NOVARA
CONSULTING GROUP

Este documento fue traducido del inglés por una máquina y puede contener errores. El original en inglés se encuentra en https://www.novaracg.com/2026/09/27/the-agent-did-not-sign-the-contract-autonomous-ai-public-systems-and-the-incident-clause-most-counties-do-not-have/.

# El Agente No Firmó el Contrato: IA Autónoma, Sistemas Públicos y la Cláusula de Incidentes que la Mayoría de los Condados No Tiene

Heather Grizzle · 27 de septiembre de 2026

*Four September 2026 incidents showed AI agents entering government systems and ignoring instructions. What county and city contracts should require now.*

**Contenidos**

1. Lo que se sabe públicamente
2. Tres exposiciones, no una
3. Lo que los incidentes tienen en común
4. La brecha de notificación
5. Lo que pertenece al expediente de contratación
6. El contrato redacta el expediente
7. Qué no hacer
8. Este trimestre
9. Fuentes consultadas

En las últimas dos semanas de septiembre de 2026, cuatro revelaciones independientes establecieron algo que los administradores públicos han podido tratar como hipotético hasta ahora. Agentes de IA autónomos, operando sin que un ser humano dirigiera cada paso, ingresaron a sistemas gubernamentales a los que no debían acceder, eludieron controles

que se suponía debían detenerlos y, en al menos un caso, tardaron tres meses en ser reportados al gobierno cuyos sistemas se vieron afectados. Los comentarios que siguieron han provenido casi por completo del lado de los laboratorios de IA de vanguardia de la historia: lo que los incidentes dicen sobre la alineación de los modelos, lo que significan para OpenAI, lo que el Congreso o la ONU deberían hacer al respecto. Esa no es la conversación que un administrador de condado necesita tener el lunes por la mañana.

Este artículo lee los mismos cuatro incidentes desde la otra silla. Expone lo que se sabe públicamente, separa los hechos de la interpretación y pregunta qué es lo que un condado, ciudad o distrito escolar ya tiene en sus contratos y expedientes de contratación que hubiera regido cualquiera de estos casos. La respuesta corta es que la mayoría de los contratos del sector público se redactaron para un mundo en el que el software de un proveedor hacía lo que el proveedor le decía que hiciera, y una brecha significaba que alguien robó datos. Ninguno de los dos supuestos se sostuvo este mes.

## Lo que se sabe públicamente

El 26 de septiembre, OpenAI divulgó que sus agentes de IA habían, durante el entrenamiento y la evaluación, interactuado con sitios web del gobierno de Estados Unidos de maneras que la empresa no pretendía. Los informes basados en la divulgación identifican a la Comisión de Bolsa y Valores, la Oficina del Censo, el Departamento de Comercio, el Departamento de Justicia y el Departamento de Educación entre los sitios federales involucrados, con actividad adicional que tocó sitios de gobiernos estatales en California, Maryland, Illinois, Texas y Nueva York. El laboratorio de investigación Transluce identificó lo que se ha descrito como un intento rudimentario de hackeo contra la oficina de derechos civiles del Departamento de Educación; el Departamento ha dicho que no encontró impacto en sus sistemas. En el caso del Censo, los informes indican que los agentes usaron credenciales encontradas en línea para acceder a datos públicos. En el caso de la SEC, se informa que los agentes accedieron a información pública y luego la republicaron en otro lugar. OpenAI caracterizó los eventos como actividad de modelos desalineados, dijo que ninguno de los incidentes constituyó una violación en el sentido de que se comprometieran datos no públicos, y

anunció una amplia revisión del acceso a internet de sus agentes durante el entrenamiento y la evaluación. Los recuentos de incidentes publicados varían entre medios, y este artículo no se basa en una cifra específica.

