

// documento complementario

Metodología.

Cómo construimos y analizamos el corpus de **Después de la brecha** · Ene 2024 – Sep 2026

Este documento acompaña la nota de investigación **Después de la brecha**. Explica de dónde vienen los datos, cómo se clasificaron y validaron, qué pueden y qué no pueden decir, y las reglas éticas que seguimos.

// 01

Fuente y periodo.

Escucha social con una consulta única y versionada para todo el periodo (diccionario v0.3.0), del 1 de enero de 2024 al 30 de septiembre de 2026, sobre contenido en español ubicado en México. La consulta se basa en términos fuertes de incidentes de ciberseguridad: «hackeo» solo cuenta cuando aparece junto a una institución, y Pegasus solo en contexto de espionaje.

// 02

Flujo muestral.

PASO	MENCIONES	NOTA
Capturadas	258 069	3 exportaciones, una sola versión de la consulta
– duplicados exactos	9 111	Mismo identificador en dos exportaciones
– contenido basura	56	Plantillas de precio de criptomonedas
Corpus procesado	248 902	95 801 publicaciones (base de las tendencias) y 153 101 comentarios (se analizan aparte)

Las tendencias se calculan con las publicaciones. Los comentarios se analizan aparte porque unas pocas publicaciones virales concentran decenas de miles; sumarlos distorsionaría cualquier tendencia.

// 03

Clasificación y validación.

- **Etiquetas léxicas.** Los temas (por ejemplo, CURP, datos biométricos, registro de líneas celulares o transparencia), las instituciones y el tipo de respuesta (desmentido, reconocimiento, notificación) se asignan por términos en el texto.
- **Validación humana por muestras.** Muestras aleatorias revisadas por personas, con criterios escritos de inclusión y exclusión. La precisión observada va de 0,43 a 0,91 según el grupo de términos; las muestras todavía son pequeñas.

// 04

Rama de robo y venta de datos.

- Consulta separada (v0.4.0) sobre ofertas para robar cuentas, espiar conversaciones y vender bases de datos: unas 268 mil menciones en español sin filtro de país; unas 38 mil vienen de perfiles ubicados en México.
- El vínculo con México se asigna por señales en el contenido (teléfonos +52, instituciones, lugares, moneda y comercios mexicanos), no por la ubicación que asigna la herramienta, que resultó poco confiable. Esta regla se sigue validando con revisión humana.
- Nunca se descargan, compran ni abren datos robados.

// 05

Cronología de incidentes.

- Se reconstruye con la cobertura de prensa del corpus: cada uno de los 40 episodios tiene al menos dos medios; los casos clave se verificaron con declaraciones oficiales. La lista no es exhaustiva.
- **Marcas de análisis.** ♦ señala los incidentes en los que, según lo reportado, el daño recae en particular en mujeres, niñez, personas mayores, víctimas, denunciantes o informantes, o periodistas. Para comparar respuestas, separamos los incidentes en que los datos salieron de la institución de los demás (sitios alterados, cuentas secuestradas, caídas).
- La verificación corrige: un episodio resultó ser de Argentina y se excluyó; un hallazgo que no resistió las pruebas se retiró.
- Los alcances son los reportados, no cifras validadas por las autoridades.

// 06

Alcance y límites.

- **La clasificación es automática y la validación humana, parcial.** Las cifras sirven para comparar, no como magnitudes absolutas.
- **Medimos conversación, no incidentes.** Más menciones significan más atención pública, no más ataques.
- **Las fechas coinciden, pero eso no prueba causas.** La extinción del INAI coincide con un nuevo gobierno y nuevas políticas de identidad digital. Además, el INAI ya estaba sin quórum desde abril de 2023: 2024 no es una línea base de supervisión plena.
- **Cobertura incompleta.** Con el filtro de país casi no se captura TikTok, y los espacios privados o cerrados no son observables con este método.

// 07

Auditabilidad, ética y transparencia.

- **Auditabilidad.** Cada decisión se registra con fecha, responsable y justificación; los datos originales se conservan con huellas de integridad; cada exclusión tiene un motivo y los scripts son reproducibles.
- **Ética.** No reproducimos textos íntegros ni nombres de personas afectadas, y no enlazamos datos filtrados.
- **Transparencia.** Libro de códigos, bitácora de decisiones y flujo muestral, a solicitud en info@cyber-cats.org.

Licencia: © 2026 CyberCats Academy A.C. · [Creative Commons Atribución-NoComercial-SinDerivadas 4.0 Internacional \(CC BY-NC-ND 4.0\)](#).

Cómo citar: CyberCats Academy A.C. (2026). *Después de la brecha: metodología*. Octubre de 2026.