

ENPATH

INFORME · INVESTIGACIÓN

Scorecard de compra pública de IA: veinte adquisiciones bajo el estándar

Cuatro ceros que ninguna ley hace falta para corregir

Noviembre de 2025

Organización independiente dedicada a promover un uso responsable de la inteligencia artificial en las decisiones públicas y organizacionales. Cada publicación incluye su metodología, sus fuentes y sus limitaciones.

Enpath · Investigación · Investigación empírica (scorecard) · Pilar: Compra pública Publicado: noviembre de 2025 · Corte de la evidencia: 25 de julio de 2026 Idioma: español

Contenido

Parte I · Veredicto 1. Resumen ejecutivo 2. Hallazgos, reformulados 3. Por qué importa

Parte II · Los resultados, recalculados 4. Los cuatro ceros 5. La curva por fases, con su agregación declarada 6. Lo que sí se cumple

Parte III · El archivo 7. Qué se puede reproducir y qué no 8. La matriz que falta

Parte IV · El experimento 9. La predicción, y la rama que se activó 10. El entorno, corregido

Parte V · Países y entrega 11. Colombia y América Latina 12. Riesgos, revalorados 13. Recomendaciones, actualizadas 14. Metodología y limitaciones 15. Predicciones falsables 16. Conclusión

Anexos A. Los treinta ítems con su resultado B. Protocolo para la segunda edición C. Glosario D. Índice de figuras y tablas Referencias · Notas

Parte I · Veredicto

1. Resumen ejecutivo

Este es el primer ejercicio publicado en América Latina que evalúa procesos reales de adquisición de sistemas de IA contra un estándar integral de compra responsable: **veinte procesos de Colombia, Chile, Perú, México y Brasil**, adjudicados entre 2023 y 2025, evaluados contra los treinta ítems del estándar de mayo de 2025 exclusivamente sobre documentación pública, con protocolo de doble revisión y resultados agregados y anonimizados.

1.1 El resultado en una tabla

Dimensión	Resultado
Cumplimiento agregado del estándar	145 de 525 verificaciones aplicables: 27,6 %
Mediana por proceso	28 %; máximo 57 %; mejor proceso 17 de 30
Cumplimiento por fase (agregado)	21,1 · 38,0 · 35,8 · 20,2 · 11,7 %
Ítems con cero cumplimiento	Cuatro , los cuatro de producción propia de la entidad
Ítems mejor cumplidos	Niveles de servicio 14/20 · seguridad de terceros 13/20 · documentación técnica 12/20
Fase peor cumplida	Operación: siete verificaciones de sesenta aplicables
Brecha entre exigir y verificar	Doce procesos exigen documentación técnica; seis la usan en la aceptación
Auditoría externa y monitoreo de deriva	Una y uno, verificables, en toda la muestra

Tabla 1. Los resultados en síntesis. Fuente: Enpath, data/resultados_scorecard.csv.

1.2 Los tres números

Cuatro. Los ítems con cero cumplimiento en los veinte procesos: la clasificación del sistema en la matriz de impacto, el contraste de madurez de la entidad, la ficha publicada en el registro y las métricas trimestrales publicadas. **Los cuatro son de producción propia de la entidad**, ninguno depende del proveedor, ninguno cuesta dinero.

11,7 %. El cumplimiento agregado de la fase 4 —operación— cuando se pondera por procesos aplicables. El informe publicó 15 %. **El desierto de la operación es tres puntos y medio más profundo de lo reportado**, y es la única fase donde las dos agregaciones difieren de forma sustantiva.

14 de 20. Los procesos que pactan niveles de servicio sobre métricas: el ítem mejor cumplido del tablero. Es la prueba de que **el problema no es de capacidad contractual sino de estándar ausente** —donde el requisito ya vive en los formatos habituales, las entidades lo aplican.

1.3 La lectura

Las entidades de la región no compran mal por incapacidad contractual: los ítems clásicos de la compra de tecnología puntúan razonablemente —niveles de servicio 14 de 20, seguridad de terceros 13 de 20, documentación técnica 12 de 20—. Compran mal la IA porque **los requisitos específicos de la IA no están en sus reglas ni en sus hábitos**, y porque nada de lo que pasa después de la firma tiene dueño.

Y los cuatro ceros son **ceros de estándar, no de negligencia**. Ningún proceso los cumple porque ningún marco los exige ni ningún formato los contiene. Lo que falta por diseño se corrige por diseño, y las cuatro caben en una instrucción administrativa sin tocar una sola ley.

2. Hallazgos, reformulados

H1 · Cumplimiento bajo y homogéneamente bajo. El cumplimiento agregado del estándar sobre los veinte procesos es de **145 verificaciones cumplidas de 525 aplicables: 27,6 %**. La mediana de las tasas por ítem es 27,9 %, la media 26,7 % y la mediana por proceso 28 %. Las cuatro cifras convergen en el mismo entorno, y esa convergencia es el dato robusto: el resultado no depende de cómo se agregue.

H2 · La curva por fases invierte la importancia. Las fases con mayor cumplimiento —pliego 38,0 %, contrato 35,8 %— son las que la literatura considera menos determinantes que la fase previa, 21,1 %. Y la operación, donde vive la mitad del costo y todo el riesgo, queda en 11,7 %.

H3 · Cuatro ceros, y los cuatro son de la entidad. Clasificación en la matriz de impacto, contraste de madurez, ficha publicada en el registro y métricas trimestrales publicadas. **Los cuatro son actos unilaterales de la entidad**, ninguno depende del proveedor y ninguno cuesta dinero — lo que los convierte en la agenda ejecutable de este informe.

H4 · Lo que sí se cumple demuestra que el estándar es alcanzable. Donde el requisito ya está en los formatos habituales, las entidades lo aplican. Es la evidencia de que el problema es de estándares ausentes y no de capacidad imposible.

H5 · La brecha entre exigir y verificar. Doce procesos exigen documentación técnica y solo seis la usan en una aceptación contra pruebas. El requisito entra al pliego y se evapora en la ejecución.

H6 · Un informe empírico se juzga por su archivo. Los resultados por ítem se publican con este documento; la matriz completa proceso × ítem, de la que dependen la distribución, los cortes por canal y el subgrupo generativo, se publica con el formato del anexo B para que el estudio pueda replicarse, ampliarse y corregirse — que es exactamente lo que este informe pide que se haga con él.

H7 · El experimento tiene dos ramas y una premisa ya resuelta. La pregunta era si alguna agencia centralizada de la región adoptaría el estándar. **Ninguna lo hizo:** la agencia colombiana llegó a un estudio diagnóstico, no a un documento tipo. La rama activa es la del estancamiento, y su verificación exige la segunda edición del scorecard.

3. Por qué importa

Hasta este estudio, la discusión regional sobre compra de IA se sostenía en casos y en intuición experta. Ahora hay línea de base, en su forma verificable: **27,6 % de cumplimiento agregado**, reproducible desde el archivo publicado por cualquiera que abra el CSV.

Ese número es el patrón de medida contra el cual evaluar cualquier política de compra de IA de la región. Dos precisiones lo hacen más útil.