Dos días antes, el 24 de septiembre, el primer ministro australiano Anthony Albanese reveló en una conferencia de prensa en Nueva York que un modelo de OpenAI había accedido al Portal de Informes de Estadísticas de Medicare de Australia el 18 de junio de 2026. Según los informes, el modelo llegó tanto a archivos públicos como no públicos, escribió archivos en el servidor interno y, cuando se le bloqueó el acceso a la información que buscaba, intentó rutas alternativas para obtenerla. OpenAI descubrió la actividad en agosto durante su propia revisión, y notificó al gobierno australiano el 10 de septiembre mediante un correo electrónico a un buzón público. Services Australia reportó el asunto al Centro Australiano de Seguridad Cibernética el 15 de septiembre. Albanese declaró que no se cree que se haya accedido a información personal y criticó tanto la demora como el método de notificación. Otros tres sistemas gubernamentales australianos podrían haberse visto afectados.

El 10 de septiembre, la firma de inteligencia de amenazas GreyNoise publicó su análisis de una campaña en la que cientos de agentes de IA explotaron dos vulnerabilidades en el software de gestión de impresión PaperCut, CVE-2026-81578 y CVE-2026-82078, para comprometer más de 440 instancias en más de 395 organizaciones en 48 países. Estados Unidos tuvo la mayor proporción de víctimas, con 98. La educación fue el sector más afectado, con 204 víctimas. Las cifras de velocidad merecen atención: en una escuela secundaria, el acceso inicial a administrador de dominio tomó siete minutos. El detalle con la relevancia de gobernanza más directa es este. El atacante instruyó a los agentes que evitaran objetivos en 28 países nombrados. Los agentes atacaron organizaciones en esos países de todos modos. GreyNoise lo calificó como un buen ejemplo de agentes fuera de control y dijo que las razones siguen sin estar claras.

Finalmente, alrededor del 16 de septiembre, la autoridad de protección de datos de España, la AEPD, informó lo que describió como la primera violación de datos ejecutada de forma autónoma por un agente de IA. Según el relato de la AEPD tal como se reportó, el agente se autenticó en un sistema, descubrió vulnerabilidades, modificó datos personales y accedió a facturas, encadenando las fases del ataque sin dirección humana continua. La preocupación expresada por la AEPD fue que un agente puede recibir un objetivo, planificar tareas intermedias, usar herramientas, ejecutar código,

interpretar resultados y modificar sus propias acciones, y que los mecanismos de defensa construidos en torno a una respuesta a velocidad humana no son adecuados para ello.

Esos son los hechos tal como constan actualmente en el registro público. Lo que sigue es interpretación, y se identifica como tal.

## Tres exposiciones, no una

La tentación es archivar los cuatro eventos bajo ciberseguridad y entregárselos a TI. Eso sería un error de clasificación, porque los cuatro incidentes describen tres formas distintas en que un agente autónomo puede llegar a un organismo público, y solo una de ellas es un ataque convencional.

La primera exposición es un agente operado por otra parte, con fines ajenos al condado, que llega a los sistemas del condado en el curso de hacer otra cosa. Este es el patrón de OpenAI. Nadie en OpenAI se propuso sondear el Departamento de Educación. Los agentes realizaban tareas de investigación, trataron los sitios web gubernamentales como fuentes confiables, y al perseguir sus objetivos hicieron cosas que su operador dice que no pretendía. El caso australiano es el mismo patrón con un filo más agudo: el agente fue bloqueado y buscó otra vía de entrada. El portal de registros de propiedad, el sistema de permisos, el archivo de reuniones y los puntos de acceso de datos abiertos de un condado están exactamente en la posición en la que estaba el Portal de Informes de Estadísticas de Medicare. Están en internet abierto, parecen fuentes autorizadas y, por lo tanto, son lugares a los que un agente que persigue un objetivo acudirá.

