Source dossier

It did not break in, it logged in — the first agent-executed breach notification, and the limb of Article 4(12) nobody plans for

Edition 2026-09-18 · narrative: The control the AEPD names is token scope, and the widest-scoped tokens in a normal estate belong to the reader's own agents; the breach was alteration, which Article 33(3)(a) still requires you to describe. · compiled 2026-09-18

Claims index

Claim	Statement	Status	Sources
CLAIM-2026-09-18-01	The AEPD published on 14 September 2026 that it received the first notification of a personal data breach in which the incident was executed by means of an AI agent using a well-known language model.	verified	agenciaespa2026primera
CLAIM-2026-09-18-02	The AEPD describes the chain as: a search for vulnerabilities in generic files, a successful login, then an autonomous search for vulnerabilities inside the application, which allowed the agent to modify personal data and access invoices.	verified	agenciaespa2026primera
CLAIM-2026-09-18-03	The AEPD states that use of a particular AI model does not imply the model or its provider's infrastructure was compromised, nor that the tool was designed for malicious activity; what matters is that a third party used the agent as an instrument to chain the phases of the attack.	verified	agenciaespa2026primera
CLAIM-2026-09-18-04	The AEPD states that this first notification does not permit the assertion of a statistical trend.	verified	agenciaespa2026primera
CLAIM-2026-09-18-05	The AEPD's third observation is that an agent obtaining an account, an API key or a token with excessive permissions can operate at machine speed and reach different services before the organisation detects anomalous behaviour.	verified	agenciaespa2026primera
CLAIM-2026-09-18-06	The AEPD received 2,765 personal data breach notifications in 2025, 80 per cent from the private sector, and escalated 11 of them for further investigation.	verified	agenciaespa2026la
CLAIM-2026-09-18-07	Article 4(12) GDPR defines a personal data breach to include alteration of personal data, not only unauthorised disclosure or access.	verified	eurlex2016regulatio

Claim	Statement	Status	Sources
CLAIM-2026-09-18-08	Article 33(1) GDPR requires notification to the supervisory authority without undue delay and where feasible not later than 72 hours after the controller becomes aware of the breach, and Article 33(3)(a) requires the notification to describe the nature of the breach including the categories and approximate number of personal data records concerned.	verified	eurlex2016regulatio
CLAIM-2026-09-18-09	The AEPD blog post cites CCN-CERT BP/36, the Centro Criptologico Nacional's good-practice guide on offensive AI, published in June 2026.	verified	agenciaespa2026primera, laboratorio2026gua

Only claims marked **verified** may carry a post. **single-source** may appear as hedged context; **disputed** must not be published at all.

Primera notificación de una brecha de datos personales causada por un ataque ejecutado mediante un agente de IA

PRIMARY

Publisher	Agencia Española de Protección de Datos
Published	2026-09-14
Retrieved	2026-09-18
Cite key	agenciaespa2026primera
URL	https://www.aepd.es/prensa-y-comunicacion/blog/primera-notiviacion-brecha-datos-personales-causada-por-ataque-ejecutado-mediante-agente-ia

Claims drawn from this source. CLAIM-2026-09-18-01, CLAIM-2026-09-18-02, CLAIM-2026-09-18-03, CLAIM-2026-09-18-04, CLAIM-2026-09-18-05, CLAIM-2026-09-18-09

AEPD blog, 14 September 2026 — "Primera notificación de una brecha de datos personales causada por un ataque ejecutado mediante un agente de IA", signed Francisco Pérez Bes.

Key passages (original Spanish, working translation follows each):

> "*La Agencia Española de Protección de Datos ha recibido la primera notificación de una brecha de datos personales en la que el incidente habría sido ejecutado mediante un agente de inteligencia artificial que utilizó un conocido modelo de lenguaje.*"

The AEPD has received the first notification of a personal data breach in which the incident was reportedly executed by means of an AI agent that used a well-known language model.

> "*El agente atacante inició una búsqueda de vulnerabilidades en archivos genéricos, y realizó un login correcto. Una vez accedió al sistema, comenzó a buscar, de forma autónoma, vulnerabilidades en la aplicación, lo que, una vez conseguido, le permitió modificar datos personales y acceder a facturas.*"

The attacking agent began a search for vulnerabilities in generic files, and performed a successful login. Once it had accessed the system it began to search, autonomously, for vulnerabilities in the application, which once achieved allowed it to modify personal data and access invoices.

> "*Asimismo, la utilización de un concreto modelo de IA tampoco implica que el modelo o la infraestructura de su proveedor hayan sido comprometidos ni que la herramienta haya sido diseñada para desarrollar actividades maliciosas. Ahora bien, lo relevante desde la perspectiva de la protección de datos es que un tercero habría utilizado un*

> *agente de IA como instrumento para encadenar exitosamente distintas fases del ataque.*"

Likewise, the use of a particular AI model does not imply that the model or its provider's infrastructure was compromised, nor that the tool was designed to carry out malicious activities. What is relevant from the data protection perspective is that a third party used an AI agent as an instrument to successfully chain together the different phases of the attack.