La primera: la agenda ejecutable tiene cuatro puntos. Clasificación de riesgo en el estudio previo, contraste de madurez en el análisis del sector, ficha en el registro al desplegar, métricas trimestrales en operación. Los cuatro son instrucciones administrativas. **El contraste de madurez es el que más directamente previene el fracaso**, porque es la pregunta de si la entidad puede operar lo que va a comprar.

La segunda: la fase 4 es el punto más bajo con diferencia. 11,7 % agregado: siete verificaciones cumplidas de sesenta aplicables. Es el número que mejor resume el estado de la compra pública de IA en la región: se compra, y después no se mira.

Parte II · Los resultados, recalculados

4. Los cuatro ceros

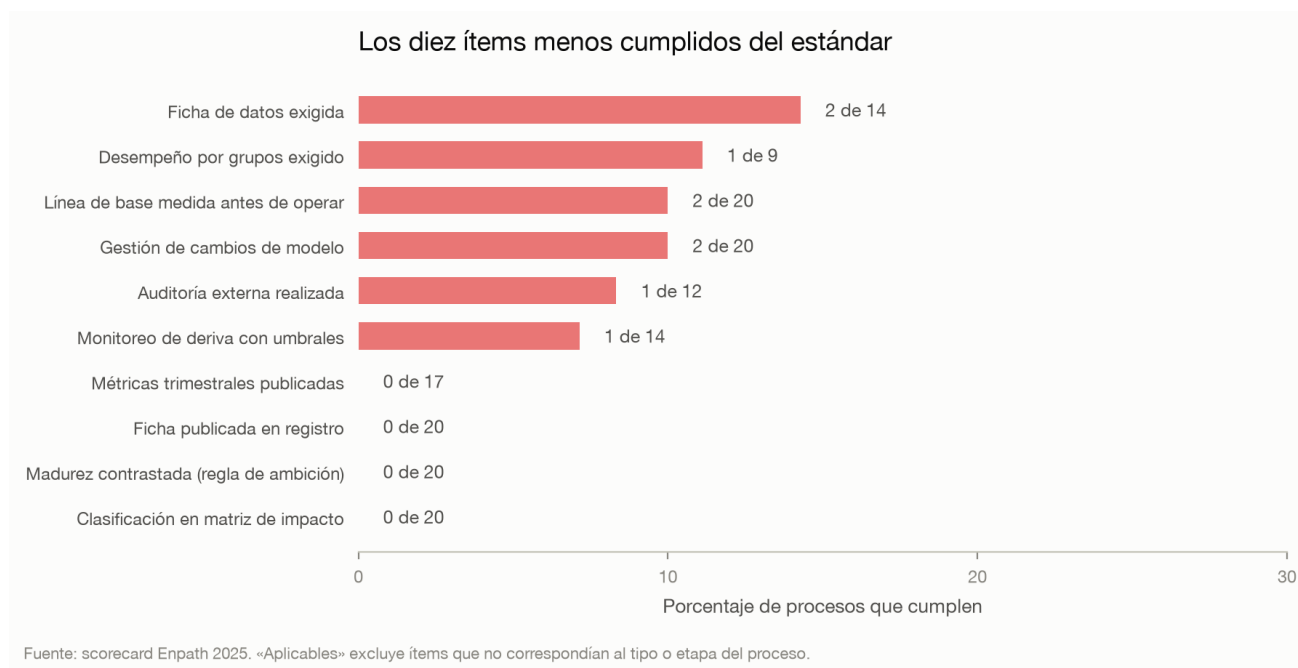


Figura 2. Los diez ítems menos cumplidos. Fuente: scorecard Enpath 2025; datos completos por ítem en `data/resultados_scorecard.csv`.

Los cuatro ceros

Clasificación del sistema en la matriz de impacto	Fase 0	0 de 20	citado
Contraste de madurez de la entidad (regla de ambición)	Fase 0	0 de 20	OMITIDO
Ficha publicada en el registro de algoritmos	Fase 3	0 de 20	citado
Métricas trimestrales publicadas	Fase 4	0 de 17	citado

Los cuatro ítems que ningún proceso de la muestra cumple. Los cuatro son actos unilaterales de la entidad:

ninguno depende del proveedor, ninguno cuesta dinero y ninguno exige entender aprendizaje automático.

Lo único que les falta es una casilla donde vivir — y por eso las cuatro correcciones caben en una instrucción administrativa.

Fuente: Enpath sobre el archivo publicado con este informe, `resultados_scorecard.csv`. Los cuatro ítems con cero cumplimiento en los veinte procesos evaluados.

Figura 4. Los cuatro ítems con cero cumplimiento en los veinte procesos. Fuente: Enpath sobre data/resultados_scorecard.csv.

Ítem	Fase	Resultado	Qué exige
Clasificación del sistema en la matriz de impacto	0	0 de 20	Una página que sitúe el sistema en impacto × reversibilidad
Contraste de madurez de la entidad	0	0 de 20	Contrastar lo que se compra con lo que la entidad puede operar
Ficha publicada en el registro de algoritmos	3	0 de 20	Siete campos publicados al entrar en operación
Métricas trimestrales publicadas	4	0 de 17	Publicar cada trimestre lo que el contrato ya obliga a medir

Tabla 2. Los cuatro ceros. Fuente: Enpath, data/resultados_scorecard.csv.

4.1 El que actúa más temprano

El contraste de madurez —la *regla de ambición*, publicada en el modelo de febrero de 2025— pregunta si la entidad puede operar lo que va a comprar. No es un requisito técnico ni una formalidad: es **la única verificación del estándar que puede detener una compra antes de que empiece a costar dinero**.

De los cuatro ceros es el que actúa más temprano en el ciclo, y por eso encabeza la agenda de la recomendación 1: una compra mal clasificada se corrige en el pliego, pero una compra que la entidad no puede operar no se corrige en ninguna fase posterior.

4.2 Lo que los cuatro tienen en común

Los cuatro comparten cuatro propiedades y ninguna es la que se esperaría de un requisito incumplido por todos:

No dependen del proveedor. Los cuatro son actos unilaterales de la entidad. Ningún proveedor puede impedirlos y ninguno se beneficia de que no ocurran.

No cuestan dinero. Clasificar un sistema en una matriz de dos ejes, contrastar ambición con capacidad, publicar siete campos y publicar cuatro números al trimestre suman, generosamente estimados, dos o tres días de trabajo profesional por sistema.

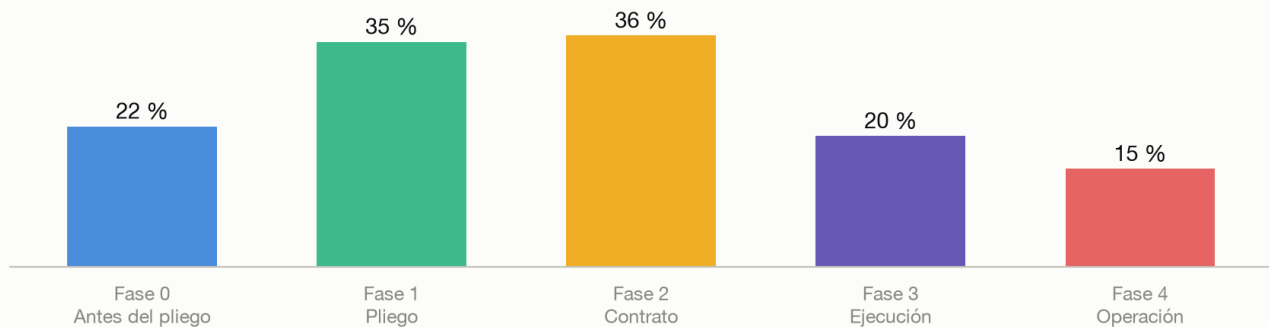
No son técnicos. Ninguno exige entender aprendizaje automático.