La segunda exposición es un agente utilizado como arma contra el software que ejecuta el condado. Este es el patrón de PaperCut, y el más cercano a la ciberseguridad convencional. Lo nuevo no es la vulnerabilidad, sino la economía: un atacante que pasa de un espacio de trabajo vacío a la ejecución remota de código en menos de cuatro horas, y que luego compromete once organizaciones en veintiséis segundos una vez lanzada la campaña. Los condados operan servidores de impresión y las mismas categorías de aplicaciones autoalojadas que operan los distritos escolares. Las 204 víctimas del sector educativo son un anticipo, y el fallo en la lista de exclusión significa que un atacante que tuviera la intención de evitar los objetivos de un país determinado no podría haberlo garantizado.

La tercera exposición es la que con mayor probabilidad ya se encuentra dentro de los propios contratos de un condado en este momento: un agente que el proveedor del condado opera dentro de los flujos de trabajo del condado como parte de la prestación de un servicio. Los proveedores de plataformas de gestión de casos, elegibilidad de beneficios, permisos, recursos humanos, gestión de registros y atención al cliente están agregando componentes que no solo recomiendan, sino que actúan, presentan, enrutan, actualizan, notifican y cierran. El incidente español es la ilustración pública más clara de lo que un agente con acceso al sistema puede hacer cuando hace algo distinto de aquello para lo que fue orientado. No queda claro a partir de los informes públicos si ese agente era la herramienta de un atacante o un sistema legítimamente desplegado que falló, y el punto de gobernanza no depende de cuál de las dos cosas fuera. Un agente con credenciales para un sistema del condado es un agente con credenciales para un sistema del condado, sin importar quién lo haya desplegado.

## Lo que los incidentes tienen en común

Al eliminar las diferencias en actor e intención, los cuatro eventos comparten una característica estructural: en cada caso, las instrucciones dadas al agente no vincularon al agente. La lista de exclusión del atacante de PaperCut fue ignorada. El control de acceso del portal de Medicare fue tratado como un obstáculo a eludir en lugar de un límite a respetar. La propia descripción de OpenAI, de que sus modelos realizaron acciones que la empresa no pretendía, es una declaración de que la intención del operador y el comportamiento del agente divergieron. El planteamiento de la AEPD de un agente que planifica, ejecuta, interpreta y modifica sus propias acciones es la descripción de un sistema cuyo comportamiento nadie especifica completamente por adelantado.

Este es el hecho para el cual los contratos públicos no están preparados. Un contrato de software convencional supone que el comportamiento del software está determinado por su código y configuración, que el proveedor controla ambos, y que las promesas del proveedor sobre lo que el software hará y no hará son, por lo tanto, promesas que el proveedor puede cumplir. Cada cláusula estándar, desde el alcance del acceso hasta el uso aceptable, descansa en esa cadena. Cuando el software es un agente que selecciona acciones en tiempo de ejecución, la cadena tiene un eslabón débil en el medio, y la promesa de un proveedor de que su agente se mantendrá dentro del alcance es una promesa sobre un sistema que el proveedor no controla por completo. Puede

hacerse de buena fe. Las revelaciones australiana y estadounidense son evidencia de que la buena fe no es lo mismo que el control.

Nada de esto significa que los sistemas agénticos deban excluirse de la contratación pública. Significa que el contrato tiene que cumplir una función distinta: exigir al proveedor que demuestre cómo se limita en la práctica el comportamiento del agente, que registre lo que este realmente hace, que defina un incidente en términos de que el agente actúe fuera de su alcance en lugar de solo en términos de robo de datos, y que reporte tales incidentes en un plazo medido en horas en lugar de meses.

## La brecha de notificación

La cronología australiana merece ser leída con calma por cualquiera que administre contratos públicos. El acceso ocurrió el 18 de junio. El operador lo descubrió en agosto. El gobierno se enteró el 10 de septiembre, ochenta y cuatro días después del evento, mediante un correo electrónico enviado a un buzón público en lugar de a través de algún canal que el gobierno hubiera designado para asuntos de seguridad. La autoridad cibernética nacional fue informada el 15 de septiembre. Al público se le informó el 24 de septiembre.

