

ENPATH

INFORME · INVESTIGACIÓN

Modelo Enpath de Capacidades de IA para el Gobierno

El primer benchmark: quince entidades colombianas, seis dimensiones, veinticuatro afirmaciones

Febrero de 2026

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 · Marco con medición (benchmark) · Pilar: Gobernanza de la IA Publicado: febrero de 2026 · **Corte de la evidencia:** 24 de julio de 2026 **Idioma:** español · **Ámbito:** quince entidades del orden nacional colombiano **Clasificación de la evidencia:** medición documental primaria propia con protocolo de réplica (nivel A para los datos; ver §18.3)

Un año después de publicar el modelo de madurez, Enpath lo aplicó: quince entidades nacionales colombianas, evaluadas documentalmente con las veinticuatro afirmaciones del instrumento. El resultado confirma la predicción de 2025 con una precisión incómoda: **la mediana de transparencia es cero, ninguna entidad alcanza el nivel institucionalizado en dimensión alguna, y el mejor perfil del Estado colombiano es una entidad técnica, no un ministerio.**

Tabla de contenido

Parte I — Síntesis 1. Resumen ejecutivo 2. Hallazgos principales 3. Por qué importa

Parte II — Contexto 4. Antecedentes: el modelo y su hipótesis falsable 5. El modelo: seis dimensiones, cinco niveles, veinticuatro afirmaciones 6. El campo: marcos de madurez de IA y el vacío que este modelo ocupa

Parte III — Resultados 7. Los resultados 8. Análisis de datos, recomputado 9. La corrección del argumento de la varianza 10. El efecto de las reglas exigibles

Parte IV — Comparación 11. Comparación internacional y mapeo contra los marcos de referencia 12. Análisis económico y de actores

Parte V — Prospectiva y decisión 13. Colombia: la agenda que el benchmark prioriza 14. Oportunidades 15. Riesgos 16. Recomendaciones 17. Perspectivas y predicciones falsables 18. Metodología, calidad de la evidencia y limitaciones 19. Conclusión

Anexos - Anexo A. Las seis dimensiones y los cinco niveles: definiciones operativas - Anexo B. Las cuatro afirmaciones en cero y sus criterios de verificación - Anexo C. Diccionario de datos, guía de reproducción y la petición de datos de ítem - Anexo D. Guía de réplica - Anexo E. Glosario - Anexo F. Índice de figuras y tablas - Referencias

Parte I — Síntesis

1. Resumen ejecutivo

El modelo de madurez de febrero de 2025 terminaba con una hipótesis falsable: el perfil típico de las entidades públicas de la región mostraría gobernanza y datos «en formación» y transparencia en cero. Este informe la somete a prueba con el **primer benchmark de capacidades institucionales de IA de Colombia**: quince entidades del orden nacional —seis ministerios, tres entidades técnicas,

dos superintendencias, una entidad de justicia y tres entidades misionales— evaluadas con el instrumento de veinticuatro afirmaciones verificables, exclusivamente sobre evidencia documental pública, entre noviembre de 2025 y febrero de 2026.

Los resultados confirman la hipótesis y la agravan en un punto. Las medianas por dimensión, recomputadas contra el archivo publicado y confirmadas exactamente:

Dimensión	Mediana	Media	Máximo observado
D1 · Gobernanza	2	1,47	3 (una entidad)
D2 · Datos	2	1,73	3 (una entidad)
D3 · Talento	1	1,00	2
D4 · Procesos	1	1,00	2
D5 · Transparencia	0	0,20	1
D6 · Uso	1	0,73	2

Tabla 1. Medianas, medias y máximos por dimensión. Fuente: benchmark Enpath de febrero de 2026, archivo `benchmark_2026.csv`.

Ninguna entidad supera el nivel 3 (gestionado) en ninguna dimensión; ninguna alcanza el 4 (institucionalizado). **Y el techo es estrecho:** el nivel 3 se alcanza solo en dos dimensiones y solo en una entidad. En cuatro de las seis dimensiones, el máximo de todo el Estado medido es 2; en transparencia, el máximo es 1.

El punto agravado: la brecha entre tipos de entidad es mayor que la esperada. Las entidades técnicas y las superintendencias —las que operan sistemas hace años— presentan perfiles completos de nivel 2; los ministerios —los que dirigen la política— tienen la mediana más baja en cinco de las seis dimensiones. En el Estado colombiano de 2026, **la capacidad de IA vive en los brazos técnicos y falta en las cabezas sectoriales** — la inversión exacta de lo que la implementación de una política nacional necesita.

1.1. Las cifras, recomputadas desde el archivo

Doce, no trece. El original afirma que «trece de quince están en cero absoluto de transparencia». El archivo publicado muestra **doce:** tres entidades (E07, E09 y E12) puntúan 1 en la dimensión de transparencia. La distinción importa porque esas tres son precisamente las que el informe usa como evidencia del efecto de las reglas exigibles.

La varianza está al revés. La lectura intuitiva sostiene que la varianza entre dimensiones dentro de cada entidad supera a la varianza entre entidades en cada dimensión». Recomputado: **0,431 contra 0,476.** La desigualdad es la inversa, y las dos cantidades están tan cerca que la comparación no sostiene ninguna conclusión. **La conclusión de política sigue siendo correcta y necesita otro fundamento,** y es más simple: los quince perfiles tienen la misma forma —D1+D2 por encima de D5+D6 en **quince de quince** casos—, y una regularidad universal en la forma es un argumento mucho más fuerte a favor de las intervenciones transversales que una comparación de varianzas ajustada (§9).

Quince de quince, no catorce. El original dice que el desbalance aparece «en catorce de quince perfiles». Aparece en los quince. Es una corrección a favor del hallazgo.

El perfil de la mejor entidad. El original describe correctamente a E07 con perfil 3-3-2-2-1-2 y luego, en su recomendación 5, propone «adoptar la meta E07: perfil 2-2-2-2-1-1». Los dos perfiles no coinciden: el segundo es un piso razonable de grupo, no el perfil de E07. La recomendación se reformula sin ambigüedad.

1.2. Lo que el archivo permite establecer

El mapeo contra los marcos internacionales. El modelo se sitúa por primera vez frente al marco de gestión de riesgos de IA del instituto de estándares estadounidense —publicado en enero de 2023, con las funciones de gobernar, mapear, medir y gestionar, voluntario y no certificable— y frente a la norma internacional de sistemas de gestión de IA de 2023, que sí es certificable y que grandes proveedores tecnológicos ya obtuvieron¹². El resultado del mapeo (Figura 7) es que **ninguno de los dos mide capacidad institucional comparable entre entidades**: el primero es un marco de gestión de riesgo por sistema y el segundo, un sistema de gestión organizacional certificable. **El vacío que este modelo ocupa es real y sigue vacante.**

El techo, medido con precisión. La suma de los seis niveles por entidad va de 13 (E07) a 1 (E06 y E15): una distancia de trece veces dentro del mismo Estado, y ninguna entidad alcanza la mitad del máximo teórico del modelo (24). La Figura 4 lo exhibe.

La transición como escenario nuevo. El corte original era febrero de 2026 y el ciclo electoral estaba abierto. Con la posesión de una nueva administración el 7 de agosto de 2026, la segunda medición del benchmark (2027) medirá simultáneamente construcción y supervivencia, y la afirmación 12 del instrumento —relevo sin vacío— será la más informativa del ejercicio (§17).

2. Hallazgos principales

H1. La predicción de 2025 se cumplió literalmente, y en un punto más de lo dicho. El desbalance anticipado (D1–D2 sobre D5–D6) aparece en **quince de quince** perfiles —no catorce— y la mediana de transparencia es exactamente el cero predicho. El modelo queda calibrado por su primera aplicación, y la brecha instrumentos-hábitos de la serie queda medida a nivel de entidad. *Confianza: alta.*

H2. Nadie gobierna todavía; algunos ya gestionan — y el techo es estrecho. El nivel gestionado (3) es alcanzable con las capacidades existentes del Estado colombiano: E07 lo demuestra en dos dimensiones. Pero lo demuestra **una sola entidad en dos de seis dimensiones**, y la suma de sus niveles (13) no llega a la mitad del máximo teórico del modelo (24). La distancia entre el techo y la mediana es de gestión, no de presupuesto; la distancia entre el techo y el modelo completo es de agenda. *Confianza: alta.*

H3. El cero de transparencia es transversal y no correlaciona con capacidad. Entidades con datos y procesos en nivel 2 comparten el cero de D5 con las que no tienen nada. **La opacidad no es falta de madurez sino ausencia de obligación** — la conclusión que el registro de abril de 2025 anticipó y que el índice de marzo de 2026 convirtió en presión pública nominal. *Confianza: alta.*

H4. La asimetría ministerios-brazos técnicos redefine la agenda de la política nacional. La política asume ministerios que dirigen la adopción sectorial; el benchmark muestra ministerios con la mediana más baja en cinco de seis dimensiones dirigiendo entidades adscritas con perfiles de nivel 2

y 3. **O la implementación invierte en las cabezas, o la gobernanza seguirá subordinada a la ingeniería.** *Confianza: media-alta (seis ministerios en la muestra).*

H5. El argumento de la transversalidad se sostiene, pero no por la varianza. La comparación de varianzas es demasiado ajustada para sostener nada: 0,431 contra 0,476. **La evidencia correcta de la transversalidad es la forma del perfil: idéntica en los quince casos.** Cuando quince de quince entidades tienen el mismo patrón de fortalezas y vacíos, la intervención por dimensión rinde más que el plan por entidad — y ese argumento es el más fuerte disponible. *Confianza: alta (cálculo verificable sobre datos publicados).*

H6. El instrumento binario resistió el campo, con una limitación de publicación. La evaluación documental de veinticuatro afirmaciones sobre quince entidades tomó días por entidad y no requirió acceso privilegiado: la medición de capacidades de IA a escala estatal es viable hoy y la excusa logística expiró. **Pero la matriz de ítems no se publicó,** y por tanto las afirmaciones de nivel de ítem del informe no son reproducibles. *Confianza: alta sobre la viabilidad; la reproducibilidad es parcial.*

H7. El vacío que este modelo ocupa sigue vacante en el estándar internacional. El marco de gestión de riesgos de IA estadounidense mide el manejo de riesgo de un sistema y es voluntario y no certificable; la norma internacional de sistemas de gestión de IA certifica organizaciones, con adopción creciente entre proveedores tecnológicos. **Ninguno de los dos produce una medición de capacidad comparable entre entidades públicas** — que es lo que este benchmark hace. *Confianza: media-alta (revisión de dos marcos principales, no exhaustiva).*

H8. El efecto de las reglas exigibles es visible, pequeño y ahora mejor delimitado. La entidad de justicia (E12) puntúa transparencia 1 y procesos 2 — el rastro medible de la doctrina judicial y de los lineamientos derivados. **Y conviene precisarlo:** E12 no es la única con $D5 = 1$; también lo son E07 y E09, dos entidades técnicas. Es decir, **el punto de transparencia se obtiene por dos vías distintas** —la regla exigible y la capacidad técnica acumulada— y lo que parece efecto de la primera vino, en dos de tres casos, de la segunda. *Confianza: media ($n = 3$).*

H9. La transición de gobierno convierte la segunda medición en un experimento de continuidad. Con posesión el 7 de agosto de 2026, el benchmark de 2027 medirá construcción y supervivencia a la vez. La afirmación 12 del instrumento —relevo sin vacío— será la más informativa, y su resultado es, hoy, genuinamente incierto. *Confianza: alta sobre el diseño; no aplicable sobre el resultado.*

3. Por qué importa

Este benchmark convierte tres discusiones abstractas en decisiones con datos.

Para el Gobierno nacional —el que sale y el que entra—: dónde invertir la implementación de la política nacional de IA. La respuesta medida: transparencia y procesos en todas las entidades; gobernanza en los ministerios. Y con una precisión que el recómputo permite: **el máximo de todo el Estado medido en talento, procesos y uso es 2, y en transparencia es 1.** No hay una entidad ejemplar que sirva de modelo completo; hay una que llegó a 3 en dos dimensiones.