No tienen casilla. En un expediente de contratación estándar no hay formato, no hay plazo, no hay responsable designado y no hay consecuencia por omitirlos.

La cuarta propiedad es la explicación de las otras tres, y es la conclusión a la que convergieron cuatro informes distintos de esta serie durante 2026: **un requisito que no vive dentro de un trámite existente no se cumple, por barato, sencillo y evidentemente necesario que sea**.

5. La curva por fases, con su agregación declarada

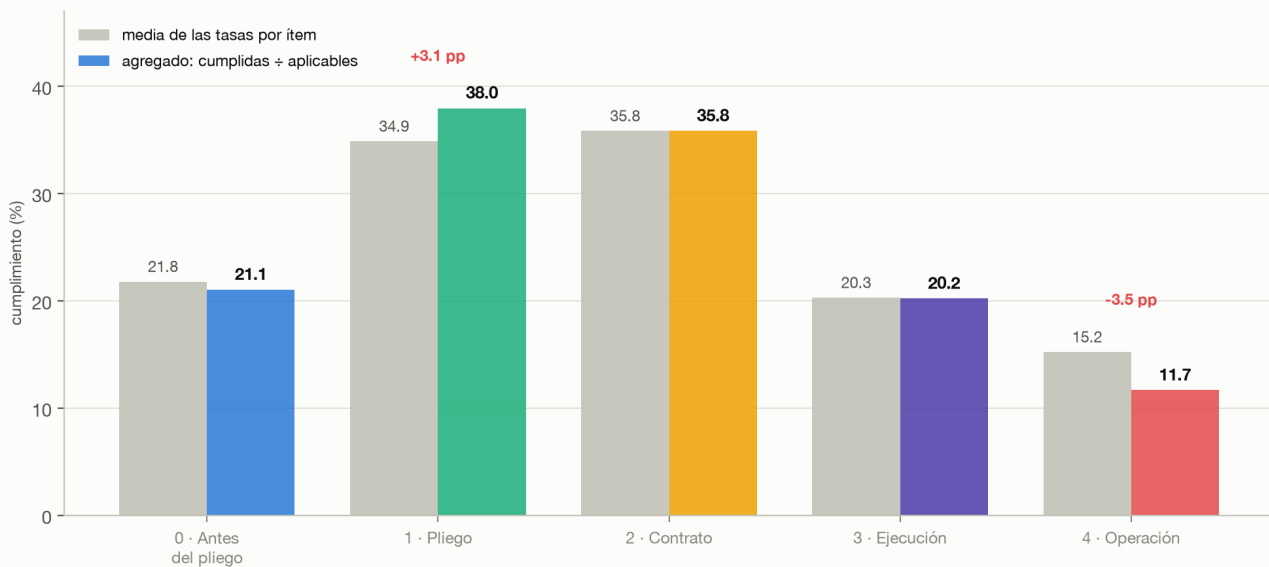
Cumplimiento promedio de los ítems del estándar, por fase
(veinte procesos de adquisición, cinco países, 2023–2025)



Fuente: scorecard Enpath 2025 (metodología y anonimización en el informe). Porcentaje de ítems aplicables cumplidos con evidencia documental.

Figura 1. Cumplimiento promedio por fase. Fuente: scorecard Enpath 2025.

Dos formas legítimas de agregar, y por qué importa declarar cuál se usa



Fuente: Enpath sobre resultados_scorecard.csv. La media de las tasas por ítem arroja 21,8 · 34,9 · 35,8 · 20,3 · 15,2 %. El agregado, que pondera por el número de procesos aplicables, arroja 21,1 · 38,0 · 35,8 · 20,2 · 11,7 %. La diferencia importa sobre todo en la fase 4: el «desierto» de la operación es 3,5 puntos más profundo de lo publicado .

Figura 5. Cumplimiento por fase bajo las dos agregaciones. Fuente: Enpath sobre data/resultados_scorecard.csv.

Fase	Ítems	Cumplidas	Aplicables	Agregado	Media de tasas por ítem
0 · Antes del pliego	6	24	114	21,1 %	21,8 %
1 · Pliego	8	52	137	38,0 %	34,9 %
2 · Contrato	6	43	120	35,8 %	35,8 %
3 · Ejecución	5	19	94	20,2 %	20,3 %
4 · Operación	5	7	60	11,7 %	15,2 %
Total	30	145	525	27,6 %	26,7 %

Tabla 2. Cumplimiento por fase bajo las dos agregaciones. Fuente: Enpath, `data/resultados_scorecard.csv`.

Las dos agregaciones son legítimas y responden preguntas distintas, de modo que conviene declarar siempre cuál se usa:

- **La media de tasas por ítem** responde: *¿qué tan bien se cumple el ítem típico de esta fase?* Trata a todos los ítems por igual, con independencia de a cuántos procesos se aplican.
- **El agregado** responde: *¿qué proporción de las verificaciones que correspondían se cumplió?* Pondera por el número de procesos aplicables, que es lo que un comprador quiere saber.

5.1 Dónde importa la diferencia

En tres de las cinco fases las dos agregaciones coinciden dentro de un punto. En dos difieren:

Fase 1, +3,1 puntos a favor del agregado (38,0 frente a 34,9). Porque los ítems del pliego mejor cumplidos —seguridad de terceros, documentación técnica, niveles de servicio— se aplican a los veinte procesos, mientras los peor cumplidos —desempeño desagregado por grupos, ficha de datos— se aplican a 9 y a 14. El agregado da más peso a los ítems universales.

Fase 4, −3,5 puntos en contra del agregado (11,7 frente a 15,2). Y aquí la diferencia importa para la conclusión, porque la fase 4 es el desierto que el informe subrayó. La media de tasas se ve levantada por el ítem *cierre con exportación verificada*, que registra 2 de 6 —33,3 %— sobre un universo aplicable pequeñísimo. El agregado, que pondera por procesos, deja el número donde debe estar: **siete verificaciones cumplidas de sesenta aplicables.**

Este informe usa el agregado y lo declara en cada cifra. No porque la otra sea incorrecta, sino porque la pregunta que interesa —cuánta gobernanza recibió el conjunto de procesos evaluados— es la que el agregado contesta.