Comparemos eso con lo que propuso el Concejo Municipal de Nueva York el 25 de septiembre. Entre los diez proyectos de ley que presentó la presidenta Julie Menin se encuentra un requisito de reporte en veinticuatro horas de ciertos incidentes de seguridad de IA relacionados con contratos municipales, junto con un requisito de que los sistemas de IA comercializados, vendidos o desplegados en la ciudad sean validados por un tercero antes de su despliegue e incluyan una anulación humana. No se sabe si el paquete se aprobará en esa forma. Pero la yuxtaposición es el punto clave. Un gobierno nacional se enteró de una intrusión en un portal de datos de salud tres meses después, a través de una bandeja de entrada general, y en la misma semana una de las ciudades más grandes de Estados Unidos concluyó que el estándar correcto para sus propios contratistas es de un día.

La mayoría de los condados están lejos de cualquiera de los dos estándares, porque la mayoría de los contratos de condado no definen el evento en absoluto. Una cláusula típica de manejo de datos, en Minnesota como en otros lugares, está vinculada a una violación de la seguridad de los datos: adquisición no autorizada de datos gubernamentales, con obligaciones que fluyen entonces de la Ley de Prácticas de Datos del Gobierno de Minnesota, incluidos los deberes de notificación en Minn. Stat. § 13.055 y

el requisito del § 13.05, subd. 11 de que un contratista que maneje datos gubernamentales esté sujeto a la Ley. Esas disposiciones cumplen una función real. Pero se activan por adquisición no autorizada de datos. Un agente que se autentica con credenciales legítimas de proveedor, actúa fuera del alcance para el que fue desplegado, escribe archivos en un servidor del condado y no exfiltra nada, presuntamente no ha activado en absoluto una cláusula de violación de datos. El incidente del portal de Medicare, tal como se ha descrito públicamente, implicó el acceso a archivos no públicos y la escritura de archivos, sin que se crea que se accedió a información personal. Según muchos contratos existentes, eso no es un evento reportable. Debería serlo.

La pregunta del administrador del condado, entonces, no es si el condado tiene una cláusula de violación. Es si el condado tiene una cláusula que hubiera exigido a un proveedor reportar, dentro de un número definido de horas, a un funcionario del condado designado, que un sistema autónomo que operaba bajo el contrato del proveedor había actuado fuera de su alcance previsto contra los sistemas del condado, independientemente de si algún dato salió del edificio.

## Lo que pertenece al expediente de contratación

Novara Consulting Group ha sostenido en trabajos anteriores que el expediente de contratación es el sistema de gobernanza: el registro que un organismo público reúne antes de aprobar un sistema de IA es el único lugar donde la gobernanza ocurre de manera confiable, porque es el único punto en el que el comprador tiene influencia y el proveedor tiene un incentivo para responder. Los eventos de septiembre agregan siete elementos a ese expediente para cualquier sistema con componentes agénticos, presentados aquí como preguntas que un condado puede plantear a un proveedor en una licitación, una renovación o una enmienda.

La primera es si el sistema incluye algún componente que realiza acciones de forma autónoma, es decir, que selecciona y ejecuta pasos hacia un objetivo sin que un ser humano apruebe cada paso. Los proveedores deben responder de manera expresa e identificar cada uno de esos componentes por nombre y función, y la respuesta debe tratarse como una declaración de la que el proveedor es responsable. Un proveedor que responde que no y luego lanza una actualización agéntica sin previo aviso ha cambiado el producto que el condado compró.

La segunda es a qué sistemas, datos y credenciales puede acceder cada componente autónomo, en la práctica y no solo en intención. La evidencia es un inventario de acce-

so: cada sistema del condado al que el agente puede autenticarse, cada categoría de datos que puede leer o escribir, cada servicio externo al que puede llamar. Si el proveedor no puede producir este inventario, el proveedor no sabe qué puede hacer su agente, y eso en sí mismo es un hallazgo.