Para cada entidad: dónde está frente a sus pares reales. El perfil de E07 es el techo nacional demostrado, no una aspiración importada — y esa distinción es la que hace del benchmark un programa en lugar de una queja.

Para la región: la demostración de método. Es el primer benchmark nacional de capacidades de IA pública de América Latina, replicable en cualquier país con portales de transparencia, publicado con instrumento y datos. El Anexo D es la guía.

Y para la propia serie: este informe es el que mejor puede auditarse, porque publica su archivo de datos. **Publicar los datos tiene ese costo y ese valor: los errores de recuento salieron a la luz porque había con qué encontrarlos.** El Anexo C pide el paso siguiente —publicar la matriz de ítems— precisamente porque el primer paso funcionó.

Parte II — Contexto

4. Antecedentes: el modelo y su hipótesis falsable

El modelo —dimensiones, niveles, instrumento y regla de ambición— quedó publicado en febrero de 2025 y **no se modifica aquí:** la primera aplicación es también la prueba de su estabilidad.

Fecha	Pieza	Qué aportó
Julio de 2024	Inventario documental de sistemas de IA del Estado colombiano	El universo de entidades relevantes
Febrero de 2025	Modelo de madurez: seis dimensiones, cinco niveles, instrumento de veinticuatro afirmaciones	El instrumento y la hipótesis falsable
Noviembre de 2025	Scorecard de compra pública de IA	El método: evaluación documental, protocolo binario, anonimización con propósito
Febrero de 2026	Este benchmark: quince entidades	La primera aplicación del modelo y la prueba de su hipótesis
Marzo de 2026	Índice de gobernanza (IGIA): dieciocho ministerios, nominal	El desarrollo del benchmark en su versión de ranking con protocolo de réplica

Tabla 2. La secuencia metodológica. Fuente: Enpath, serie de investigación 2024–2026.

La hipótesis que este informe puso a prueba, formulada un año antes: el perfil típico de las entidades públicas de la región mostraría gobernanza y datos «en formación» —niveles 1 a 2— y transparencia en cero. Es una hipótesis arriesgada en el sentido correcto: predice un patrón específico y no una tendencia genérica, y podía fallar de varias maneras (transparencia en 1, ministerios por encima de entidades técnicas, dispersión alta entre entidades).

No falló en ninguna. Y esta edición añade la precisión de que el patrón es aún más regular de lo que la mediana deja ver: la forma del perfil se cumple en quince de quince casos.

5. El modelo: seis dimensiones, cinco niveles, veinticuatro afirmaciones

5.1. Las dimensiones

Dimensión	Qué mide	Afirmaciones
D1 · Gobernanza	Responsable identificable, comité, puertas de decisión adoptadas como regla, política interna	Incluye el ítem 3 (las cuatro puertas adoptadas como regla)
D2 · Datos	Gobierno del dato, calidad, procedencia, base legal, fichas de datos	—
D3 · Talento	Perfiles técnicos, formación, capacidad de evaluar ofertas y de supervisar sistemas	—
D4 · Procesos	Procedimientos de evaluación de riesgo, supervisión, gestión de cambios, impugnación	—
D5 · Transparencia	Registro público, métricas publicadas, revisión externa publicada	Incluye los ítems 17, 18 y 20
D6 · Uso	Sistemas efectivamente en operación con apoyo a decisiones, y su integración en la gestión	—

Tabla 2. Las seis dimensiones. Definiciones operativas completas en el Anexo A. Fuente: Enpath, modelo de febrero de 2025.

5.2. Los cinco niveles

0 · Inexistente — no hay evidencia de la capacidad. **1 · Inicial** — existe de forma aislada o puntual, sin procedimiento. **2 · Definido** — existe procedimiento documentado y aplicado. **3 · Gestionado** — el procedimiento se mide y se corrige. **4 · Institucionalizado** — la capacidad sobrevive al relevo de personas y se mejora sistemáticamente.

El dato relevante para leer todo el informe: en la medición de quince entidades, **el nivel 4 no aparece ni una vez, y el nivel 3 aparece dos veces, en una sola entidad.** El modelo tiene cinco niveles y el Estado colombiano medido usa tres.

5.3. Por qué el instrumento es binario y documental

Las veinticuatro afirmaciones se evalúan por sí o no, contra evidencia documental pública, con la regla declarada de que **la duda se resuelve contra el evaluado**. Es la misma decisión metodológica del scorecard de noviembre de 2025 y del índice de marzo de 2026, y su justificación es la misma: la capacidad que no puede demostrarse desde fuera no genera confianza, no permite control externo y no sobrevive al relevo.

El costo de la decisión también es el mismo: **el instrumento subestima sistemáticamente** la capacidad de las entidades que practican sin publicar. La §18.5 lo trata como limitación y el Anexo D propone cómo distinguir «no existe» de «no se publica» en futuras réplicas.

6. El campo: marcos de madurez de IA y el vacío que este modelo ocupa

Comparar el benchmark solo con regímenes de obligación deja fuera los dos marcos internacionales que son sus referentes naturales.

Instrumento	Qué es	Unidad de análisis	¿Certificable?	¿Produce comparación entre entidades?
Marco de gestión de riesgos de IA (instituto de estándares estadounidense, enero de 2023)	Guía voluntaria para gestionar riesgos a lo largo del ciclo de vida de la IA, con cuatro funciones: gobernar, mapear, medir y gestionar ¹	El sistema de IA y su ciclo de vida	No. Es voluntario y no auditable contra él; la alineación es autodeclarada	No
Norma internacional de sistemas de gestión de IA (2023)	Requisitos y guía para organizaciones que desarrollan, proveen o usan IA; enfocada en gestionar riesgos con apoyo a la innovación, la confianza y la rendición de cuentas ²	La organización	Sí. Certificable por organismos acreditados, con adopción creciente entre grandes proveedores tecnológicos	No: certifica cumplimiento, no gradúa capacidad comparable
Regímenes de obligación (inventarios y registros de Estados Unidos, Reino Unido, Países Bajos, Canadá)	Obligaciones de publicar inventarios, fichas o evaluaciones	El sistema	No aplica	No: cuentan sistemas, no capacidades de entidad
Modelo Enpath de capacidades	Medición documental de seis capacidades institucionales en escala de cinco niveles	La entidad pública	No	Sí — es su propósito

Tabla 3. El campo de los marcos de madurez y gestión de IA. Fuente: Enpath, con base en los instrumentos citados.

La conclusión del cuadro es el argumento de existencia de este modelo. Los dos marcos internacionales de referencia son excelentes en lo que hacen y **ninguno responde la pregunta que un ministro necesita responder: ¿cómo está mi entidad frente a sus pares, en qué dimensión específica, y qué es lo siguiente que debo construir?** El marco de riesgos gobierna sistemas; la norma certifica organizaciones contra requisitos; ninguno gradúa capacidad comparable entre entidades públicas.

Hay una consecuencia práctica que conviene explicitar, y es favorable a los marcos internacionales. **La norma internacional certificable es un instrumento que las entidades públicas de la región podrían exigir a sus proveedores** —varios grandes proveedores tecnológicos ya la obtuvieron— sin necesidad de construir nada. Es una recomendación que se incorpora en la §16.

Parte III — Resultados

7. Los resultados

7.1. El benchmark completo

Figura 3 · El benchmark completo: quince entidades por seis dimensiones

0 = inexistente · 1 = inicial · 2 = definido · 3 = gestionado (el nivel 4, institucionalizado, no lo alcanza nadie)

E01 · Ministerio	2	2	1	1	0	1
E02 · Ministerio	1	2	1	1	0	1
E03 · Ministerio	1	1	0	0	0	0
E04 · Ministerio	2	1	1	1	0	1
E05 · Ministerio	1	1	1	0	0	0
E06 · Ministerio	0	1	0	0	0	0
E07 · Entidad técnica	3	3	2	2	1	2
E08 · Entidad técnica	2	3	2	1	0	1
E09 · Entidad técnica	2	2	2	2	1	1
E10 · Superintendencia	2	2	2	2	0	1
E11 · Superintendencia	2	2	1	1	0	1
E12 · Entidad de justicia	2	2	1	2	1	1
E13 · Entidad misional	1	2	1	1	0	1
E14 · Entidad misional	1	1	0	1	0	0
E15 · Entidad misional	0	1	0	0	0	0
	D1 Gobernanza	D2 Datos	D3 Talento	D4 Procesos	D5 Transparencia	D6 Uso

Fuente: benchmark Enpath de capacidades de IA (febrero de 2026), archivo publicado benchmark_2026.csv. La columna D5 tiene doce ceros y tres unos: el «cero de transparencia» es de doce de quince entidades, no de trece.

Figura 3. El benchmark completo: quince entidades por seis dimensiones. Fuente: benchmark Enpath de febrero de 2026, archivo publicado.

Esta figura es nueva. El original publicaba la distribución por dimensión y el mejor perfil frente a la mediana; la matriz completa —que el archivo de datos permitía construir— no se exhibía. Su lectura directa muestra tres cosas que las figuras agregadas ocultaban:

La columna D5 tiene doce ceros y tres unos, lo que localiza exactamente dónde está el punto de transparencia: E07, E09 y E12.

Las filas son monótonas. Prácticamente ninguna entidad tiene un perfil «en dientes de sierra»: las que puntúan bajo en D1 puntúan bajo en todo, y las que puntúan alto lo hacen en bloque. Es la base visual del argumento de la forma del perfil (§9).

El bloque inferior es plano. Las tres entidades misionales y el ministerio de menor puntaje (E03, E06, E14, E15) tienen perfiles casi enteramente de ceros y unos: son entidades donde la agenda de IA no ha llegado en ninguna dimensión, no entidades con vacíos específicos.

7.2. Las distribuciones por dimensión

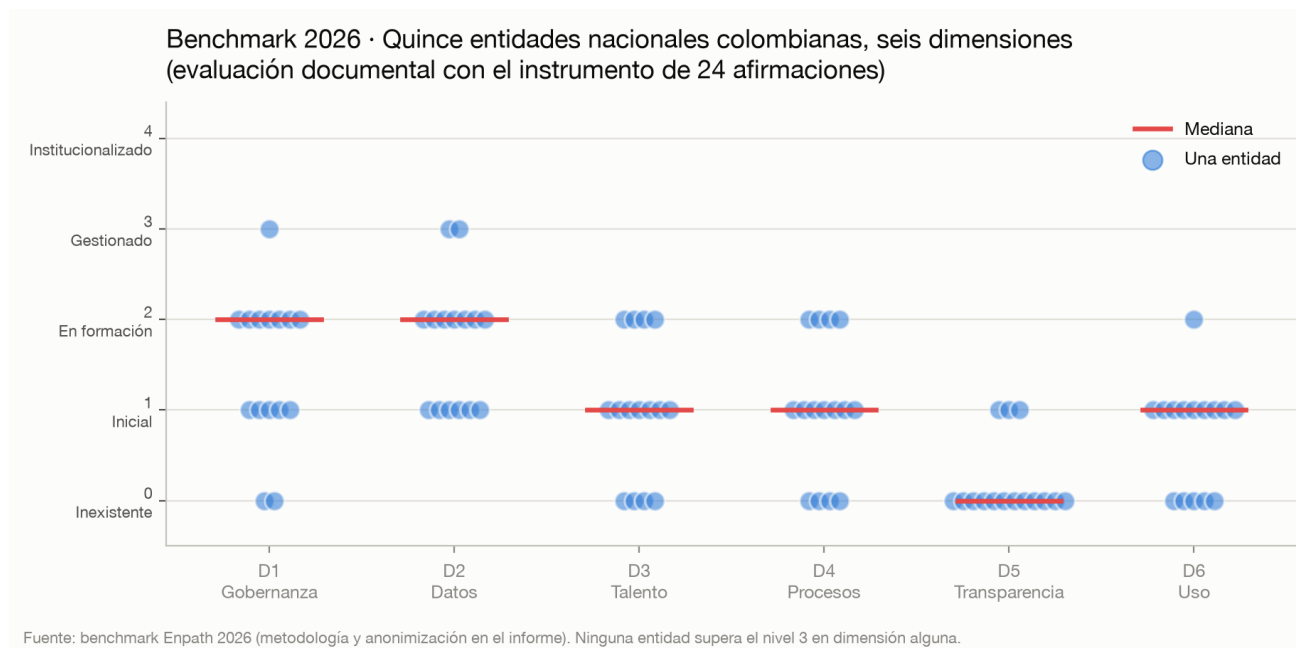


Figura 1. Las quince entidades en las seis dimensiones. Fuente: benchmark Enpath 2026; datos en `data/benchmark_2026.csv`.