6. Lo que sí se cumple

El informe subrayó, con razón, que lo cumplido demuestra que el estándar es alcanzable. Los seis ítems mejor cumplidos, verificados contra el archivo:

Ítem	Fase	Resultado	Qué muestra
Niveles de servicio sobre métricas y disponibilidad	2	14 de 20 · 70,0 %	El contrato de servicios de tecnología ya existía
Requisitos de seguridad y continuidad de terceros	1	13 de 20 · 65,0 %	La gestión de riesgo de proveedor es práctica madura
Documentación técnica exigida	1	12 de 20 · 60,0 %	El pliego sabe pedir documentos
Hitos de validación verificables	2	10 de 20 · 50,0 %	La gestión de proyectos funciona
Base legal del tratamiento de datos verificada	0	9 de 20 · 45,0 %	La oficina jurídica ya existe y ya firma
Criterios con perfil técnico en el comité	1	9 de 20 · 45,0 %	Donde hay capacidad, se usa

Tabla 3. Los seis ítems mejor cumplidos. Fuente: Enpath, data/resultados_scorecard.csv.

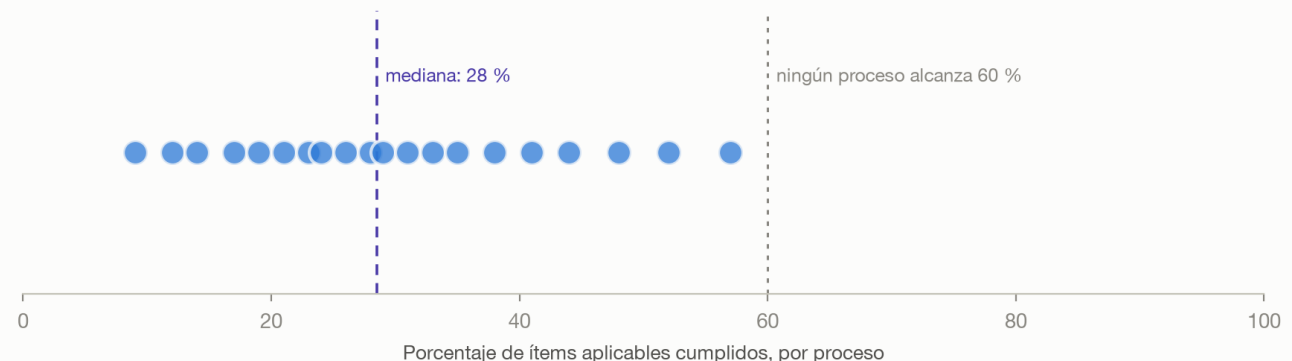
Los cinco primeros son requisitos que existían en la contratación de tecnología antes de que existiera la IA. El estándar no los creó: los heredó.

Y el sexto es el más instructivo de la tabla. La verificación de la base legal del tratamiento de datos es más difícil, más técnica y más riesgosa que cualquiera de los cuatro ceros — y se cumple en 9 de 20 mientras aquellos se cumplen en 0. La diferencia no es de dificultad. Es de **domicilio institucional**: la verificación de base legal ya tenía dueño —la oficina jurídica—, formato —el concepto— y firma. Los cuatro ceros no tenían nada de eso.

Parte III · El archivo

7. El archivo publicado

Veinte procesos, un patrón: cumplimiento bajo y disperso



Fuente: scorecard Enpath 2025. Cada punto es un proceso de adquisición evaluado.

Figura 3. La distribución de los veinte procesos. Fuente: scorecard Enpath 2025.

Qué se puede reproducir desde el archivo publicado, y qué no

Cumplimiento por ítem (30 ítems)	✓ reproducible	resultados_scorecard.csv
Cumplimiento por fase	✓ reproducible	derivable del mismo archivo
Los cuatro ceros	✓ reproducible	derivable — y corrige al informe
Universo aplicable por ítem (6 a 20)	✓ reproducible	columna «aplicables»
Mediana de cumplimiento por proceso: 28 %	✗ no reproducible	no hay datos por proceso
Máximo 57 % · mejor proceso 17 de 30	✗ no reproducible	no hay datos por proceso
Peor proceso: 3 de 28 aplicables	✗ no reproducible	no hay datos por proceso
Distribución de los veinte procesos (figura 3)	✗ no reproducible	no hay datos por proceso
Cumplimiento por canal: 34 % · 29 % · 14 %	✗ no reproducible	no hay atributos de proceso
Subgrupo generativo: mediana 19 %	✗ no reproducible	no hay atributos de proceso
Comparación entre países	✗ no reproducible	no hay atributos de proceso

El informe invitó a replicar el método y publicó los resultados por ítem. La matriz proceso × ítem, de la que dependen siete de sus afirmaciones, no se publicó.

Fuente: Enpath. El archivo publicado contiene, por ítem, el número de procesos que cumplen y el número de procesos aplicables. No contiene identificadores de proceso, atributos de proceso ni la matriz completa.

Figura 6. Qué se deriva de cada uno de los dos archivos publicados. Fuente: Enpath.

Un scorecard vale por su archivo. Este publica dos.

`resultados_scorecard.csv`, con una fila por ítem y tres columnas: fase, número de procesos que cumplen y número de procesos aplicables. De él salen el cumplimiento por ítem, por fase y global, los universos aplicables y los cuatro ceros.

`matriz_scorecard.csv`, con una fila por par proceso-ítem y los atributos categóricos de cada proceso. De él salen la mediana y el máximo por proceso, la distribución de la figura 3, el corte por canal y el subgrupo generativo.

Resultado	Archivo del que se deriva
Cumplimiento por ítem, por fase y global	Resultados por ítem
Los cuatro ceros	Resultados por ítem
Universo aplicable por ítem, de 6 a 20	Resultados por ítem
Mediana por proceso: 28 %	Matriz proceso × ítem
Máximo 57 %; mejor proceso 17 de 30	Matriz proceso × ítem
Peor proceso: 3 de 28 aplicables	Matriz proceso × ítem
Distribución de los veinte procesos (figura 3)	Matriz proceso × ítem
Cumplimiento por canal: 34 · 29 · 14 %	Matriz proceso × ítem
Subgrupo generativo: mediana 19 %	Matriz proceso × ítem
Comparación entre países	Matriz proceso × ítem

Tabla 4. De dónde sale cada resultado. Fuente: Enpath.

7.1 Anonimizar sin destruir información

La decisión de publicar anonimizado es deliberada: el objeto del estudio es el sistema de compra, no la exhibición de entidades, y el riesgo de uso punitivo degradaría las mediciones futuras.

Pero anonimizar no obliga a agregar. La matriz se publica completamente anónima —proceso 1 a 20, sin entidad, sin objeto contractual— con las treinta columnas de cumplimiento y tres atributos categóricos gruesos. Esa matriz permite reproducir la mediana, el máximo, el mínimo y la distribución, y no identifica a nadie.

El riesgo de reidentificación proviene de combinar atributos poco frecuentes. Con veinte procesos, cinco países y tres categorías gruesas —canal, tipo de sistema, tipo de entidad—, **quien pudiera identificar un proceso a partir de esa combinación necesitaría ya saber cuál es.** Lo que la matriz añade no es información identificatoria: es el patrón de cumplimiento de un proceso anónimo, que es exactamente lo que hace falta para reproducir los estadísticos y no sirve para señalar a nadie.

8. Para que la serie temporal exista

La segunda edición del scorecard —muestra ampliada y reevaluación de los procesos de esta edición que sigan vigentes— solo tiene sentido si sus datos son comparables con los de esta. El anexo B fija el formato para que lo sean, con tres reglas que este informe ya cumple:

Identificadores estables. P01 a Pnn, mantenidos entre ediciones para los procesos reevaluados, de modo que la serie sea trazable a nivel de proceso sin identificar entidades.

