

DETECCIÓN DE INFORMACIÓN DE IDENTIFICACIÓN PERSONAL (PII) EN ARGENTINA A TRAVÉS DE MODELOS DE INTELIGENCIA ARTIFICIAL OPTIMIZADOS PARA EL ESPAÑOL

Zea Cardenas Milagros, Barrionuevo Ileana, Gibellini Fabián, Parisi Germán, Ruhl Analía, Notreni Juliana, Ciceri Leonardo, Auquer Marcelo, Bertola Federico, Sanchez Ignacio

1 CONTEXTO

La creciente digitalización ha incrementado el volumen de datos sensibles, entre los cuales la Información Personal Identificable (PII) representa un activo crítico. Su exposición puede generar consecuencias legales, económicas y reputacionales significativas.

Si bien existen marcos regulatorios como la Ley 25.326 en Argentina, la detección de PII en documentos organizacionales resulta compleja debido a su presencia en formatos no estructurados y a la posibilidad de inferencia a partir de datos contextuales.

En este contexto, surge la necesidad de desarrollar soluciones tecnológicas que permitan identificar y proteger la PII de manera automatizada.

2 LÍNEAS DE I+D

El proyecto se enmarca en el área de la ciberseguridad, con foco en la protección de la PII mediante técnicas de Inteligencia Artificial. Se adopta un enfoque empírico basado en el método científico, integrando:

- Procesamiento de Lenguaje Natural (NLP)
- Modelos de Machine Learning
- Desarrollo de soluciones adaptadas al idioma español y contexto argentino

La propuesta se desarrolla bajo el enfoque de Investigación, Desarrollo e Innovación (I+D+i), combinando teoría y aplicación práctica.

3 RESULTADOS

Se desarrolló una prueba de concepto funcional basada en:

- Integración de Microsoft Presidio para detección y anonimización de PII
- Arquitectura contenerizada mediante Docker
- Implementación de una API REST para consumo de servicios

El sistema incorpora:

- Motor NLP para análisis de texto
- Detectores de entidades sensibles
- Reconocedores específicos para el contexto argentino (DNI, CUIL/CUIT)

Los resultados permiten:

- Detectar automáticamente PII en documentos
- Adaptar la solución al idioma español
- Validar la viabilidad de herramientas basadas en IA para protección de datos

4 RECURSOS HUMANOS

El proyecto involucra docentes, investigadores y estudiantes de la UTN – Facultad Regional Córdoba, integrantes del GISSIC.

Los estudiantes participan como becarios en actividades de investigación y desarrollo, fortaleciendo su formación en ciberseguridad, inteligencia artificial y desarrollo de software.

Asimismo, se promueve la difusión del conocimiento mediante charlas, talleres y conferencias a nivel académico.