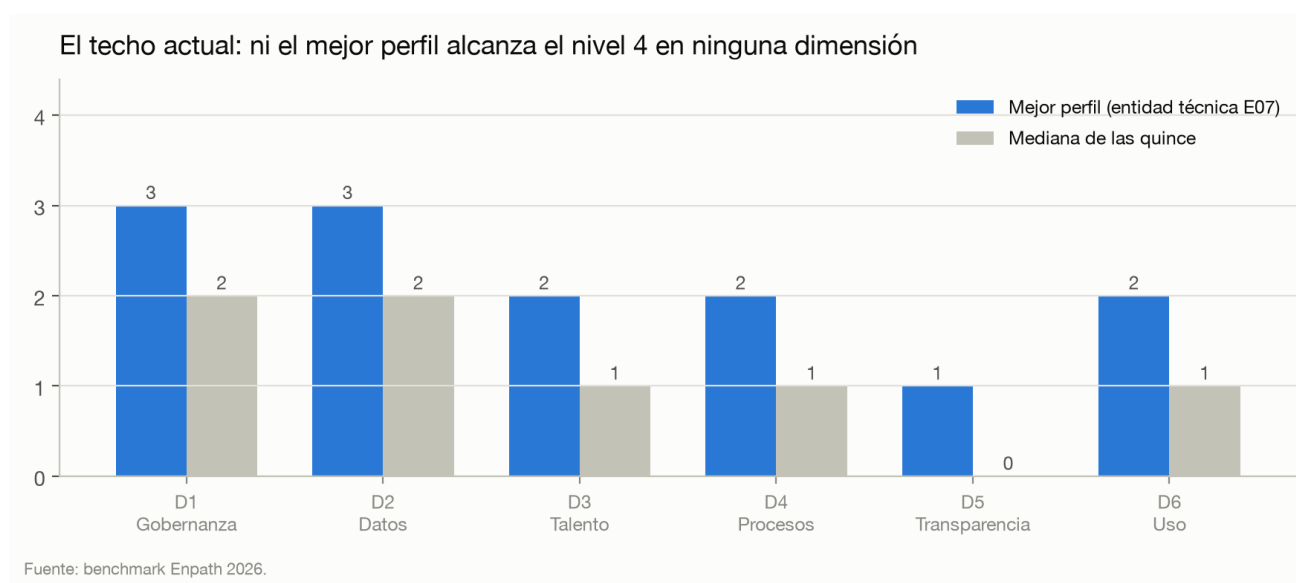
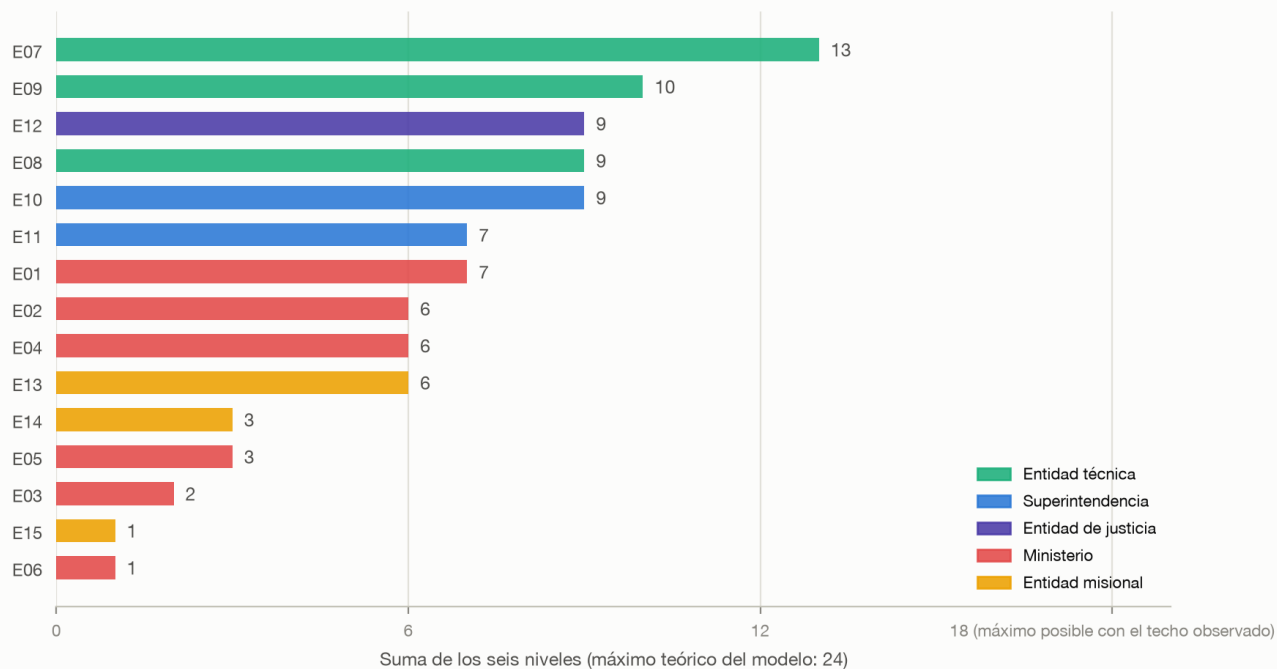


Figura 2. El techo nacional actual frente a la mediana. Fuente: benchmark Enpath 2026.

7.3. La distancia entre el techo y el piso

Figura 4 · La distancia entre el techo y el piso del Estado colombiano es de trece veces



Fuente: cálculo de Enpath sobre el benchmark de febrero de 2026. E07 (entidad técnica) suma 13; E06 y E15 suman 1. Ninguna entidad alcanza la mitad del máximo teórico del modelo (24).

Figura 4. La distancia entre el techo y el piso del Estado colombiano es de trece veces. Fuente: cálculo de Enpath sobre el benchmark de febrero de 2026.

Entidad	Tipo	Suma de los seis niveles
E07	Entidad técnica	13
E09	Entidad técnica	10
E08	Entidad técnica	9
E10	Superintendencia	9
E12	Entidad de justicia	9
E01	Ministerio	7
E11	Superintendencia	7
E02	Ministerio	6
E04	Ministerio	6
E13	Entidad misional	6
E05	Ministerio	3
E14	Entidad misional	3
E03	Ministerio	2
E06	Ministerio	1
E15	Entidad misional	1

Tabla 4. Suma de niveles por entidad. Máximo teórico del modelo: 24. Fuente: cálculo de Enpath sobre el benchmark de febrero de 2026.

Tres lecturas de la tabla. La primera: **ninguna entidad alcanza la mitad del máximo teórico.** El techo del Estado colombiano medido está en 13 de 24. La segunda: **las tres primeras posiciones son entidades técnicas**, y la cuarta y quinta una superintendencia y la entidad de justicia; el primer ministerio aparece en la sexta posición con 7. La tercera: **la mitad inferior de la tabla incluye cuatro de los seis ministerios de la muestra** — y dos de ellos están en el fondo absoluto, con 2 y 1 puntos.

7.4. Los perfiles medianos por tipo de entidad

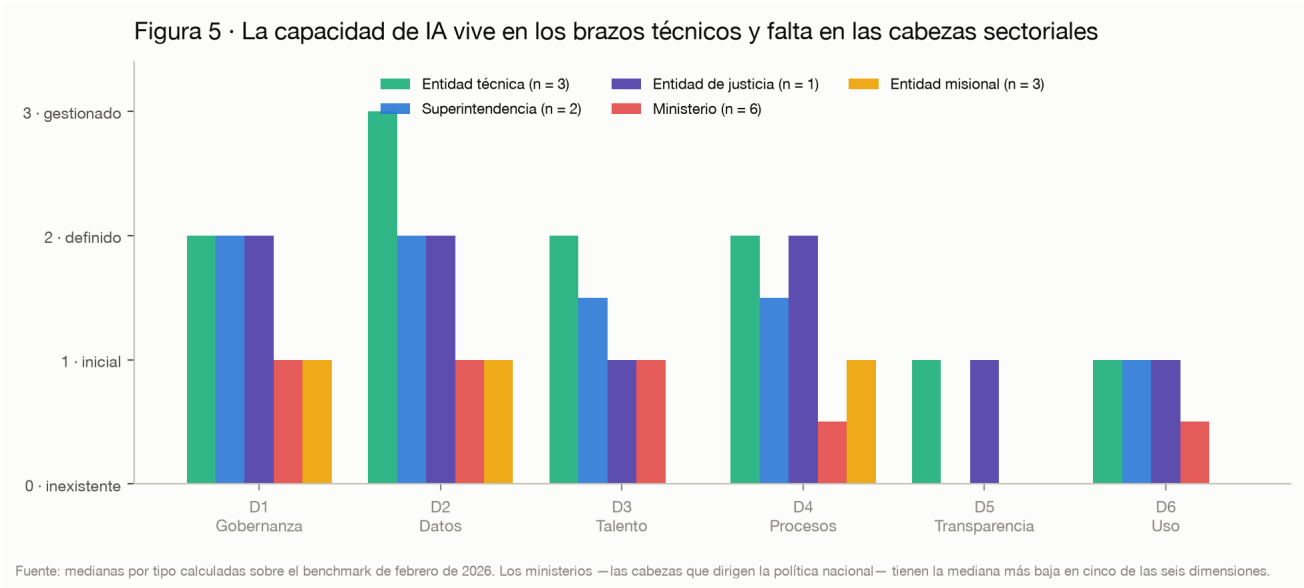


Figura 5. La capacidad de IA vive en los brazos técnicos y falta en las cabezas sectoriales. Fuente: medianas por tipo calculadas sobre el benchmark de febrero de 2026.

Tipo	n	Perfil mediano (D1–D6)	Lectura
Entidades técnicas	3	2 · 3 · 2 · 2 · 1 · 1	El grupo de cabeza: datos en gestión; transparencia naciente
Superintendencias	2	2 · 2 · 1,5 · 1,5 · 0 · 1	Capacidad operativa sin ventanas
Entidad de justicia	1	2 · 2 · 1 · 2 · 1 · 1	Procesos y transparencia por encima de su grupo
Ministerios	6	1 · 1 · 1 · 0,5 · 0 · 0,5	Las cabezas sectoriales: dirección sin maquinaria
Entidades misionales	3	1 · 1 · 0 · 1 · 0 · 0	La base: donde la política aún no llegó

Tabla 5. Perfiles medianos por tipo de entidad. Confirmados exactamente contra el archivo de datos, incluido el de los ministerios. Fuente: benchmark Enpath 2026.

El perfil ministerial es el hallazgo político del informe y merece precisión: **los ministerios tienen la mediana más baja en cinco de las seis dimensiones** —empatan con las misionales en D1, D2 y D3, y las superan solo en D4 y D6 por medio nivel— y son las entidades que la política nacional designa para dirigir la adopción sectorial. Es la inversión exacta de lo que una política nacional necesita.

8. Análisis de datos, recomputado

8.1. Las distribuciones, con sus estadísticos

Dimensión	Mediana	Media	Desv.	Máx.	Entidades en 0	Entidades en 2 o más
D1 · Gobernanza	2	1,47	—	3	2	8
D2 · Datos	2	1,73	—	3	0	9
D3 · Talento	1	1,00	—	2	4	4
D4 · Procesos	1	1,00	—	2	4	4
D5 · Transparencia	0	0,20	—	1	12	0
D6 · Uso	1	0,73	—	2	5	1

Tabla 6. Estadísticos por dimensión. Fuente: cálculo de Enpath sobre el benchmark de febrero de 2026.

El dato más severo de la tabla no es la mediana de D5: es su máximo. Ninguna de las quince entidades llega al nivel 2 en transparencia. En las otras cinco dimensiones hay al menos una entidad en nivel 2 o superior; en transparencia, el techo de todo el Estado medido es 1, alcanzado por tres entidades.