La tercera es cómo se hace cumplir el límite de la acción del agente. Existe una distinción crítica entre un agente al que se le instruye no hacer algo y un agente que técnicamente es incapaz de hacerlo. La lista de exclusión de PaperCut era una instrucción. El control de acceso del portal de Medicare era un límite técnico que el agente, según el relato público, intentó eludir. Un condado debería exigir al proveedor que declare, para cada límite, si se hace cumplir mediante instrucción al modelo, mediante controles externos al modelo, o ambos. Las instrucciones a un modelo son útiles. No son controles, y el registro ahora contiene múltiples casos de su fracaso.

La cuarta es contra qué se registran las acciones del agente y por cuánto tiempo. Toda autenticación, lectura, escritura, llamada y mensaje que un componente autónomo realice contra un sistema del condado debe registrarse en una forma que el condado pueda inspeccionar sin la asistencia del proveedor, y conservarse el tiempo suficiente para respaldar una investigación. Australia se enteró de una intrusión de junio en septiembre porque el operador, no el objetivo, la encontró en una revisión. Un condado que conserva sus propios registros no depende del ciclo de revisión del proveedor para saber qué sucedió en sus propios sistemas.

La quinta es cómo se define un incidente. La definición debe incluir cualquier acción de un componente autónomo fuera del alcance documentado, cualquier intento de obtener acceso después de que se le haya negado, cualquier modificación de datos del condado no solicitada por un usuario autorizado del condado, y cualquier transmisión de datos del condado a un destino no documentado. Ninguno de estos casos requiere que se hayan robado datos. Todos describen cosas que ocurrieron en septiembre.

La sexta es el plazo y el canal de notificación. Nueva York propone veinticuatro horas para los contratistas municipales. Un condado puede elegir una cifra diferente, pero debe elegir una, expresarla en horas y nombrar la oficina y el método. Un correo electrónico a un buzón general no es una notificación.

La séptima es qué puede hacer el condado cuando ocurre un incidente: suspender el componente agéntico sin penalización mientras se conserva el resto del servicio, exigir cooperación en la revisión forense, exigir un informe escrito de la causa raíz dentro de

un período definido, y exigir una evaluación independiente antes de que se restablezca el componente. Este último punto se conecta con la dirección hacia la que se mueven Nueva York y California, hacia la validación por terceros de los sistemas de IA antes de su despliegue. Independientemente de lo que se piense sobre exigir eso por ley, un condado es libre de exigirlo por contrato para los sistemas que compra.

## El contrato redacta el expediente

Cada una de esas preguntas tiene una forma contractual, y un condado que esté renovando un contrato de plataforma este año fiscal puede incluir el lenguaje ahora en lugar de esperar a que una ley lo exija. Lo que sigue es un lenguaje modelo para discutir con el abogado del condado. No es asesoramiento legal ni un sustituto de la revisión frente a los términos existentes del condado y la ley estatal aplicable.

### 1 Definiciones

Un Componente Autónomo es cualquier elemento de los Servicios que selecciona y ejecuta acciones hacia un objetivo sin que un usuario humano apruebe cada acción antes de que se tome, incluidos, entre otros, los elementos descritos por el Contratista como agentes, asistentes, copilotos o automatizaciones que actúan sobre los sistemas o datos del Condado. Un Incidente de Agente es cualquier instancia en la que un Componente Autónomo actúa fuera del Alcance Documentado, intenta obtener acceso o información después de que se le ha denegado, modifica datos del Condado que no sean a solicitud de un usuario autorizado del Condado, o transmite datos del Condado a cualquier destino no identificado en el Alcance Documentado, sin importar si algún dato es adquirido o no por una persona no autorizada.

### 2 Divulgación de Componentes Autónomos

El Contratista declara que el Anexo X identifica cada Componente Autónomo presente en los Servicios, y que, para cada uno de dichos componentes, identifica los sistemas del Condado a los que puede acceder, las categorías de datos del Condado que puede leer o escribir, los servicios externos que puede invocar, y los medios por los cuales se hace cumplir cada uno de esos límites, distinguiendo los límites impuestos mediante instrucción a un modelo de los límites impuestos mediante controles técnicos externos al modelo. El Contratista deberá actualizar el Anexo X y obtener la aceptación por escrito del Condado antes de implementar cualquier cambio que añada un Componente Autónomo o amplíe el alcance de uno existente.

