DIARIO DIGITAL DEV / RECURSO DE LECTURA

Datos sintéticos: cómo probar software e IA con casos útiles

Generar registros ficticios puede ayudar a ensayar una integración o un asistente. El valor está en los errores que permiten detectar y en comprobar por separado su utilidad y el riesgo de exposición.

Ideas clave del artículo

- Definir el objetivo: probar reglas y excepciones es distinto de reproducir una población para análisis estadístico.

- Conservar relaciones entre registros y resultados esperados, incluyendo casos inválidos identificados.

- Evaluar privacidad y utilidad por separado; los datos sintéticos pueden conservar valores del origen.

- Guardar escenarios de evaluación independientes y verificar el comportamiento fuera de los ejemplos usados para ajustar el sistema.

Notas para tu equipo

_____

_____

_____

_____

Artículo y fuentes

https://www.digital-dev.cl/blog/datos-sinteticos-pruebas-software-ia/

Johnny Greco y colaboradores, arXiv: NeMo Data Designer: An Extensible Framework for Multimodal Synthetic Data Generation
https://arxiv.org/abs/2609.17699

Snowflake: Using synthetic data in Snowflake
https://docs.snowflake.com/en/user-guide/synthetic-data

National Institute of Standards and Technology: Differentially Private Synthetic Data
https://www.nist.gov/blogs/cybersecurity-insights/differentially-private-synthetic-data