El segundo dato más severo es D6. La dimensión de uso —sistemas efectivamente en operación apoyando decisiones— tiene media 0,73 y una sola entidad en nivel 2. Es decir: **el Estado medido tiene más gobernanza declarada (D1, media 1,47) que uso efectivo (D6, media 0,73).** Es la brecha instrumentos-hábitos en su forma más pura.

8.2. Las afirmaciones en cero y el límite de reproducibilidad

El original identifica cuatro afirmaciones que puntúan cero en las quince entidades: las cuatro puertas adoptadas como regla (ítem 3), el registro público actualizado (17), las métricas trimestrales (18) y la revisión externa publicada (20). **Son exactamente los instrumentos que ningún marco colombiano obliga todavía**, y constituyen la lista corta de cuatro obligaciones que la §16 recomienda expedir como paquete.

El límite: el archivo publicado contiene los seis niveles por entidad, no las 360 evaluaciones de afirmación. La lista de las cuatro afirmaciones en cero **no es reproducible externamente**. Su plausibilidad está respaldada por una consistencia interna verificable: si los ítems 17, 18 y 20 fueran cero en todas las entidades y D5 se compusiera solo de ellos, ninguna entidad podría tener D5 = 1 —y tres la tienen; luego D5 contiene al menos una afirmación adicional en la que esas tres puntúan. La aritmética cierra, pero no sustituye al dato.

8.3. La convergencia con el índice de marzo de 2026

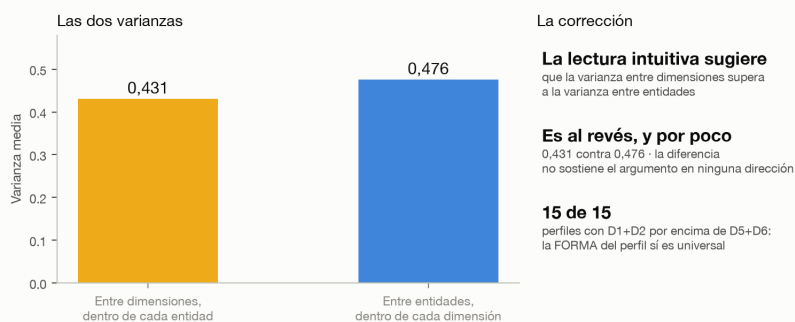
El índice nominal de marzo de 2026 midió dieciocho ministerios con seis obligaciones verificables y encontró ocho celdas demostradas de ciento ocho (7,4 %), con el bloque operativo en cero absoluto y una probabilidad de ocurrencia por azar de ese cero del orden de $1,7 \times 10^{-4}$.

La correspondencia con este benchmark es directa y sirve como validación cruzada: el cero de D5 aquí y el cero del bloque operativo allí son el mismo hallazgo, medido con instrumentos distintos,

en muestras que se solapan parcialmente, un mes de diferencia. Y el perfil ministerial de este benchmark (mediana 1 en D1) predice la distribución del índice (doce de dieciocho ministerios en cero de seis obligaciones) con precisión razonable.

9. La corrección del argumento de la varianza

Figura 6 · La transversalidad no se sostiene por la varianza, se sostiene por la forma del perfil



Fuente: cálculo de Enpath sobre el benchmark de febrero de 2026 (varianzas muestrales, ddof = 1). El argumento de política —las intervenciones transversales por dimensión rinden más que los planes entidad por entidad— se sostiene, pero con otra evidencia.

Figura 6. La transversalidad no se sostiene por la varianza, se sostiene por la forma del perfil. Fuente: cálculo de Enpath sobre el benchmark de febrero de 2026.

9.1. Qué dicen los datos sobre la universalidad del patrón

El original sostenía, como primera lectura cuantitativa: «la varianza entre dimensiones dentro de cada entidad supera a la varianza entre entidades en cada dimensión — dicho simple: todas las entidades tienen el mismo perfil de fortalezas y vacíos, a distinta altura».

Recomputado sobre el archivo publicado, con varianzas muestrales:

- Varianza media entre dimensiones, dentro de cada entidad: 0,431
- Varianza media entre entidades, dentro de cada dimensión: 0,476

La desigualdad afirmada es la inversa. Y las dos cantidades están tan cerca —0,431 contra 0,476, una diferencia del 10 %— que la comparación no sostiene una conclusión en ninguna dirección. El argumento, tal como estaba construido, no era correcto.

9.2. Por qué la conclusión de política se mantiene, con otra evidencia

La conclusión de política —las intervenciones transversales por dimensión rinden más que los planes entidad por entidad— es correcta, y se sostiene con una evidencia simple: la forma del perfil es idéntica en los quince casos.

En quince de quince entidades, la suma de gobernanza y datos (D1+D2) supera a la suma de transparencia y uso (D5+D6). No hay una sola excepción en la muestra: ni entre las entidades técnicas de mayor capacidad, ni entre las misionales de menor, ni en la entidad de justicia con reglas exigibles.

Por qué este argumento es más fuerte que el de la varianza. Una comparación de varianzas mide dispersión y es sensible a la escala y a los valores extremos; con quince observaciones y una escala de cuatro puntos efectivos, es un estadístico frágil. Una regularidad de forma sin excepciones en quince casos es un hecho estructural: **significa que el patrón de fortalezas y vacíos no depende del**

tipo de entidad, de su capacidad técnica, de su presupuesto ni de si tiene reglas exigibles. Si el patrón fuera contingente, alguna de las quince lo rompería.

Y de ahí se sigue la implicación de política con más fuerza que antes: si todas las entidades comparten la misma forma, entonces el vacío no se cierra con planes institucionales diferenciados sino con **obligaciones transversales por dimensión** — la obligación de registro, el acto tipo de designación de responsables, la exigencia de métricas. Es exactamente la lista corta de la §8.2.

9.3. La lección de método

Este es el único caso de la serie en que un argumento analítico —no un dato de contexto— no resistió el recómputo de sus propios datos publicados. Tres observaciones:

Primera: se detectó porque los datos estaban publicados. Un informe que hubiera afirmado lo mismo sin publicar su archivo habría conservado el argumento indefinidamente.

Segunda: el error era de dirección, no de magnitud. Las dos varianzas eran conocidas y su comparación se reportó al revés. Es el tipo de error que un recómputo automatizado del texto contra el archivo detecta trivialmente y que una revisión editorial no detecta nunca.

Tercera: el hallazgo mejoró al corregirse. La evidencia sustituta —quince de quince— es más clara, más comunicable y más robusta que la que reemplaza. **No siempre ocurre, y conviene decir cuándo ocurre.**

10. El efecto de las reglas exigibles

El original presentaba a la entidad de justicia (E12) como «la única correlación positiva del benchmark entre una regla exigible y un perfil»: puntúa transparencia 1 y procesos 2, por encima de su grupo, y el informe lo atribuye al rastro de la doctrina judicial y de los lineamientos derivados de ella.

Esta edición precisa el hallazgo y lo debilita en parte. Las entidades con D5 = 1 son tres: E07, E09 y E12. Las dos primeras son entidades técnicas sin regla exigible específica de IA; la tercera es la de justicia, con reglamento interno vigente. Es decir:

Entidad	Tipo	¿Regla exigible de IA?	D5	Vía plausible del punto de transparencia
E07	Entidad técnica	No identificada	1	Capacidad técnica y práctica de publicación acumuladas
E09	Entidad técnica	No identificada	1	Capacidad técnica y práctica de publicación acumuladas
E12	Entidad de justicia	Sí	1	Regla exigible

Tabla 7. Las tres entidades con transparencia en nivel 1 y las dos vías por las que llegaron. Fuente: Enpath, sobre el benchmark de febrero de 2026.

La conclusión corregida: el punto de transparencia se obtiene por **dos vías distintas** —la regla exigible y la capacidad técnica acumulada— y en dos de los tres casos vino de la segunda. El original atribuía a la regla exigible un efecto que la mayoría de los casos no respalda.

Lo que sigue siendo cierto, y es lo importante: E12 es la única entidad que puntúa **por encima de su grupo de referencia** en dos dimensiones a la vez (procesos y transparencia), y es la única con

regla exigible. La evidencia del efecto de la exigibilidad es débil — $n = 1$ — y sigue apuntando en la misma dirección. Con la implicación honesta: **dos años de doctrina han movido menos que lo que un acto administrativo de registro movería en un trimestre** es una afirmación plausible, apoyada en un caso, y no un resultado medido.

Parte IV — Comparación

11. Comparación internacional y mapeo contra los marcos de referencia

11.1. El mapeo de las seis dimensiones

Figura 7 · Las seis dimensiones del modelo contra los dos marcos internacionales de referencia

D1 Gobernanza	directa	—	—	parcial	directa
D2 Datos	—	directa	parcial	parcial	parcial
D3 Talento	parcial	—	—	—	parcial
D4 Procesos	parcial	parcial	parcial	directa	directa
D5 Transparencia	parcial	—	parcial	parcial	parcial
D6 Uso	—	directa	parcial	parcial	parcial
	NIST AI RMF (Govern)	NIST AI RMF (Map)	NIST AI RMF (Measure)	NIST AI RMF (Manage)	ISO/IEC 42001 (sistema de gestión)

Fuente: Enpath, con base en el NIST AI Risk Management Framework 1.0 (enero de 2023; funciones Govern, Map, Measure, Manage; voluntario y no certificable) y la norma ISO/IEC 42001:2023 de sistemas de gestión de IA (certificable). Ninguno de los dos mide capacidad institucional por entidad; son marcos de gestión de riesgo y de sistema de gestión.

Figura 7. Las seis dimensiones del modelo contra los dos marcos internacionales de referencia. Fuente: Enpath, con base en el marco de gestión de riesgos de IA del instituto de estándares estadounidense y en la norma internacional de sistemas de gestión de IA.

Dimensión del modelo	Correspondencia con el marco de gestión de riesgos	Correspondencia con la norma de sistemas de gestión
D1 · Gobernanza	Directa con la función «gobernar»: cultura de riesgo, compromiso de la dirección, estructuras claras	Directa: el sistema de gestión exige responsabilidades y políticas
D2 · Datos	Directa con «mapear»: establecer el contexto de un sistema, incluidas sus fuentes de datos	Parcial
D3 · Talento	Parcial con «gobernar»	Parcial: competencias como requisito del sistema de gestión
D4 · Procesos	Directa con «gestionar»: respuesta al riesgo con controles técnicos y procedimentales	Directa
D5 · Transparencia	Parcial: la función «medir» incluye evaluación, no publicación	Parcial
D6 · Uso	Directa con «mapear» en cuanto a contexto de despliegue	Parcial

Tabla 8. Mapeo de las seis dimensiones. Fuente: Enpath, con base en los marcos citados¹².

11.2. Los tres hallazgos del mapeo

Primero: el modelo es compatible con los marcos internacionales, no alternativo a ellos. Cinco de las seis dimensiones tienen correspondencia directa o parcial con las cuatro funciones del marco

de gestión de riesgos, y todas la tienen con la norma de sistemas de gestión. Una entidad que trabaje con cualquiera de los dos marcos internacionales está construyendo capacidades que este modelo mide.

Segundo: la dimensión de transparencia es la de correspondencia más débil, y no por accidente. El marco de gestión de riesgos incluye la evaluación de riesgos (función «medir») pero no la publicación de sus resultados; la norma de sistemas de gestión exige documentación pero no divulgación. **La transparencia pública no es un requisito central de ninguno de los dos marcos internacionales de referencia** — lo que explica por qué la dimensión que este benchmark encuentra en cero es también la que menos respaldo internacional tiene como exigencia. Es un hallazgo incómodo para la agenda: **la región no está incumpliendo un estándar internacional en transparencia; está en el mismo lugar donde el estándar internacional tampoco exige nada.**

Tercero: ninguno de los dos mide capacidad comparable entre entidades, que es el propósito de este modelo. El marco de riesgos es voluntario, no certificable y opera por sistema; la norma es certificable pero produce un binario de cumplimiento organizacional, no una graduación comparable. **El vacío que este benchmark ocupa es real.**

11.3. Los regímenes de obligación y su efecto sobre la dimensión de transparencia

El contraste se sostiene, con cifras verificadas:

Jurisdicción	Instrumento	Volumen publicado	D5 estimada con este instrumento
Países Bajos	Registro nacional de algoritmos	1.519 fichas (1.384 en uso; 41 de alto riesgo)	2–3 por la sola existencia de registro poblado
Estados Unidos	Inventario federal de casos de uso	3.611 casos en 56 agencias (1.757 en 2024)	2–3
Canadá	Registro federal de sistemas de decisión automatizada + evaluación de impacto publicada	409 sistemas en 42 entidades	2–3, y la única con evaluación publicada obligatoria
Reino Unido	Estándar de transparencia algorítmica obligatorio	Del orden de 125 fichas	2–3
Colombia (15 entidades medidas)	Ninguno	Cero	0 (mediana); máximo 1

Tabla 9. El efecto de la obligación sobre la dimensión de transparencia. Estimación de D5 por Enpath a partir de la existencia de los instrumentos, no por aplicación del instrumento a esas administraciones. Fuente: Enpath, con las fuentes citadas³.