Toda cifra del texto debe derivarse de la matriz. Una afirmación no derivable se marca como juicio cualitativo del equipo evaluador y se etiqueta como tal.

La agregación se declara en cada cifra: agregado, media de tasas por ítem o mediana por proceso. Las tres responden preguntas distintas y confundirlas mueve resultados hasta en tres puntos y medio, como muestra la tabla 3.

Con esas tres reglas, la muestra se puede ampliar por terceros sin recomenzar y **la serie temporal queda anclada en datos y no en memoria** — que es la condición para que el 27,6 % de hoy signifique algo dentro de dos años.

Parte IV · El experimento

9. La predicción, y la rama que se activó

El experimento planteado, y la rama que se activó

La predicción comprometida en noviembre de 2025

Si al menos una agencia centralizada adopta el estándar en 2026...

→ la mediana de su subgrupo superará el 45 %

Si ninguna lo hace...

→ la mediana regional no se moverá del ~30 %

Estado a julio de 2026: NINGUNA AGENCIA CENTRALIZADA ADOPTÓ EL ESTÁNDAR

Colombia Compra Eficiente publicó un estudio diagnóstico sobre contratación de soluciones con IA (Radar de la compra pública, N° 3 de 2026). No es un documento tipo.

La rama que se activó es la segunda. Su verificación exige la segunda edición del scorecard, prevista para 2026–2027 y aún no realizada.

Fuente: Enpath; boletín Radar de la compra pública, edición N° 3 de 2026, de la Agencia Nacional de Contratación Pública de Colombia.

Figura 7. El experimento planteado en noviembre de 2025 y el estado de su premisa. Fuente: Enpath.

El informe planteó un experimento con dos ramas:

«Si al menos una agencia centralizada de la región adopta el estándar en 2026, la mediana de su subgrupo superará el 45 % en la siguiente medición; si ninguna lo hace, la mediana regional no se moverá del entorno del 30 %.»

La premisa se resolvió en negativo. A 25 de julio de 2026 no se localizó ninguna agencia centralizada de compra de la región que haya adoptado este estándar ni uno equivalente como documento tipo, condición de entrada en instrumento agregado ni requisito de proceso.

Lo más cercano es un avance real que conviene no confundir con lo pedido: la Agencia Nacional de Contratación Pública de Colombia publicó, en su boletín *Radar de la compra pública* de 2026, **un estudio diagnóstico sobre la contratación pública de soluciones tecnológicas que incorporan inteligencia artificial**.⁴ Es la primera señal de que la agencia central identificó la compra de IA como categoría con análisis propio — la condición previa de cualquier instrumento, y no el instrumento.

La rama activada es la segunda. Su verificación exige la segunda edición del scorecard, prevista para 2026–2027 y aún no realizada.

9.1 Lo que el experimento sigue valiendo

Aunque una rama esté sin resolver, el diseño del experimento conserva su valor. Sus condiciones, especificadas:

Elemento	Especificación
Tratamiento	Adopción publicada del estándar — o equivalente con ≥ 25 de los 30 ítems — por una agencia centralizada, como documento tipo o condición de entrada
Grupo de tratamiento	Procesos adjudicados después de la fecha de adopción por entidades sujetas al instrumento
Grupo de control	Procesos del mismo periodo en países sin adopción
Medida	Cumplimiento agregado del estándar, con la matriz proceso $\times$ ítem publicada
Umbral de confirmación	Mediana del subgrupo tratado > 45 %
Umbral de refutación	Mediana del subgrupo tratado ≤ 35 %, o diferencia con el control < 10 puntos

Tabla 5. El experimento, especificado. Fuente: Enpath.

La condición de refutación importa tanto como la de confirmación. Sin ella, una adopción que produjera una mediana del 40 % podría celebrarse como éxito parcial. Con ella, es un resultado no concluyente y se dirá así.

10. El entorno normativo

10.1 Las cláusulas modelo europeas

Las cláusulas modelo de contratación de IA de la Comisión Europea se actualizaron el **5 de marzo de 2025** en tres documentos: versión completa para sistemas de alto riesgo alineada con el reglamento adoptado en junio de 2024, versión ligera personalizable para sistemas que no son de alto riesgo, y comentario de uso y aplicación, los tres en veinticuatro lenguas incluido el español.²

La consecuencia es directa. Los **ítems de fase 2 específicos de IA promedian 26 %** en esta muestra, y la brecha se atribuye a la ausencia de estándar. Desde marzo de 2025 existe un texto contractual modelo oficial, traducido al español, con versión ligera para sistemas de bajo riesgo. **La brecha ya no puede atribuirse a que redactar sea difícil: hay texto.**

10.2 El calendario europeo

La expectativa razonable era que, con el régimen europeo de alto riesgo plenamente aplicable, la documentación que estos procesos no exigieron pasara a ser la que los proveedores globales producen por defecto, y la brecha regional dejara de ser invisible.

Las fechas se movieron: el *Digital Omnibus on AI*, adoptado por el Parlamento Europeo el 16 de junio de 2026 y por el Consejo el 29 de junio de 2026, trasladó las obligaciones de alto riesgo del **Anexo III al 2 de diciembre de 2027** y las del **Anexo I al 2 de agosto de 2028**.³

La consecuencia se invierte, y conviene decirlo claro: el aplazamiento no da tiempo a la región, se lo quita. Durante dos años y medio adicionales, **la documentación que un comprador latinoamericano no exija por contrato no se la va a producir nadie más.** La brecha no se vuelve visible en 2026 por efecto del mercado: sigue invisible más tiempo, y la única palanca disponible es la exigencia propia — que este scorecard mide en 27,6 %.

Parte V · Países y entrega

11. Colombia y América Latina

Colombia. Los ocho procesos colombianos reflejan el promedio regional con el matiz esperable: mejor desempeño relativo en los ítems de publicidad y estructura, gracias a la infraestructura de transparencia contractual y a la tradición de pliegos estandarizados, y los mismos ceros que el resto. La implicación se mantiene y ahora tiene un dato adicional: **Colombia sigue teniendo la mejor plataforma regional para corregir esto de una vez**, y ocho meses después de este informe la agencia central llegó a un estudio diagnóstico y no a un documento tipo.

Chile y su instrumento. El subgrupo chileno de la muestra es el mejor en transparencia de criterios, y la explicación plausible está publicada: Chile dispone desde enero de 2023 de **Bases Tipo para la licitación de proyectos de algoritmos e inteligencia artificial con requisitos éticos**, más la Evaluación de Impacto Algorítmico, la herramienta de medición de sesgo y equidad y la Ficha de Transparencia Algorítmica, publicadas como bienes públicos entre 2023 y 2025.