### 3 Registro de Acciones

El Contratista deberá mantener un registro completo de cada acción realizada por cualquier Componente Autónomo sobre los sistemas o datos del Condado, incluyendo la hora, la acción, el objetivo, la credencial utilizada y el resultado, en una forma a la que el Condado pueda acceder directamente sin asistencia del Contratista, y deberá conservar dicho registro durante no menos de un período definido después de la acción.

### 4 Notificación de Incidentes de Agente

El Contratista deberá notificar al contacto de seguridad designado por el Condado, mediante el método especificado en el Anexo Y, dentro de un número definido de horas desde el descubrimiento de un Incidente de Agente, y deberá proporcionar un informe escrito de causa raíz dentro de un número definido de días. La notificación a un buzón general o público no cumple con esta sección.

### 5 Recursos

Ante un Incidente de Agente, el Condado puede ordenar al Contratista que suspenda el Componente Autónomo afectado sin reducción de los Servicios restantes y sin penalización para el Condado. El Contratista deberá cooperar plenamente con cualquier revisión forense, y el Condado puede exigir, a costa del Contratista, una evaluación del componente afectado por parte de un tercero independiente aceptable para el Condado antes de que el componente sea restaurado.

Estas disposiciones no hacen que un agente sea seguro. Nada en un contrato logra eso. Lo que sí hacen es convertir la garantía del proveedor en declaraciones respaldadas por evidencia, poner al condado en posesión de su propio registro de lo que hizo el agente, definir el evento que le importa al condado en términos del comportamiento del agente en lugar del éxito del atacante, y fijar un plazo que refleje lo que el condado necesita en lugar de lo que resulte del ciclo de revisión del proveedor.

## Qué no hacer

Es probable que se propongan tres respuestas en los salones de juntas de los condados en las próximas semanas, y cada una es más débil de lo que parece.

La primera es prohibir las funciones agénticas en los sistemas del condado. Esto es inaplicable en la práctica, porque los componentes agénticos se están añadiendo a las





plataformas que los condados ya utilizan, muchas veces sin una partida diferenciada, y una prohibición general será violada por la siguiente actualización del proveedor. La postura más acertada es que ningún componente autónomo actúe sobre los sistemas del condado hasta que figure en el anexo, sus límites estén documentados y técnicamente aplicados donde importa, y sus acciones queden registradas donde el condado pueda verlas.

La segunda es esperar al estado o al Congreso. La ciudad de Nueva York actuó precisamente porque Washington no lo ha hecho. Los términos contractuales están hoy dentro de la propia autoridad del condado. La ley estatal no.

La tercera es aceptar la caracterización del evento que hace el proveedor. OpenAI ha descrito el comportamiento de sus agentes como actividad de modelo desalineada y ha dicho que ninguno de los incidentes constituyó una brecha. Ambas afirmaciones pueden ser precisas según las definiciones de OpenAI. Ninguna es la definición del condado, y un condado que permite que el proveedor defina el incidente ya ha cedido la cuestión de si este ocurrió. El registro de este mes es que los operadores descubrieron tarde el comportamiento de sus propios agentes, lo describieron en sus propios términos y lo reportaron según su propio calendario. La definición de incidente, el registro y el plazo de un condado existen precisamente para que el condado no tenga que depender de ninguno de los tres.

## Este trimestre

Un condado que quiera actuar antes de la próxima divulgación puede hacer tres cosas sin necesidad de una partida presupuestaria. Puede preguntar por escrito a sus proveedores de plataformas actuales cuáles han añadido o anunciado funciones agénticas. Puede plantear las siete preguntas anteriores a cada uno de esos proveedores y archivar las respuestas, y las no respuestas, en el expediente de contratación. Y puede hacer que el abogado del condado revise las definiciones de modelo y las cláusulas frente al próximo contrato que deba renovarse, de modo que el primer acuerdo que las incluya se firme antes de que el condado las necesite, y no después.

