers leak LinkedIn 700 million data scrape. The Record. https://therecord.media/hackers-leak-linkedin-700-million-data-scrape

Cisco. (2023). Data privacy benchmark study 2023. https://www.cisco.com/c/dam/en_us/about/doing_business/trust-center/docs/cisco-privacy-benchmark-study-2023.pdf

Claude. (2026). Claude [Modelo de lenguaje generativo]. Anthropic. https://claude.ai/

Docker. (s.f.). Documentation and containerization best practices. https://docs.docker.com/

Explosion AI. (s.f.). es_core_news_md [Modelo de procesamiento del lenguaje natural]. spaCy. https://spacy.io/models/es#es_core_news_md

Facebook AI Research (FAIR). (2019). Detectron2 [Software]. GitHub. https://github.com/facebookresearch/detectron2

FastAPI. (s.f.). Modern web framework for APIs. https://fastapi.tiangolo.com/

General Data Protection Regulation (GDPR). (s.f.). https://gdpr-info.eu/

Gobierno de Argentina. (2000). Ley 25.326 - Protección de datos personales. https://www.argentina.gob.ar/normativa/nacional/ley-25326-64790/actualizacion

Heiligenstein, M. X. (2024). Microsoft data breaches: Full timeline through 2024. https://firewalltimes.com/microsoft-data-breach-timeline/

ISO. (2013). ISO/IEC 27001: Information security management systems – Requirements.

Jaided AI. (2020). EasyOCR [Software]. GitHub. https://github.com/JaidedAI/EasyOCR

Khurana, D., Koli, A., Khatter, K., & Singh, S. (2017). Natural language processing: State of the art, current trends and challenges. arXiv. https://doi.org/10.48550/arXiv.1708.05148

Kocaman, V., & Talby, D. (2021). Spark NLP: Natural language understanding at scale. Software Impacts, 8, 100058. https://doi.org/10.1016/j.simpa.2021.100058

Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., & Stoyanov, V. (2019). RoBERTa: A robustly optimized BERT pretraining approach. arXiv. https://arxiv.org/abs/1907.11692

Microsoft. (s.f.). Presidio: Open-source PII protection tool. https://github.com/microsoft/presidio

Mora-Delgado, J., Ramos-Ruperto, L., Pardilla, M. J., Sicilia, M. Á., Rodríguez-González, A., Sempere, J. M., & Puchades, R. (2026). Generative AI: Foundational models, natural language processing and large language models. Revista Clínica Española, 226(1), 502413. https://doi.org/10.1016/j.rceng.2025.502413

National Institute of Standards and Technology (NIST). (2010). Guide to protecting the confidentiality of personally identifiable information (PII) (NIST SP 800-122).

OpenCV Foundation. (s.f.). OpenCV [Software]. GitHub. https://github.com/opencv/opencv

OpenPipe. (2026). Documentación oficial. https://docs.openpipe.ai/introduction

Rajgarhia, H., et al. (2025). An evaluation study of hybrid methods for multilingual PII detection. arXiv.

Revista Innovación Seguridad. (2024, abril 25). Fuga de datos en el RENAPER. https://revistainnovacion.com/nota/12307/fuga_de_datos_en_el_renaper

Schelb, J. (s.f.). roberta-ner-multilingual [Modelo de aprendizaje automático]. Hugging Face. https://huggingface.co/JorgeSchelb/roberta-ner-multilingual

Segu-Info. (2013, octubre 28). Descargar todas las fotos del padrón electoral argentino. https://blog.segu-info.com.ar/2013/10/descargar-todas-las-fotos-del-padron.html

Segu-Info. (2024, abril). RENAPER solo expone las deficiencias del Estado argentino en ciberseguridad. https://blog.segu-info.com.ar/2024/04/renaper-solo-expone-la-deficiencias-del.html

Seow, W. L., Chaturvedi, I., Hogarth, A., Mao, R., & Cambria, E. (2025). A review of named entity recognition: From learning methods to modelling paradigms and tasks. Artificial Intelligence Review, 58, 315. https://doi.org/10.1007/s10462-025-11321-8

Silva, P., Gonçalves, C., Godinho, C., Antunes, N., & Curado, M. (2020). Using NLP and machine learning to detect data privacy violations. En IEEE INFOCOM 2020 Workshops (pp. 972–977). IEEE.

Skim AI. (2020). Tutorial: Cómo ajustar BERT para el reconocimiento de entidades con nombre (NER).

spaCy. (s.f.). Industrial-strength NLP in Python. https://spacy.io/

Splunk. (s.f.). Defining and detecting personally identifiable information (PII) in log data.

Tesseract OCR project. (s.f.). Tesseract OCR [Software]. GitHub. https://github.com/tesseract-ocr/tesseract

Ultralytics. (2023). YOLOv8 [Software]. GitHub. https://github.com/ultralytics/ultralytics

World Economic Forum. (2024). The Global Risks Report 2024. https://www3.weforum.org/docs/WEF_The_Global_Risks_Report_2024.pdf