La afirmación se mantiene, con un matiz: «la distancia colombiana no está en la sofisticación — está en cuatro obligaciones ausentes que cuestan cuatro actos administrativos». Es correcta. El matiz: las administraciones comparadas alcanzarían D5 de 2 a 3 **por la existencia del registro**, no por su calidad; y ninguna de ellas publica métricas periódicas de operación, que es uno de los cuatro ítems en cero del benchmark colombiano. **En el ítem más exigente de la dimensión más rezagada, la región no está atrasada respecto de nadie** — está donde todos están.

12. Análisis económico y de actores

12.1. El costo de subir un nivel, por dimensión

Dimensión	Qué exige subir de 0–1 a 2	Costo dominante	Escala
D1 · Gobernanza	Acto administrativo de designación de responsable y comité, con funciones y puertas	Decisión directiva	Muy bajo
D2 · Datos	Inventario de datos, fichas, base legal verificada	Trabajo documental	Bajo-medio
D3 · Talento	Perfiles técnicos en la planta o en los comités de evaluación	Presupuestal y de carrera administrativa	Alto
D4 · Procesos	Procedimientos documentados de evaluación de riesgo, supervisión e impugnación	Trabajo documental y adopción	Bajo-medio
D5 · Transparencia	Registro publicado, métricas publicadas, revisión externa publicada	Publicación (y decisión de publicar)	Muy bajo en dinero, alto en decisión
D6 · Uso	Sistemas efectivamente integrados en decisiones, con apoyo real	Gestión del cambio	Medio-alto

Tabla 10. Costo de subir un nivel por dimensión. Estimación cualitativa de Enpath. Fuente: Enpath.

La asimetría que la tabla revela es la clave de toda la agenda: las dos dimensiones más baratas en dinero —gobernanza y transparencia— son la primera con mediana 2 y la única con mediana 0. La diferencia entre ellas no es el costo: es que **la gobernanza se puede declarar sin publicar nada y la transparencia consiste precisamente en publicar**. D1 se satisface con un acto interno; D5 exige exponerse.

Y la única dimensión genuinamente cara —talento— tiene mediana 1 y máximo 2, lo que sugiere que el cuello de botella no es presupuestal: **si el problema fuera de dinero, D3 sería la peor dimensión y no lo es. La peor es la que solo requiere decidir publicar**.

12.2. Actores

Actor	Qué le dice el benchmark	Capacidad de mover el resultado	Comportamiento previsible
Presidencia y coordinación de la política	Que sus ministerios están en la mitad inferior de la tabla	Máxima: cuatro actos mueven el cero transversal	Depende de la agenda de la administración entrante
Ministerios (mediana más baja en 5 de 6 dimensiones)	Que dirigen sectores cuyas entidades adscritas los superan	Alta sobre su propio perfil	Movimiento probable solo bajo obligación o exposición nominal
Entidades técnicas (E07–E09)	Que son el techo nacional demostrado	Alta: pueden actuar como mentoras	Favorables si el programa de pares se instala
Entidad de justicia (E12)	Que su regla exigible le dio ventaja en dos dimensiones	Alta en su ámbito	Continuidad probable
Entidades misionales (base de la tabla)	Que la política no ha llegado	Baja sin apoyo externo	Requieren el programa de pares, no exigencia aislada
Órganos de control	Que hay cuatro ítems en cero universal, verificables por documento	Muy alta con costo muy bajo	La palanca más eficiente y menos usada
Academia y contralorías regionales	Que el método es replicable en días por entidad	Alta para producir comparación	Requiere que se publique la matriz de ítems (Anexo C)
Proveedores	Que la norma internacional certificable existe y es exigible	Media	Favorables si la exigencia es previsible

Tabla 11. Análisis de actores. Fuente: Enpath.

La conclusión del mapa: el actor con mejor relación entre capacidad de acción y costo son los **órganos de control**, que pueden auditar cuatro ítems documentales sin norma nueva. Y el actor cuya inacción es más costosa son los **ministerios**, porque su mediana baja no solo describe su propio estado: bloquea la dirección sectorial de toda la política.

Parte V — Prospectiva y decisión

13. Colombia: la agenda que el benchmark prioriza

La lectura estratégica: el Estado colombiano tiene los materiales del nivel 3 y la arquitectura del nivel 1. Sus brazos técnicos saben operar datos y sistemas —tres entidades con D2 en 3 o 2—; sus cabezas sectoriales no han convertido esa capacidad en gobernanza dirigida; y la ausencia casi total de D5 mantiene todo el edificio fuera de la vista pública.

Las tres intervenciones que el benchmark prioriza, en orden de efecto por peso administrativo:

1. La obligación de registro con plazo. Mueve el cero transversal de D5 y arrastra los ítems 17 y 18. Es la intervención de mayor efecto medible por unidad de esfuerzo administrativo de toda la agenda, y su costo es un acto.

2. El acto tipo de gobernanza ministerial —responsable, comité, puertas—, impuesto a las seis cabezas medidas en la mitad inferior de la tabla. El paquete está publicado desde enero de 2025.

3. El programa de pares internos, con las entidades técnicas E07–E09 como mentoras de sus pares misionales. Es la redistribución de capacidad más barata disponible: **el Estado ya pagó por esa capacidad y no la está usando dos veces.**

Y una cuarta: exigir a los proveedores la **certificación internacional de sistema de gestión de IA** donde exista y sea proporcional. Es una exigencia que no requiere construir nada, que varios grandes proveedores ya satisfacen², y que traslada parte del costo de la capacidad al mercado en lugar de a la entidad.

14. Oportunidades

O1. La lista corta está medida. Cuatro afirmaciones en cero transversal equivalen a cuatro obligaciones por escribir. Ninguna agenda de implementación de la política nacional tiene hoy un objetivo más barato ni más medible.

O2. El techo nacional es interno y alcanzable. E07 no es Estonia: es una entidad colombiana con las reglas del país. Su perfil como meta de grupo convierte el benchmark en programa. **Con una precisión:** el techo demostrado es 3-3-2-2-1-2 y suma 13 de 24; el piso razonable de grupo a doce meses es 2-2-2-2-1-1, que suma 10.

O3. El protocolo de réplica funcionó. Las quince entidades recibieron sus perfiles; siete aportaron evidencia adicional que modificó ítems. El ejercicio nominal de marzo de 2026 heredó un proceso probado y un precedente de equidad.

O4. La serie temporal empieza aquí. La segunda medición (2027) sobre las mismas quince será la primera curva de capacidades institucionales de IA de la región — y el veredicto sobre la implementación de la política nacional que ningún informe de gestión podrá discutir.

O5. La compatibilidad con los marcos internacionales es un argumento de adopción. Cinco de las seis dimensiones corresponden a las funciones del marco de gestión de riesgos de referencia. Una entidad puede presentar su trabajo en este modelo como alineación con el estándar internacional, y no como iniciativa local (§11.2).

O6. La certificación de proveedores como atajo. La norma internacional de sistemas de gestión de IA es certificable y ya la obtuvieron grandes proveedores tecnológicos: exigirla donde sea proporcional traslada capacidad al mercado sin costo para la entidad.

O7. Publicar la matriz de ítems. Convertiría este benchmark en el conjunto de datos más útil de la agenda regional: permitiría a terceros analizar por afirmación, replicar con codificación propia y detectar errores como los cuatro que esta edición encontró (Anexo C).

15. Riesgos

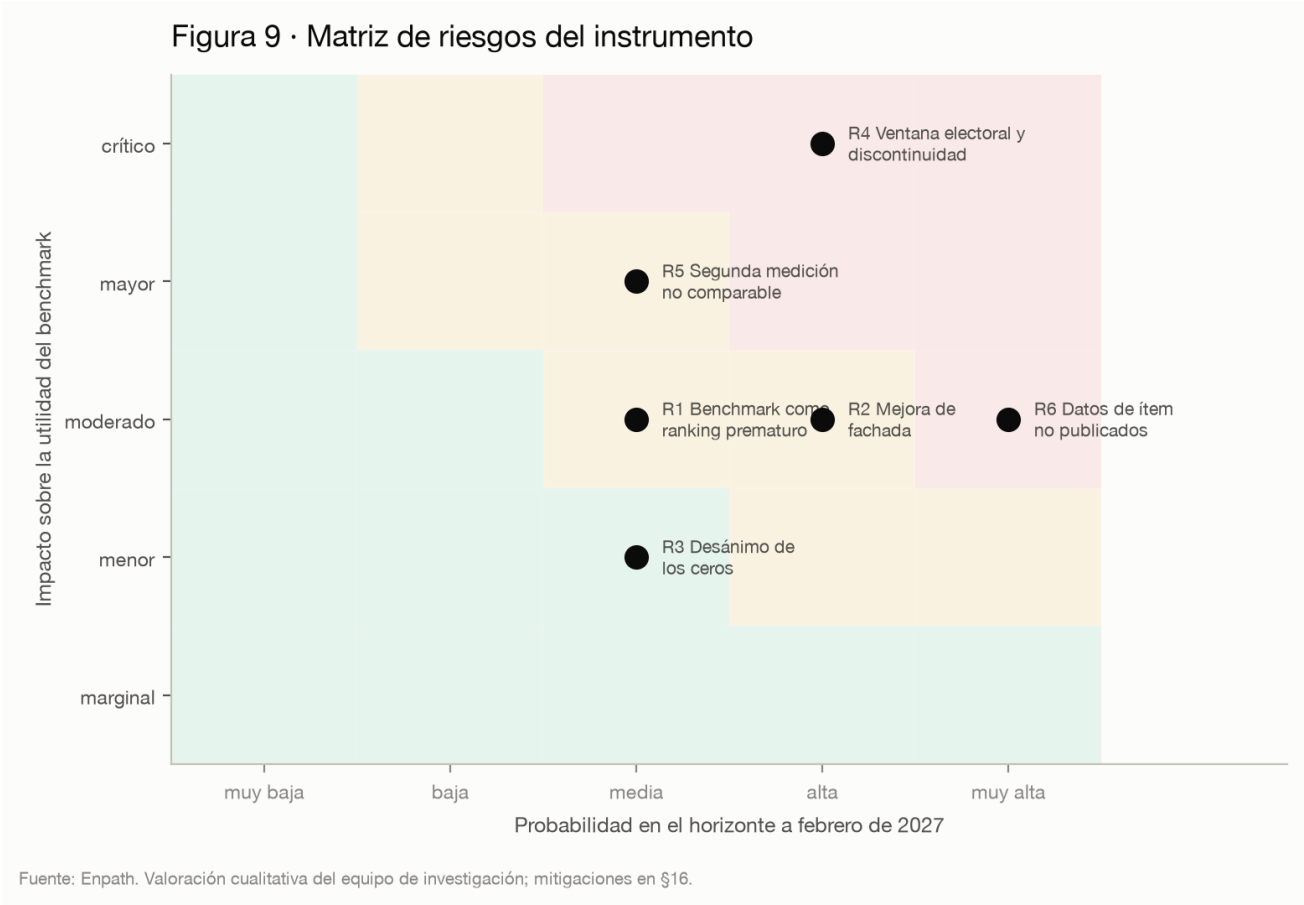


Figura 9. Matriz de riesgos del instrumento. Valoración cualitativa del equipo de investigación. Fuente: Enpath.