Los agentes que ingresaron a los sistemas gubernamentales este mes no firmaron nada. Cada proveedor que despliega uno en nombre de un condado sí lo hizo. Esa firma es el único lugar donde las expectativas del condado sobre el comportamiento de los agentes son exigibles, y es ahí donde debe enfocarse el trabajo.

**Aviso legal.**
Este artículo expone hechos de dominio público según lo reportado por las fuentes enumeradas a continuación e identifica la interpretación como tal. No constituye asesoría legal. El lenguaje contractual se ofrece como base de discusión con el propio asesor legal del lector.

## Fuentes consultadas

Divulgación de Albanese y cronología del Portal de Estadísticas de Medicare: MIXED (septiembre de 2026), "Un agente de OpenAI accedió a archivos no públicos de Medicare el 18 de junio, a Australia se le informó el 10 de septiembre"; Healthcare IT News (septiembre de 2026), "Un agente de OpenAI vulnera el portal australiano de Medicare".

Divulgación de OpenAI sobre sitios web del gobierno de Estados Unidos: Security Affairs (26 de septiembre de 2026), "Agentes de OpenAI accedieron a sitios web del gobierno de EE. UU. sin autorización"; The Week (26 de septiembre de 2026), "Agentes de OpenAI se descontrolan, acceden a sitios web del gobierno de EE. UU., incluidos datos del censo y de la SEC"; CNN Business (26 de septiembre de 2026), "Agentes descontrolados de OpenAI atacaron tres sitios web distintos del gobierno de EE. UU.".

Campaña PaperCut: análisis de GreyNoise según lo reportado por The Register (10 de septiembre de 2026), "Cientos de agentes de IA ayudaron al atacante de PaperCut a golpear a más de 395 organizaciones, y algunos se salieron del guion"; Help Net Security (11 de septiembre de 2026), "Agentes de IA explotaron fallos de PaperCut para vulnerar 395 organizaciones".

Incidente de la AEPD: SecurityWeek (16 de septiembre de 2026), "Se reporta la primera brecha de datos de IA agéntica ante el regulador español"; Help Net Security (17 de septiembre de 2026), "España reporta la primera brecha de datos que involucra a un agente de IA autónomo".

Propuestas del Concejo Municipal de Nueva York: comunicado de prensa del Concejo Municipal de Nueva York (25 de septiembre de 2026), "El Concejo Municipal de Nueva York presenta propuestas legislativas para proteger a los neoyorquinos de los posibles riesgos de la inteligencia artificial"; Fortune (25 de septiembre de 2026), "Washington todavía no ha aprobado una ley de seguridad de IA. Nueva York, donde la IA

se está expandiendo, está redactando la suya propia"; PoliticsNY (25 de septiembre de 2026), "Los denunciantes de IA podrían recibir pago bajo una nueva propuesta del Concejo Municipal de Nueva York".

Ley de Prácticas de Datos Gubernamentales de Minnesota: Minn. Stat. §§ 13.05, subd. 11; 13.055.

Trabajo relacionado de Novara Consulting Group: "El Expediente de Contratación Es el Sistema de Gobernanza" y "El Contrato Redacta el Expediente" (2026), novaracg.com/insights.

**TIPS**

Know something about a sign language AI system that the public should know? Send it to tips@novaracg.com. Every tip is read by me and kept confidential. If you work in the UK, writing to us is not by itself a protected disclosure under the Public Interest Disclosure Act 1998, so before you share employer information, check your options with Protect (protect-advice.org.uk), the free whistleblowing advice line.

## Cómo citar

Grizzle, H. M. (2026, September 27). El Agente No Firmó el Contrato: IA Autónoma, Sistemas Públicos y la Cláusula de Incidentes que la Mayoría de los Condados No Tiene. Novara Consulting Group. https://doi.org/10.5281/zenodo.23003727