> "*Esta primera notificación no permite afirmar una tendencia estadística, aunque sí constituye una señal significativa de que los ataques apoyados en inteligencia artificial han dejado de ser un riesgo teórico.*"

This first notification does not permit the assertion of a statistical trend, although it does constitute a significant signal that AI-supported attacks have ceased to be a theoretical risk.

The four observations ("¿Por qué es relevante esta primera notificación?")

1. Risk analyses must expressly incorporate attacks assisted or executed by AI: "no basta con incluir una referencia genérica a malware, phishing o acceso no autorizado".
2. Response times must be revised: "Los procedimientos diseñados para ataques ejecutados manualmente pueden resultar insuficientes cuando un agente analiza simultáneamente múltiples activos, prueba distintas vías de acceso y adapta su comportamiento con rapidez."
3. Digital identities and credentials increase in importance: "Un agente que obtiene una cuenta, una clave API o un token con permisos excesivos puede operar a la velocidad de una máquina y acceder a diferentes servicios antes de que la organización detecte un comportamiento anómalo."
4. Security of processing cannot depend on manual intervention alone.

The post cites CCN-CERT BP/36, *Buenas prácticas frente al modelo de IA ofensiva*, published by the Centro Criptológico Nacional.

The post names neither the affected organisation, nor the agent, nor the model.

La AEPD recibió en 2025 más de 2.700 notificaciones de brechas de datos personales

PRIMARY

Publisher	Agencia Española de Protección de Datos
Published	2026-01-23
Retrieved	2026-09-18
Cite key	agenciaespa2026la
URL	https://www.aepd.es/prensa-y-comunicacion/notas-de-prensa/la-aepd-recibio-en-2025-mas-2.700-notificaciones-brechas

Claims drawn from this source. CLAIM-2026-09-18-06

AEPD press release, 23 January 2026.

- 2,765 personal data breach notifications received in 2025 (a record).
- 80% from the private sector, 20% from the public sector.
- 11 of the 2,765 were escalated for further investigation — high-severity cases showing "indicios de falta de diligencia".
- Organisations sent over 200 million communications to affected individuals about high-risk incidents during 2025.
- Largest-impact incident types: ransomware and system intrusion with exfiltration; compromised credentials giving access to corporate VPNs and web applications; CRM platform breaches via service-provider compromise; administrative error (misdirected transmissions, accidental exposure).

Regulation (EU) 2016/679 (General Data Protection Regulation)

PRIMARY

Publisher	EUR-Lex
Published	2016-04-27
Retrieved	2026-09-18
Cite key	eurlex2016regulatio
URL	https://eur-lex.europa.eu/eli/reg/2016/679/oj

Claims drawn from this source. CLAIM-2026-09-18-07, CLAIM-2026-09-18-08

Regulation (EU) 2016/679 (GDPR) — verbatim extracts from the consolidated EUR-Lex text.

Article 4(12):

> "'*personal data breach*' *means a breach of security leading to the accidental or unlawful destruction, loss, alteration, unauthorised disclosure of, or access to, personal data transmitted, stored or otherwise processed*"

Article 32(1)(b):

> "*the ability to ensure the ongoing confidentiality, integrity, availability and resilience of processing systems and services*"

Article 33(1):

> "*In the case of a personal data breach, the controller shall without undue delay and, where feasible, not later than 72 hours after having become aware of it, notify the personal data breach to the supervisory authority competent in accordance with Article 55, unless the personal data breach is unlikely to result in a risk to the rights and freedoms of natural persons. Where the notification to the supervisory authority is not made within 72 hours, it shall be accompanied by reasons for the delay.*"

Article 33(3)(a):

> "*describe the nature of the personal data breach including where possible, the categories and approximate number of data subjects concerned and the categories and approximate number of personal data records concerned*"

Guía práctica del CCN-CERT para hacer frente a la IA ofensiva

SECONDARY

Publisher	Laboratorio de Innovación AEPD
Published	2026-06-25
Retrieved	2026-09-18
Cite key	laboratorio2026gua
URL	https://laboratorio.aepd.es/novedades/guia-practica-del-ccn-cert-para-hacer-frente-la-ia-ofensiva

Claims drawn from this source. CLAIM-2026-09-18-09

CCN-CERT BP/36, Guía de buenas prácticas frente al modelo de IA ofensiva, Centro Criptológico Nacional, June 2026 — summarised by the AEPD's own innovation lab (laboratorio.aepd.es), page dated 25 June 2026.

Scope: how AI lets attackers "automatizar, acelerar y ampliar a gran escala ataques ya conocidos". Threat areas analysed include advanced phishing and deepfakes, automated malicious code generation, accelerated vulnerability exploitation, mass reconnaissance, attacks on language-model systems, and agentic AI systems with autonomous execution.

Recommendations centre on vulnerability management, identity protection, secure development lifecycle, supplier oversight and operational environment security, promoting "supervisión continua, la resiliencia y la seguridad desde el diseño".

Note: the CCN-CERT's own page (ccn-cert.cni.es) returned HTTP 403 to this run; the content above is taken from the AEPD laboratory's summary, which is an aepd.es property. Treated as secondary.