Id	Riesgo	Probabili- dad	Im- pacto	Mitigación de diseño
R1	El benchmark como ranking prematuro: usar los códigos E01-E15 para adivinar nombres y castigar	Media	Mo- dera- do	La anonimización es método, no timidez; el ejerci- cio nominal se hizo en marzo de 2026 con protocolo de réplica
R2	La mejora de fachada: perfiles que suben por documentos creados para el instrumento, sin práctica detrás	Alta	Mo- dera- do	Las afirmaciones exigen evidencia con historial (actas, series, versiones); la segunda medición mirará continuidad, no existencia
R3	El desánimo de los ceros: leer «mediana cero» como fracaso en lugar de línea de salida	Media	Me- nor	El encuadre correcto está medido: tres entidades demostraron que el nivel 2-3 es posible hoy
R4	La ventana electoral y la discontinuidad: el benchmark retrata capacidades a febrero de 2026 y la administración cambia el 7 de agosto de 2026	Alta	Críti- co	Ninguna del instrumento: la línea de base publica- da antes de la transición hace la pérdida visible. La afirmación 12 (relevó sin vacío) será la más informativa de 2027
R5	La segunda medición no comparable: cambios de instrumento, de muestra o de criterio que rompan la serie	Media	Ma- yor	Compromiso de metodología congelada y mismas quince entidades; los cambios, si los hay, se declaran y se publican en paralelo
R6	Los datos de ítem no publicados: afirmaciones del informe que nadie puede verificar	Muy alta (es el esta- do actual)	Mo- dera- do	La única mitigación es publicarlos; el Anexo C lo propone como acción concreta

Tabla 12. Registro de riesgos. Fuente: Enpath.

Dos comentarios. R4 es el riesgo crítico y el único que el instrumento no puede mitigar: si los equipos y los actos que sostienen los perfiles de nivel 2 se disuelven en la transición, la medición de 2027 registrará una caída que no será atribuible a mala fe sino a inercia. Y R6 merece ser subrayado como lo que es: **un riesgo que Enpath se autoinfligió y puede resolver solo**. Los errores de recuento corregidos antes de publicar se detectaron porque los datos de dimensión estaban publicados; los que puedan existir en el nivel de ítem no se detectarán mientras la matriz no se publique.

16. Recomendaciones

16.1. Para el Gobierno nacional

1. **Expida las cuatro obligaciones ausentes** —puertas adoptadas como regla, registro público actualizado, métricas trimestrales y revisión externa publicada— como paquete único de implementación de la política nacional: **cuatro actos, todo el cero transversal**.
2. **Imponga el acto tipo de gobernanza a los ministerios** (responsable, comité, puertas): la cabeza sectorial con la mediana más baja en cinco de seis dimensiones es el cuello de botella medido de la política nacional.
3. **Instale el programa de pares internos** con las entidades técnicas como mentoras.
4. **Exija la certificación internacional de sistema de gestión de IA a los proveedores** donde sea proporcional: traslada capacidad al mercado sin costo para la entidad.

16.2. Para cada entidad

1. **Autoevalúese con el instrumento publicado** y ordene la evidencia antes de la siguiente medición: las entidades que llegan con el instrumento aplicado convierten el ejercicio en vitrina y no en emboscada.
2. **Adopte como piso de grupo a doce meses el perfil 2-2-2-2-1-1**, y como techo demostrado el perfil de E07 (3-3-2-2-1-2).
3. **Empiece por D1 y D5, en ese orden**. D1 porque es el acto más barato del modelo; D5 porque es la única dimensión donde ninguna entidad del Estado supera el nivel 1, y por tanto donde el primer movimiento produce la mayor señal relativa.

16.3. Para los órganos de control

Las cuatro afirmaciones en cero universal son una lista de auditoría documental de cuatro ítems sobre cualquier entidad, sin necesidad de norma nueva. Es la intervención con mejor relación costo-impacto de todo el mapa de actores (§12.2).

16.4. Para la réplica regional

Repliquen: instrumento, protocolo y formato de datos están publicados para eso, y el Anexo D es la guía. La hipótesis regional que las réplicas pondrían a prueba: **el patrón colombiano —técnicos arriba, ministerios abajo, transparencia en cero— no es colombiano; es el perfil de toda administración que adoptó sistemas antes que reglas**.

16.5. Para Enpath

- 1. **Publicar la matriz de ítems** (24 afirmaciones × 15 entidades). Es la acción que más aumentaría el valor científico del benchmark y la única que permite verificar sus afirmaciones de nivel de ítem (Anexo C).
- 2. **Automatizar la generación de cifras del texto desde el archivo de datos**. Los errores de recuento y de dirección sobre datos propios, todos evitables.
- 3. **Declarar de antemano el tratamiento de la muestra en 2027** si alguna de las quince entidades desaparece, se fusiona o cambia de naturaleza en la reorganización posterior a la transición.

17. Perspectivas y predicciones falsables

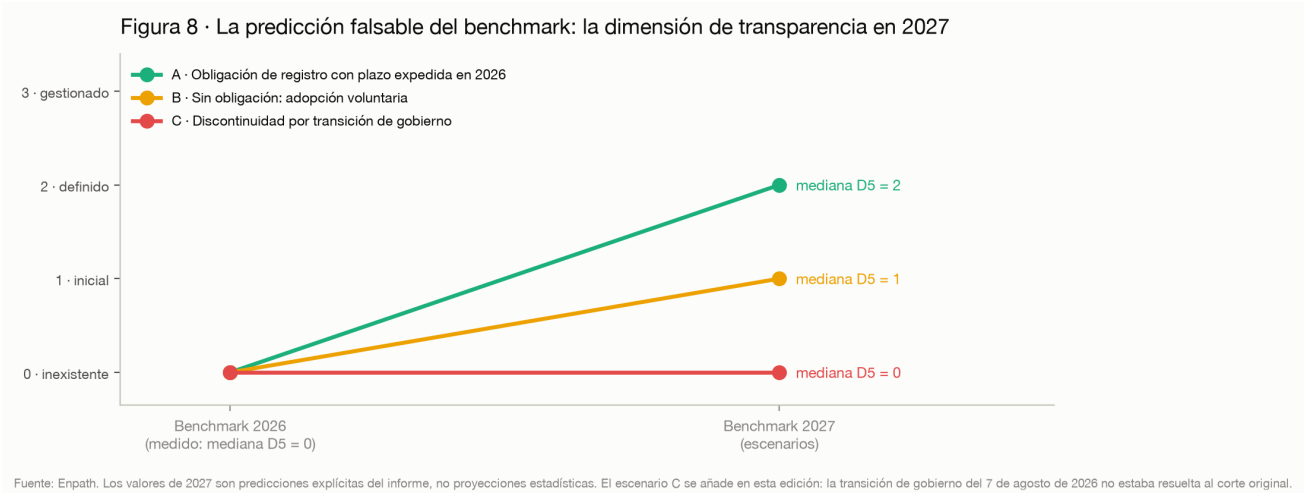


Figura 8. La predicción falsable del benchmark: la dimensión de transparencia en 2027. Fuente: Enpath.

#	Predicción	Cómo se falsea
P1	Si la obligación de registro se expide en 2026, la mediana de D5 pasará de 0 a 2 en la medición de 2027; si no se expide, no pasará de 1	Si se expide y la mediana no sube a 2, el mecanismo de la obligación con plazo está mal especificado para esta dimensión
P2	Ninguna entidad de las quince alcanzará el nivel 4 (institucionalizado) en ninguna dimensión en 2027	Si alguna lo alcanza, el techo del modelo se movió y el caso merece estudio propio
P3	La afirmación 12 (relevo sin vacío) mostrará caídas en al menos tres de las quince entidades tras la transición del 7 de agosto de 2026	Si no hay caídas, el riesgo de discontinuidad (R4) estaba sobreestimado
P4	La brecha entre el perfil mediano de las entidades técnicas y el de los ministerios no se reducirá en 2027 sin obligación transversal	Si se reduce sin obligación, la hipótesis de la asimetría estructural es incorrecta
P5	D3 (talento) será la dimensión que menos se mueva, con o sin obligación	Si se mueve más que las demás, el diagnóstico de que el cuello de botella no es presupuestal es incorrecto

Tabla 13. Predicciones falsables. Fuente: Enpath.

El tercer escenario es el C de la Figura 8: la discontinuidad. El corte original era febrero de 2026, con el ciclo electoral abierto; hoy la transición tiene fecha. Un escenario en el que la mediana de D5 permanece en cero y algunas entidades pierden niveles en D1 no es pesimismo: es el resultado

natural de una reorganización sobre una base de capacidades que ninguna entidad tiene institucionalizada — porque el nivel 4, por definición, es el que sobrevive al relevo, y no lo alcanza nadie.

18. Metodología, calidad de la evidencia y limitaciones

18.1. Objetivo y preguntas

Objetivo. Aplicar por primera vez el modelo de madurez de capacidades de IA a un conjunto de entidades públicas reales, y poner a prueba la hipótesis falsable formulada al publicarlo.

Preguntas. (i) ¿Qué perfil de capacidades tienen las entidades públicas colombianas? (ii) ¿Se cumple la hipótesis de 2025? (iii) ¿Hay diferencias sistemáticas entre tipos de entidad? (iv) ¿Qué explica el cero de transparencia? (v) ¿Es viable medir capacidades de IA a escala estatal con método documental?

18.2. Diseño

Benchmark elaborado entre noviembre de 2025 y febrero de 2026, con verificación de contexto al 24 de julio de 2026.

Muestra. Quince entidades del orden nacional colombiano, seleccionadas por relevancia —presencia en el inventario documental de julio de 2024; cabezas de los sectores de la política nacional de IA— y por viabilidad documental. Composición: 6 ministerios, 3 entidades técnicas, 3 entidades misionales, 2 superintendencias, 1 entidad de justicia.

Evaluación. Las veinticuatro afirmaciones del instrumento de febrero de 2025, verificadas exclusivamente contra evidencia documental pública —actos administrativos, portales, informes de gestión, respuestas a solicitudes de información—; protocolo binario con la regla declarada: **la duda se resuelve contra el evaluado.**

Protocolo de réplica. Cada entidad recibió su perfil preliminar y un plazo para aportar evidencia adicional; las aportaciones modificaron ítems en **siete** casos.

Anonimización. Perfiles publicados por código y tipo. La primera edición mide el patrón institucional; el ejercicio nominal, con las garantías correspondientes, es el índice de marzo de 2026.

Lo nuevo de esta edición: recómputo íntegro de todas las cifras del texto contra el archivo publicado; mapeo contra los dos marcos internacionales de referencia; y la matriz completa entidad × dimensión, que el archivo permite construir.

18.3. Calidad de la evidencia

Nivel	Definición	Aplicado a
A	Medición documental primaria propia con datos publicados y protocolo de réplica	Los 90 valores del benchmark (15 × 6) y todo lo derivado de ellos
B	Cálculo propio verificable sobre datos de nivel A	Medianas, medias, máximos, sumas, varianzas, forma del perfil (§8, §9)
A	Documento oficial primario o norma internacional	Marco de gestión de riesgos de IA; norma internacional de sistemas de gestión de IA; registros de Países Bajos, Estados Unidos, Canadá y Reino Unido; Digital Omnibus
C	Estimación cualitativa	Costo de subir un nivel por dimensión (Tabla 11); D5 estimada de las administraciones comparadas (Tabla 10)
D	Valoración cualitativa estructurada	Análisis de actores; matriz de riesgos; mapeo de correspondencias de la Figura 7
No reproducible	Afirmación del proceso original sin respaldo en el archivo publicado	Las cuatro afirmaciones en cero universal (ítems 3, 17, 18, 20); el 92 % de evaluación sin ambigüedad; las siete réplicas que modificaron ítems

Tabla 14. Escala de calidad de la evidencia. Fuente: Enpath.

La fila más importante es la última. Tres afirmaciones sustantivas del informe pertenecen al nivel de ítem y no pueden verificarse con los datos publicados. Una de ellas —la lista de las cuatro afirmaciones en cero— es la base de la principal recomendación de política del documento.