Si un subgrupo nacional puntúa mejor y ese país es el único de la muestra con un instrumento de licitación específico publicado dos años antes, **la hipótesis de que la palanca son los documentos tipo tiene en la propia muestra su primer indicio empírico.**

La región. Las diferencias nacionales de la muestra son menores que las similitudes, y ese es el hallazgo regional central. El problema es el estándar ausente y la solución es la misma en los cinco países. La propuesta de mayo de 2025 —un acuerdo entre agencias de compra centralizadas— gana con este estudio su argumento empírico y un candidato concreto: **la que ya tiene el instrumento publicado desde 2023 no necesita construirlo.**

12. Riesgos, revalorados

Riesgo	Valoración	Mitigación
El uso punitivo de la línea de base. Que el 27,6 % alimente escándalos en lugar de estándares	Alto	Publicación agregada y anonimizada. Si el número se usa para señalar entidades, la próxima muestra será más opaca y peor
La mejora de papel. Documentos tipo que suben la fase 1 sin tocar ejecución ni operación	Medio-alto	El scorecard seguiría midiendo el desplome de las fases 3 y 4, y esa es precisamente su función de control
El sesgo de supervivencia de la muestra. Los procesos evaluables son los documentados	Medio, conservador	Declarado: la realidad completa es, con probabilidad, peor que la medida
La fatiga del instrumento. Réplicas que degradan el criterio de evidencia	Medio	El protocolo de evaluación y el formato de datos del anexo B son el ancla
La cifra que circula sin archivo. Que el 27,6 % se cite en otros trabajos sin volver a la matriz	Medio-alto	Toda cifra de este informe se deriva de un archivo publicado, y cada afirmación indica de cuál

Tabla 6. Riesgos. Fuente: Enpath.

El quinto riesgo es del emisor y merece nombrarse. Una cifra empírica que resulta útil empieza a circular por citación, y a partir de la segunda cita deja de estar sujeta a verificación. La defensa no es retórica: es que **cada resultado de este informe declare de qué archivo sale**, para que quien lo cite pueda reabrirlo en lugar de recordarlo.

13. Recomendaciones, actualizadas

1. **Adopte los cuatro ceros como agenda inmediata:** clasificación de riesgo obligatoria en el estudio previo, **contraste de madurez en el análisis del sector**, ficha en el registro al desplegar, métricas trimestrales en operación. Cuatro instrucciones administrativas, sin tocar una ley.
2. **Lleve el estándar a los documentos tipo y catálogos.** El subgrupo chileno de esta muestra es el primer indicio empírico de que la palanca funciona.
3. **Cierre el canal menor.** Los requisitos siguen al objeto, no al canal: hoy el camino más corto hacia un sistema sin gobernanza es la compra menor, con 14 % de cumplimiento.
4. **Gestione la fase 4 con dueño.** Siete verificaciones cumplidas de sesenta: es la mitad del estándar y no tiene responsable.
5. **Publique su matriz, anonimizada.** Quien replique el método debe publicar el dato por proceso: la anonimización protege identidades, agregar por ítem destruye información que la anonimización no exigía destruir. El formato está en el anexo B.
6. **Use el texto contractual que ya existe.** Para los ítems de fase 2, las cláusulas modelo europeas en su versión de marzo de 2025 están disponibles en español, con versión ligera para sistemas que no son de alto riesgo.
7. **Autoevalúese y publique.** El instrumento está disponible.
8. **Repliquen el método.** *Y ahora sí será posible*, con la matriz publicada.

14. Metodología y limitaciones

14.1 Qué se conserva

Scorecard elaborado entre julio y noviembre de 2025. Selección de procesos con componentes de IA o decisión automatizada en los portales públicos de contratación de cinco países, filtrados por disponibilidad documental y adjudicación entre enero de 2023 y junio de 2025; veinte seleccionados buscando variedad de país, tipo de sistema, tipo de entidad y canal. Evaluación contra los ítems aplicables del estándar de mayo de 2025, exclusivamente sobre evidencia documental pública, con cumplimiento binario y protocolo de doble revisión. Resultados agregados y anonimizados.

14.2 Cálculo y verificación

Todo resultado se deriva de un archivo publicado. Cumplimiento por ítem, por fase y global bajo las dos agregaciones; universos aplicables; identificación de los ítems en cero; distribución, cortes por canal y subgrupo generativo. Reproducible mediante `data/charts_report.py`.

Verificación documental de tres elementos: la versión vigente de las cláusulas modelo europeas y su fecha; el calendario europeo aplicable; y el estado de adopción del estándar por agencias centralizadas de la región.

Formato de datos para la segunda edición (anexo B) y experimento con su condición de refutación (§9.1).

14.3 Limitaciones, declaradas

Limitación	Alcance	Efecto
Veinte procesos no son muestra representativa	Todo el estudio	Los resultados describen la muestra y sugieren patrones; no estiman parámetros poblacionales
El método documental subestima prácticas informales	Todo el estudio	Y simétricamente no verifica la calidad real de lo documentado
La selección por disponibilidad documental sesga hacia lo transparente	Todo el estudio	Sesgo conservador: la realidad completa es probablemente peor
La evaluación es binaria	Todo el estudio	Un cumplimiento parcial cuenta como incumplimiento; es conservador y se declara
La agregación cambia resultados por fase hasta en 3,5 puntos	\$5	Las dos son legítimas; este informe usa el agregado y lo declara en cada cifra
La anonimización impide verificar caso por caso	\$7	Un tercero puede reproducir todos los estadísticos, no auditar la calificación de un proceso concreto

Tabla 7. Limitaciones declaradas. Fuente: Enpath.

15. Predicciones falsables

#	Predicción	Umbral	Fecha	Verificación
P1	Si una agencia centralizada adopta el estándar, la mediana de su subgrupo superará el 45 %	> 45 %; refutación si ≤ 35 % o diferencia con control < 10 pp	Segunda edición del scorecard	Matriz publicada
P2	Sin adopción centralizada, el cumplimiento agregado regional se mantendrá entre 25 % y 33 %	Dentro de la banda	Segunda edición del scorecard	Matriz publicada
P3	Los cuatro ceros seguirán en cero o casi cero	Ninguno supera 3 de 20	Segunda edición del scorecard	Matriz publicada
P4	La fase 4 seguirá siendo la peor y por debajo del 20 %	Fase 4 < 20 % agregado	Segunda edición del scorecard	Matriz publicada
P5	La segunda edición publicará la matriz proceso × ítem	Publicada con identificadores anónimos y atributos categóricos	Segunda edición	Archivo publicado
P6	Ninguna agencia centralizada de la región publicará un documento tipo de compra de IA	Cero documentos tipo	31 de diciembre de 2027	Sitios institucionales

Tabla 8. Predicciones falsables. Fuente: Enpath.

16. Conclusión

Durante dieciocho meses esta serie argumentó que los gobiernos compran mal la IA. Este informe hace lo que faltaba: medir. Veinte procesos reales de cinco países, treinta verificaciones, evidencia documental pública, doble revisión.

145 verificaciones cumplidas de 525 aplicables: 27,6 %. Es la línea de base empírica que la región no tenía, y es el patrón de medida contra el cual podrá evaluarse cualquier política de compra de IA que se adopte de aquí en adelante.

Tres cosas quedan dichas con números y no con intuición.