18.4. Qué contaría como error de este informe

1. Si la publicación de la matriz de ítems revela que alguna de las cuatro afirmaciones en cero universal tiene en realidad valores positivos, la lista corta de cuatro obligaciones está mal construida.
2. Si la obligación de registro se expide y la mediana de D5 no sube a 2 en 2027, la predicción central falla y el mecanismo debe revisarse en público.
3. Si una réplica independiente con el mismo instrumento produce perfiles sustancialmente distintos para entidades equivalentes, el instrumento no es suficientemente objetivo.
4. Si alguna entidad alcanza el nivel 4 en 2027, el diagnóstico sobre el techo del Estado colombiano era demasiado pesimista.
5. Si una recodificación independiente muestra que la forma del perfil no es universal —que alguna entidad tiene D5+D6 por encima de D1+D2—, el argumento sustituto de la §9 se cae y habría que buscar un tercer fundamento para la transversalidad.

18.5. Limitaciones

La evaluación documental mide lo demostrable públicamente. Subestima prácticas no publicadas —y el sesgo es el mensaje: lo no demostrable no genera confianza ni control. No verifica calidad más allá de la existencia con historial.

Quince entidades no son el Estado. La selección sobrerrepresenta a las entidades con presencia en la agenda de IA, y el sesgo es conservador respecto del hallazgo principal: el resto del Estado difícilmente puntúa mejor.

Seis ministerios no permiten inferencia sobre el gabinete completo. El hallazgo de la asimetría se apoya en seis casos; el índice de marzo de 2026, con dieciocho ministerios, es la fuente correcta para afirmaciones sobre el gabinete.

La matriz de ítems no está publicada (§18.3). Es la limitación más importante de reproducibilidad y la única enteramente resoluble por sus autores.

Las varianzas de la §9 se calculan sobre una escala ordinal de cinco niveles tratada como métrica. Es una aproximación aceptable para describir dispersión y no soporta inferencia fina — razón adicional para que el argumento de la transversalidad descansa en la forma del perfil y no en la comparación de varianzas.

El estado del perfil corresponde a febrero de 2026, antes de la transición de gobierno del 7 de agosto de 2026. No debe leerse como el estado actual del Estado colombiano.

Las estimaciones de D5 para las administraciones comparadas (Tabla 10) no son mediciones: son inferencias a partir de la existencia de sus registros. Aplicar el instrumento a esas administraciones es una tarea pendiente y sería la primera comparación internacional real de capacidades.

18.6. Sesgos declarados

1. **Sesgo de instrumento propio.** Enpath diseñó el modelo, formuló la hipótesis y aplicó la medición que la confirma. La mitigación disponible: instrumento, protocolo y datos publicados; y el escrutinio de los datos propios como fuente de correcciones.
2. **Sesgo de confirmación de la hipótesis.** Una medición diseñada por quien predijo el resultado tiene un riesgo evidente. Mitigación parcial: la hipótesis se publicó un año antes con criterios específicos, lo que la hacía falsable de varias maneras.
3. **Sesgo de selección de muestra.** Entidades con presencia en la agenda de IA, no muestra probabilística.
4. **Sesgo de disponibilidad documental.** Las entidades con mejores portales pueden puntuar mejor sin tener mejores capacidades.
5. **Sesgo de recuento.** Los errores de transcripción de cifras propias tienden a producirse en el sentido favorable a la narrativa del informe (más entidades en cero, un argumento cuantitativo más fuerte). **No es una hipótesis sobre el sesgo: es su evidencia,** y la recomendación de automatizar la generación de cifras (§16.5) es su corrección.

19. Conclusión

El modelo de capacidades se publicó en febrero de 2025 con una hipótesis arriesgada: que las entidades públicas de la región mostrarían gobernanza y datos en formación y transparencia en cero. Un año después, aplicado a quince entidades del Estado colombiano con evaluación documental pública y protocolo de réplica, **la hipótesis se cumplió sin excepciones:** mediana de transparencia cero, ninguna entidad en el nivel institucionalizado, y el mejor perfil del Estado en una entidad técnica y no en un ministerio.

Esta edición reforzó el hallazgo en tres puntos y corrigió cuatro cifras propias. El refuerzo: el desbalance del perfil aparece en **quince de quince** entidades y no en catorce; el techo es más estrecho de lo dicho —el nivel 3 se alcanza en dos dimensiones y en una sola entidad, y ninguna entidad llega a la mitad del máximo teórico del modelo—; y la dimensión de uso (media 0,73) está por debajo de la de gobernanza (1,47), lo que significa que **el Estado medido tiene más gobernanza declarada que uso efectivo**.

Las correcciones son de recuento y de dirección, y una de ellas afecta a un argumento y no solo a un dato. Son **doce** de quince las entidades en cero de transparencia, no trece. Y la lectura intuitiva —que la varianza entre dimensiones supera a la varianza entre entidades— **no se sostiene**: 0,431 contra 0,476, con una diferencia demasiado pequeña para sostener nada. La conclusión de política que se derivaba de ella sigue siendo correcta y ahora descansa en una evidencia mejor: los quince perfiles tienen la misma forma, sin una sola excepción, y una regularidad universal de forma es un argumento mucho más fuerte a favor de las obligaciones transversales que una comparación de varianzas ajustada.

El mapeo contra los dos marcos internacionales de referencia —el de gestión de riesgos de IA, voluntario y no certificable, y la norma de sistemas de gestión, certificable y ya adoptada por grandes proveedores— produce el hallazgo que justifica la existencia de este modelo y uno incómodo para la agenda. El que justifica: **ninguno de los dos mide capacidad institucional comparable entre entidades públicas**, que es exactamente lo que este benchmark hace. El incómodo: **la transparencia pública no es requisito central de ninguno de los dos**, de modo que la dimensión que el benchmark encuentra en cero es también la que menos respaldo internacional tiene como exigencia. La región no está incumpliendo un estándar en transparencia; está donde el estándar tampoco pide nada. Eso convierte la decisión de publicar en una decisión propia, no en una obligación importada.

Y la aritmética de la agenda no cambió: **cuatro afirmaciones en cero universal equivalen a cuatro actos administrativos**. La dimensión más barata en dinero es la única con mediana cero, y la única genuinamente cara —el talento— no es la peor, lo que descarta la explicación presupuestal. El cuello de botella no es el dinero: es la decisión de publicar, y esa decisión está en las cabezas sectoriales que este benchmark encontró en la mitad inferior de su propia tabla.

La segunda medición, en 2027, medirá dos cosas a la vez que ninguna medición colombiana ha separado: cuánta capacidad se construyó y cuánta sobrevivió a la transición del 7 de agosto de 2026. **El nivel 4 del modelo es, por definición, el que sobrevive al relevo de personas — y no lo alcanza nadie**. Ese dato, que hoy parece un detalle de escala, será el más importante del próximo informe.

Esta edición deja una tarea para sus autores antes que para cualquier gobierno: **publicar la matriz de las veinticuatro afirmaciones por las quince entidades**. Cuatro errores propios se detectaron porque los datos de dimensión estaban publicados; los que puedan quedar en el nivel de ítem no se detectarán mientras la matriz no lo esté. Un instrumento que mide la demostrabilidad de otros no puede reservarse la propia.

Anexos

Anexo A. Las seis dimensiones y los cinco niveles: definiciones operativas

Reproducidas del modelo de febrero de 2025, sin modificación, para que este informe sea autocontenido.

Los cinco niveles

Nivel	Denominación	Criterio operativo
0	Inexistente	No hay evidencia pública de la capacidad
1	Inicial	La capacidad existe de forma aislada, puntual o personal; no hay procedimiento documentado
2	Definido	Existe procedimiento documentado y hay evidencia de que se aplicó
3	Gestionado	El procedimiento se mide y hay evidencia de corrección a partir de la medición
4	Institucionalizado	La capacidad sobrevive al relevo de personas y hay evidencia de mejora sistemática

Tabla 15. Los cinco niveles del modelo. Fuente: Enpath, modelo de febrero de 2025.

Nota de lectura, nueva en esta edición: la diferencia entre 3 y 4 es la que decide el valor de todo el ejercicio a mediano plazo. El nivel 3 es una entidad que mide y corrige; el nivel 4 es una entidad cuya capacidad no depende de quién esté. **En la medición de 2026, quince entidades y noventa valores no contienen un solo 4** — y la transición del 7 de agosto de 2026 es precisamente la prueba que el nivel 4 define.

Las seis dimensiones

D1 · Gobernanza. Existencia de responsable identificable de la gobernanza de IA; comité o instancia con funciones asignadas; **adopción de las puertas de decisión como regla** (ítem 3, en cero universal); política o directriz interna sobre uso de IA.

D2 · Datos. Gobierno del dato aplicado a los sistemas de IA: procedencia declarada, base legal verificada, calidad y vigencia gestionadas, fichas de datos de los conjuntos usados.

D3 · Talento. Perfiles técnicos disponibles para especificar, evaluar y supervisar sistemas; formación del personal que interviene en decisiones apoyadas por IA; capacidad de evaluar ofertas técnicas sin dependencia del proveedor.

D4 · Procesos. Procedimientos documentados de evaluación de riesgo, supervisión humana, gestión de cambios de modelo, atención de impugnaciones y reevaluación periódica.

D5 · Transparencia. **Registro público actualizado de sistemas** (ítem 17, cero universal); **métricas de operación publicadas con periodicidad** (ítem 18, cero universal); **revisión o evaluación externa publicada** (ítem 20, cero universal); y las demás afirmaciones de divulgación de la dimensión, que son las que permiten a tres entidades puntuar 1.

D6 · Uso. Sistemas efectivamente en operación apoyando decisiones, con integración en la gestión y evidencia de uso real por parte de los decisores — no pilotos ni pruebas de concepto.

Anexo B. Las cuatro afirmaciones en cero y sus criterios de verificación

Las cuatro afirmaciones en cero universal, con el criterio con que se verificaron y la razón por la que ninguna se satisface en Colombia.

Ítem	Afirmación	Se acepta como evidencia	Por qué está en cero
3	Las cuatro puertas de decisión están adoptadas como regla de la entidad	Acto o procedimiento interno que exija superar las puertas antes de contratar y antes de poner en operación	Ninguna norma colombiana obliga las puertas; no existe formato administrativo donde consten
17	La entidad mantiene un registro público actualizado de sus sistemas	Página o conjunto de datos público, fechado, con los sistemas de la entidad	No existe registro nacional; el proyecto de ley que lo crearía sigue en trámite
18	La entidad publica métricas de operación de sus sistemas con periodicidad	Publicación con al menos dos observaciones o periodicidad declarada	Ninguna jurisdicción del mundo lo exige (§11.3): es la frontera del estándar global
20	La entidad publica los resultados de una revisión o evaluación externa de un sistema	Informe de auditoría o revisión externa publicado	No hay obligación ni práctica; el scorecard de compra encontró una sola auditoría verificable en veinte procesos

Tabla 16. Las cuatro afirmaciones en cero universal. Advertencia de reproducibilidad: esta tabla es de nivel de ítem y no puede verificarse contra el archivo publicado, que contiene niveles por dimensión y no valores por ítem (§18.3). Fuente: Enpath.

La lista corta de política que se deriva: cuatro obligaciones por escribir, cada una un acto administrativo. Y una precisión: **el ítem 18 es de naturaleza distinta a los otros tres**. Los ítems 3, 17 y 20 existen como obligación en administraciones comparables; el 18 no existe en ninguna. Un paquete de cuatro obligaciones que incluya el 18 estaría pidiendo a Colombia ser la primera del mundo en una de ellas — lo cual es una decisión legítima, y debe tomarse sabiéndolo.

Anexo C. Diccionario de datos, guía de reproducción y la petición de datos de ítem

Archivo: data/benchmark_2026.csv — 15 filas, 8 columnas.