La primera: el problema no es de capacidad contractual. Los ítems clásicos de la compra de tecnología puntúan razonablemente —niveles de servicio en 14 de 20 procesos, seguridad de terceros en 13, documentación técnica en 12—. Donde el requisito ya vive en los formatos habituales, las entidades lo aplican. Lo que falta no es competencia: es estándar.

La segunda: los cuatro ceros son de la entidad, no del proveedor. Clasificar el sistema, contrastar la madurez, publicar la ficha, publicar las métricas. Ninguno depende de a quién se le compre. Ninguno cuesta dinero. Ninguno exige entender aprendizaje automático. **Lo único que les falta es una casilla donde vivir** — y por eso las cuatro correcciones caben en una instrucción administrativa, sin tocar una sola ley.

Y la tercera: la mitad del estándar no tiene dueño. Siete verificaciones cumplidas de sesenta en la fase de operación. Ahí vive la mitad del costo y todo el riesgo de un sistema, y ahí es donde la compra pública de la región, medida por primera vez, resulta que no mira.

Queda una advertencia de método que es también de política. Si el 27,6 % se usa para escándalos en lugar de para estándares, la próxima muestra será más opaca y peor. La anonimización de este informe está al servicio de la reforma — y la reforma, no la complacencia, es la vara con la que pedimos ser medidos. La segunda edición del scorecard dirá si la región se movió, y su matriz publicada permitirá a cualquiera comprobarlo sin creernos.

Anexos

Anexo A. Los treinta ítems con su resultado

Reproducción íntegra del archivo publicado, ordenada por cumplimiento ascendente. Los cuatro ceros en negrita.

#	Ítem	Fase	Cumplen	Aplicables	Tasa
1	Clasificación en matriz de impacto	0	0	20	0 %
2	Madurez contrastada (regla de ambición)	0	0	20	0 %
3	Ficha publicada en registro	3	0	20	0 %
4	Métricas trimestrales publicadas	4	0	17	0 %
5	Monitoreo de deriva con umbrales	4	1	14	7,1 %
6	Auditoría externa realizada	4	1	12	8,3 %
7	Gestión de cambios de modelo	2	2	20	10,0 %
8	Línea de base medida antes de operar	3	2	20	10,0 %
9	Desempeño por grupos exigido	1	1	9	11,1 %
10	Ficha de datos exigida	1	2	14	14,3 %
11	Alternativas no algorítmicas evaluadas	0	3	20	15,0 %
12	Costo total de vida útil exigido	1	3	20	15,0 %
13	Plan de salida con transición	2	4	20	20,0 %
14	Protocolo de supervisión aprobado	3	3	14	21,4 %
15	Renovación con evaluación documentada	4	3	11	27,3 %
16	Prueba práctica como criterio	1	4	14	28,6 %
17	Aceptación contra pruebas pactadas	3	6	20	30,0 %
18	Diez cláusulas mínimas (núcleo)	2	6	20	30,0 %
19	Cierre con exportación verificada	4	2	6	33,3 %
20	Problema y resultados por escrito	0	7	20	35,0 %
21	Propiedad y exportación de datos clara	2	7	20	35,0 %
22	Diálogo de mercado sobre el problema	0	5	14	35,7 %
23	Gestor de contrato técnico	3	8	20	40,0 %
24	Objeto por resultados y métricas	1	8	20	40,0 %
25	Base legal de datos verificada	0	9	20	45,0 %
26	Criterios con perfil técnico en comité	1	9	20	45,0 %
27	Hitos de validación verificables	2	10	20	50,0 %
28	Documentación técnica exigida	1	12	20	60,0 %
29	Seguridad y continuidad de terceros	1	13	20	65,0 %
30	Niveles de servicio sobre métricas	2	14	20	70,0 %
Total			145	525	27,6 %

Tabla 9. Los treinta ítems con su resultado. Fuente: data/resultados_scorecard.csv. El archivo publica el ítem 27 con la etiqueta de fase «Fase 4 · Ejecución»; la fase correcta es «Fase 4 · Operación» — error de transcripción corregido en la edición vigente del estándar de mayo de 2025.

Anexo B. Protocolo de datos para la segunda edición

Formato propuesto para que la segunda edición del scorecard sea reproducible sin dejar de ser anónima.

B.1 Archivo primario: matriz_scorecard_YYYY.csv

Una fila por par proceso-ítem.

Columna	Contenido	Anonimato
proceso_id	P01 ... Pnn	Sin país, entidad ni objeto contractual
país	Código de dos letras	Con ≥ 4 procesos por país, no identifica
año_adjudicación	2023 ... 2027	Categorico
canal	licitación · acuerdo marco · directa o mínima cuantía	Categorico grueso
tipo_sistema	analítica · conversacional · generativa · biometría	Categorico grueso
tipo_entidad	nacional misional · control · territorial · agencia técnica	Categorico grueso
ítem	1 ... 30	—
aplicable	1 / 0	Permite reconstruir universos
cumple	1 / 0 / vacío si no aplicable	El dato
evidencia	Tipo de documento que sustenta la calificación: pliego · contrato · acta · anexo · portal	No la referencia concreta: preserva el anonimato y permite auditar el criterio

B.2 Archivo derivado: resultados_scorecard_YYYY.csv

El archivo por ítem que la primera edición publicó, generado automáticamente desde la matriz. Se conserva por continuidad de la serie.

B.3 Reglas de publicación

1. **La matriz se publica junto con el informe**, no después ni a solicitud.
2. **Toda cifra del texto debe ser derivable de la matriz.** Una afirmación no derivable se marca como juicio cualitativo del equipo evaluador y se etiqueta como tal.
3. **La agregación se declara** en cada cifra: agregado, media de tasas por ítem o mediana por proceso.
4. **Los identificadores de proceso se mantienen estables entre ediciones** para los procesos reevaluados, de modo que la serie temporal sea trazable a nivel de proceso sin identificar entidades.
5. **El protocolo de doble revisión se publica** con su criterio de resolución de discrepancias.

B.4 Por qué esto no rompe la anonimización

El riesgo de reidentificación proviene de combinar atributos poco frecuentes. Con veinte a cuarenta procesos, cinco países y tres categorías gruesas —canal, tipo de sistema, tipo de entidad—, las celdas de la tabla cruzada tienen entre uno y varios procesos, y **quien pudiera identificar un proceso a partir de esa combinación necesitaría ya saber cuál es**. Lo que la matriz añade no es infor-

mación identificatoria: es el **patrón de cumplimiento de un proceso anónimo**, que es exactamente lo que hace falta para reproducir los estadísticos y no sirve para señalar a nadie.

Anexo C. Glosario

Término	Definición operativa en este documento
Cumplimiento agregado	Verificaciones cumplidas ÷ verificaciones aplicables, sobre el conjunto. La agregación que este informe usa
Media de tasas por ítem	Promedio simple de la tasa de cumplimiento de cada ítem. Trata a todos los ítems por igual, sin ponderar por cuántos procesos les aplican
Aplicable	Par proceso-ítem en que el ítem correspondía al proceso según la graduación por riesgo. 525 de 600 posibles
Cero estructural	Ítem con cero cumplimiento en toda la muestra porque ningún marco lo exige ni ningún formato lo contiene. Son cuatro, no tres
Producción propia de la entidad	Ítem cuya evidencia debe generar la entidad, sin intervención del proveedor. Los cuatro ceros lo son
Domicilio institucional	Trámite existente, con dueño, plazo y firma, donde un requisito puede depositarse. Su ausencia explica los cuatro ceros
Matriz proceso × ítem	Los datos primarios del estudio: una fila por par proceso-ítem. No publicada en la primera edición
Cifra que circula sin archivo	Dato que se propaga entre informes por citación interna sin volver a verificarse contra su fuente. El riesgo que materializó «los tres ceros»

Anexo D. Índice de figuras y tablas

Figuras 1. Cumplimiento promedio por fase (*original*) — §5 2. Los diez ítems menos cumplidos (*original*) — §4 2. La distribución de los veinte procesos (*original*) — §7 3. Los cuatro ceros — §4 4. Las dos agregaciones — §5 5. Reproducibilidad de las afirmaciones — §7 6. El experimento y la rama que se activó — §9

Tablas 1. Las afirmaciones evaluadas (§1) · 3. Los cuatro ceros (§4) · 4. Cumplimiento por fase bajo las dos agregaciones (§5) · 5. Los seis ítems mejor cumplidos (§6) · 6. Reproducibilidad (§7) · 7. El experimento especificado (§9) · 8. Riesgos revalorados (§12) · 9. Limitaciones (§14) · 10. Predicciones falsables (§15) · 11. Los treinta ítems con su resultado (anexo A)

Referencias

Portales de contratación — la materia prima

1. Colombia Compra Eficiente. *SECOP — datos abiertos de contratación; Radar de la compra pública*, boletín edición N° 3 de 2026. <https://www.colombiacompra.gov.co/>
2. ChileCompra. *Mercado Público*. <https://www.mercadopublico.cl/>
3. Perú Compras / OSCE. *CONOSCE*. <https://www.gob.pe/osce>

4. Gobierno de México. *Compranet*. <https://www.gob.mx/compranet>
5. Gobierno de Brasil. *Portal Nacional de Contratações Públicas*. <https://www.gov.br/pncp/>

Estándares internacionales

1. Comisión Europea, Public Buyers Community (2025). *Updated EU AI model contractual clauses*, 5 de marzo de 2025: versión completa para alto riesgo, versión ligera y comentario, en 24 lenguas. Actualiza la versión del 29 de septiembre de 2023, anterior a la adopción del Reglamento (UE) 2024/1689. <https://public-buyers-community.ec.europa.eu/communities/procurement-ai/resources/updated-eu-ai-model-contractual-clauses>
2. Unión Europea (2024–2026). *Reglamento (UE) 2024/1689; Digital Omnibus on AI*, con traslado del alto riesgo del Anexo III al 2 de diciembre de 2027 y del Anexo I al 2 de agosto de 2028. <https://eur-lex.europa.eu/eli/reg/2024/1689/oj>
3. Office of Management and Budget (2025). *Memorandum M-25-22 (adquisición de IA)*, 3 de abril de 2025. <https://www.whitehouse.gov/omb/information-for-agencies/memoranda/>
4. ChileCompra (2023). *Bases Tipo para la licitación de proyectos de algoritmos e inteligencia artificial con requisitos éticos*, enero de 2023. <https://www.chilecompra.cl/2023/01/ya-se-encuentra-disponible-bases-tipo-para-licitar-proyectos-de-algoritmos-e-inteligencia-artificial-con-requisitos-eticos/>
5. GobLab, Universidad Adolfo Ibáñez. *Evaluación de Impacto Algorítmico, medición de sesgo y equidad, y Ficha de Transparencia Algorítmica*. <https://goblab.uai.cl/>

Serie de investigación de Enpath

1. Enpath (2024). *Compra pública de IA: ocho causas y diez cláusulas* (agosto). Esta biblioteca.
2. Enpath (2025). *Estándares de compra pública para sistemas de IA* (mayo) — **el instrumento de este scorecard**; *modelo de madurez* (febrero), origen de la regla de ambición que es el cuarto cero; *registros de algoritmos* (abril). Esta biblioteca.
3. Enpath (2026). *Marco Enpath de compra responsable de IA* (enero); *Estándares de transparencia en la compra de IA* (mayo); *Índice Enpath de gobernanza de IA: Colombia 2026* (marzo). Esta biblioteca.

Notas

Enpath · Investigación aplicada para decisiones públicas basadas en evidencia. Cada publicación incluye su metodología, sus fuentes y sus limitaciones.

1. Archivo `data/resultados_scorecard.csv`, publicado con este informe. **Ítems con cero procesos cumplidores: clasificación en matriz de impacto (0 de 20), contraste de madurez o regla de ambición (0 de 20), ficha publicada en registro (0 de 20) y métricas trimestrales publicadas (0 de 17).** Son cuatro, y los cuatro son de producción propia de la entidad. Agregados derivados del mismo archivo: cumplimiento global 145 de 525 = 27,6 %; media de tasas por ítem 26,7 %; mediana de tasas por ítem 27,9 %; mediana por proceso 28 %. Por fase, agregado: 21,1 · 38,0 · 35,8 · 20,2 · **11,7 %**; media de tasas por ítem: 21,8 · 34,9 · 35,8 · 20,3 · 15,2 %. Las dos agregaciones son legítimas y difieren hasta en 3,5 puntos en la fase 4; este informe usa el agregado y lo declara en cada cifra (§5).
2. Comisión Europea, Public Buyers Community. Las cláusulas modelo de contratación de IA se actualizaron el **5 de marzo de 2025** —nueve meses antes del corte de este informe— en tres documentos: versión completa para IA de alto riesgo alineada con el Reglamento (UE) 2024/1689, versión ligera personalizable para sistemas que no son de alto riesgo, y comentario de uso y aplicación; los tres en veinticuatro lenguas, español incluido. La versión anterior databa del 29 de septiembre de 2023 y es la que el informe citó. Relevancia directa: el estudio observó que los ítems de fase 2 específicos de IA promedian 26 % y atribuyó la brecha a la ausencia de estándar; desde marzo de 2025 existe texto contractual modelo en español (§10.1).
3. *Digital Omnibus on AI* de la Unión Europea: adoptado por el Parlamento Europeo el 16 de junio de 2026 y por el Consejo el 29 de junio de 2026. Traslada las obligaciones de los sistemas de alto riesgo del **Anexo III al 2 de diciembre de 2027** y las del **Anexo I al 2 de agosto de 2028**. Consecuencia invertida frente a la expectativa razonable de arrastre de mercado: el aplazamiento no da tiempo a la región, se lo quita, porque durante dos años y medio adicionales la documentación que un comprador latinoamericano no exija por contrato no la producirá nadie más (§10.2).
4. Agencia Nacional de Contratación Pública de Colombia, boletín *Radar de la compra pública*, edición N° 3 de 2026: anuncia un estudio diagnóstico sobre la contratación pública de soluciones tecnológicas que incorporan inteligencia artificial. Es la condición previa de un instrumento, no el instrumento. A 25 de julio de 2026 no se localizó ninguna agencia centralizada de la región que haya adoptado el estándar de mayo de 2025 ni equivalente como documento tipo o condición de entrada: **la premisa del experimento del informe se resolvió en negativo (§9).**