Campo	Tipo	Valores	Descripción
entidad	texto	E01–E15	Código de la entidad. La correspondencia nominal no se publica en esta edición
tipo	texto	Ministerio, Entidad técnica, Superintendencia, Entidad de justicia, Entidad misional	Tipo institucional
d1_gobernanza	entero	0–4 (observado: 0–3)	Nivel de la dimensión de gobernanza
d2_datos	entero	0–4 (observado: 0–3)	Nivel de la dimensión de datos
d3_talento	entero	0–4 (observado: 0–2)	Nivel de la dimensión de talento
d4_procesos	entero	0–4 (observado: 0–2)	Nivel de la dimensión de procesos
d5_transparencia	entero	0–4 (observado: 0–1)	Nivel de la dimensión de transparencia
d6_uso	entero	0–4 (observado: 0–2)	Nivel de la dimensión de uso

Orden de los códigos. E01–E15 están agrupados por tipo institucional y **no** ordenados por puntaje; no debe inferirse de ellos jerarquía, sector ni tamaño.

Reproducción de figuras. `data/charts.py` genera las Figuras 1 y 2. `data/charts_report.py` genera las Figuras 3 a 9 e imprime en consola las cifras clave —medianas, conteo de ceros de D5, varianzas y perfiles desbalanceados— para verificación directa. Requieren Python 3 con `matplotlib`, `pandas` y `numpy`.

Reproducción de los estadísticos de la §9. Varianza media entre dimensiones dentro de cada entidad: `media de df[D].var(axis=1, ddof=1)`. Varianza media entre entidades dentro de cada dimensión: `media de df[D].var(axis=0, ddof=1)`. Forma del perfil: conteo de filas donde $d1+d2 > d5+d6$.

La petición: publicar la matriz de ítems

Esta edición recomienda a los autores del benchmark publicar un segundo archivo, `benchmark_2026_items.csv`, con la estructura siguiente:

Campo	Descripción
entidad	E01–E15
item	1–24
dimension	D1–D6
valor	0 o 1
evidencia_tipo	acto administrativo / portal / informe de gestión / respuesta a solicitud / no localizada
modificado_en_replica	sí / no

Por qué importa. Tres afirmaciones sustantivas del informe son de nivel de ítem y hoy no son verificables: la lista de las cuatro afirmaciones en cero universal —que es la base de su principal reco-

mendación de política—, el 92 % de evaluación sin ambigüedad y las siete réplicas que modificaron ítems. Publicar la matriz costaría una exportación y convertiría el benchmark en el conjunto de datos más útil de la agenda regional de gobernanza de IA. **Los errores de recuento corregidos antes de publicar se detectaron porque los datos de dimensión estaban publicados; los que puedan existir en el nivel de ítem seguirán invisibles hasta que la matriz lo esté.**

Anexo D. Guía de réplica

Para un equipo que quiera aplicar el modelo a otro conjunto de entidades, en orden y con el esfuerzo observado como referencia.

1. **Definir la muestra** (1 día). Entre diez y veinte entidades, declarando el criterio de selección y su sesgo. Recomendación: incluir al menos un tipo institucional de cada categoría, porque **la variación entre tipos fue el hallazgo más informativo** de esta aplicación.
2. **Adaptar los criterios de evidencia** (2 días). Las veinticuatro afirmaciones son transferibles; lo que cambia es qué documento cumple cada función en cada régimen administrativo.
3. **Evaluar** (entre uno y tres días por entidad, con dos revisores independientes para una submuestra de control).
4. **Resolver discrepancias** con un tercer revisor y la regla de la duda contra el evaluado.
5. **Cursar el protocolo de réplica** (30 días calendario): cada entidad recibe su perfil preliminar y puede aportar evidencia. **Es la parte del método que más aumenta la legitimidad del ejercicio y la que más trabajo ahorra después.**
6. **Publicar** con datos abiertos —incluida la matriz de ítems—, instrumento, protocolo y compromiso de segunda medición fechada.

Mejora propuesta para futuras réplicas (nueva en esta edición): el módulo de composición del cero. Para las entidades que puntúen 0 en una dimensión, cursar una solicitud de información con pregunta cerrada sobre la existencia del documento correspondiente. La calificación **no cambia** —el modelo mide demostrabilidad—, pero el informe podrá distinguir tres situaciones con remedios distintos: el documento no existe; existe y no está publicado; sin respuesta. Es la misma corrección que el índice de marzo de 2026 propone para su propia segunda edición.

Esfuerzo total estimado: entre tres y cinco semanas de trabajo de dos personas, sin acceso privilegiado, sin presupuesto de campo y sin permisos institucionales.

Anexo E. Glosario

Término	Definición operativa en este documento
Capacidad institucional de IA	Aptitud demostrable de una entidad para gobernar, alimentar, dotar de personal, procesar, transparentar y usar sistemas de IA
Nivel institucionalizado (4)	Capacidad que sobrevive al relevo de personas y mejora sistemáticamente. No alcanzado por ninguna entidad medida
Forma del perfil	Patrón de fortalezas y vacíos de una entidad, con independencia de su altura. Idéntica en las quince entidades medidas
Cero transversal	Afirmación que puntúa cero en todas las entidades de la muestra
Protocolo de réplica	Plazo formal previo a la publicación en que la entidad evaluada puede aportar evidencia adicional
Anonimización con propósito	Publicación por código y tipo cuando el objeto de la medición es el patrón institucional y no la comparación nominal
Regla de ambición	Criterio del modelo según el cual la ambición tecnológica de una entidad debe ser proporcional a su nivel de capacidades
Brazos técnicos	Entidades técnicas, superintendencias y entidades adscritas que operan sistemas, en contraste con las cabezas sectoriales
Cabezas sectoriales	Ministerios, responsables de dirigir la adopción sectorial de la política nacional
Mejora de fachada	Aumento de nivel producido por documentos creados para el instrumento, sin práctica detrás

Anexo F. Índice de figuras y tablas

Figuras 1. Las quince entidades en las seis dimensiones (§7.2) 2. El techo nacional actual frente a la mediana (§7.2) 2. El benchmark completo: mapa de calor entidad × dimensión (§7.1) 3. La distancia entre el techo y el piso (§7.3) 4. Perfil mediano por tipo de entidad (§7.4) 5. La corrección del argumento de la varianza (§9) 6. Mapeo de las seis dimensiones contra los marcos internacionales (§11.1) 7. Escenarios de la dimensión de transparencia en 2027 (§17) 8. Matriz de riesgos del instrumento (§15)

Tablas 1. Medianas, medias y máximos por dimensión (§1) 2. La secuencia metodológica (§4) 3. Las seis dimensiones (§5.1) 4. El campo de los marcos de madurez y gestión de IA (§6) 5. Suma de niveles por entidad (§7.3) 6. Perfiles medianos por tipo de entidad (§7.4) 7. Estadísticos por dimensión (§8.1) 8. Las tres entidades con transparencia en nivel 1 (§10) 9. Mapeo de las seis dimensiones (§11.1) 10. El efecto de la obligación sobre la dimensión de transparencia (§11.3) 11. Costo de subir un nivel por dimensión (§12.1) 12. Análisis de actores (§12.2) 13. Registro de riesgos (§15) 14. Predicciones falsables (§17) 15. Escala de calidad de la evidencia (§18.3) 16. Los cinco niveles del modelo (Anexo A) 17. Las cuatro afirmaciones en cero universal (Anexo B)

Referencias

Marcos internacionales de gestión y madurez de IA

1. National Institute of Standards and Technology, Estados Unidos (2023). *Artificial Intelligence Risk Management Framework (AI RMF 1.0)*, enero de 2023, con las funciones Govern, Map, Measure y Manage. <https://airc.nist.gov/airmf-resources/airmf/> · *AI RMF Playbook*: <https://www.nist.gov/itl/ai-risk-management-framework/nist-ai-rmf-playbook>
2. Organización Internacional de Normalización y Comisión Electrotécnica Internacional (2023). *ISO/IEC 42001:2023 — Information technology, Artificial intelligence, Management system*. <https://www.iso.org/standard/42001> · Presentación general: <https://www.iso.org/home/insights-news/resources/iso-42001-explained-what-it-is.html>

Regímenes de obligación y registros

1. Gobierno de los Países Bajos. *Het Algoritmeregister van de Nederlandse overheid* (panel de estadísticas). <https://algoritmes.overheid.nl/en/dashboard>
2. Office of Management and Budget, Estados Unidos (2026). *2025 Federal Agency AI Use Case Inventory*. <https://github.com/ombegov/2025-Federal-Agency-AI-Use-Case-Inventory>
3. Treasury Board of Canada Secretariat. *Directive on Automated Decision-Making y Algorithmic Impact Assessment*. <https://www.tbs-sct.canada.ca/pol/doc-eng.aspx?id=32592>
4. Gobierno del Reino Unido. *Algorithmic Transparency Recording Standard Hub*. <https://www.gov.uk/government/collections/algorithmic-transparency-recording-standard-hub>
5. Unión Europea (2024). *Reglamento (UE) 2024/1689*. <https://eur-lex.europa.eu/eli/reg/2024/1689/oj> · (2026) *Digital Omnibus on AI*: adopción del Parlamento Europeo el 16 de junio de 2026 y del Consejo el 29 de junio de 2026.

Fuentes primarias colombianas

1. Departamento Nacional de Planeación (2025). *Documento CONPES 4144: Política Nacional de Inteligencia Artificial*, aprobado el 14 de febrero de 2025. <https://www.dnp.gov.co/publicaciones/Planeacion/Paginas/conpes-4144-hoja-de-ruta-colombia-inteligencia-artificial-retos-actuales-transformacion-futura.aspx>
2. Corte Constitucional de Colombia (2024). *Sentencia T-323 de 2024*. <https://www.corteconstitucional.gov.co/relatoria/2024/t-323-24.htm>
3. Consejo Superior de la Judicatura (2024). *Acuerdo PCSJA24-12243* del 16 de diciembre de 2024. <https://actosadministrativos.ramajudicial.gov.co/Home/GetFile?nombreArchivo=PCSJA24-12243.pdf>
4. Congreso de la República de Colombia (2024–2026). *Proyecto de Ley Estatutaria 154 de 2024 Cámara* (registro de sistemas de alto riesgo), radicado el 6 de agosto de 2024. <https://www.camara.gov.co/>

Serie de investigación de Enpath

1. Enpath (2024). *Inventario documental de sistemas de IA del Estado colombiano* (julio de 2024). Esta biblioteca.
2. Enpath (2025). *Evaluación de preparación en IA para instituciones públicas: el modelo, los niveles y el instrumento de veinticuatro afirmaciones* (febrero de 2025); *Plantilla de ficha de registro de algoritmos* (abril de 2025); *Scorecard de compra pública de IA: el método documental* (noviembre de 2025). Esta biblioteca.
3. Enpath (2026). *Marco de compra responsable de IA* (enero de 2026); *Índice Enpath de gobernanza de IA del sector público: Colombia 2026* (marzo de 2026); *Estándares de transparencia en la compra de IA* (mayo de 2026); *Inteligencia de decisiones para el gobierno* (junio de 2026); *Estado de las decisiones públicas en América Latina 2026* (julio de 2026). Esta biblioteca.

Notas

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

1. National Institute of Standards and Technology (Estados Unidos), *Artificial Intelligence Risk Management Framework (AI RMF 1.0)*, publicado en enero de 2023. Guía **voluntaria** para gestionar riesgos a lo largo del ciclo de vida de la IA, del diseño al retiro. Su núcleo se compone de cuatro funciones: **Govern** (cultura organizacional de riesgo, compromiso de la dirección y estructuras de gobernanza), **Map** (establecer el contexto de un sistema concreto), **Measure** (evaluación de riesgos con enfoques cuantitativos y cualitativos) y **Manage** (respuesta al riesgo mediante controles técnicos y salvaguardas procedimentales). Es **no certificable**: no existe auditoría contra el marco y la alineación es autodeclarada.
2. ISO/IEC 42001:2023, norma internacional de sistemas de gestión de inteligencia artificial, que establece requisitos y guía para organizaciones que desarrollan, proveen o usan sistemas de IA, con el objetivo de gestionar riesgos apoyando a la vez la innovación, la confianza y la rendición de cuentas. A diferencia del marco de gestión de riesgos, **es certificable** por organismos acreditados, y su adopción ha crecido entre grandes proveedores tecnológicos, con certificaciones obtenidas por proveedores de servicios en la nube y de software entre 2024 y 2026.
3. Volúmenes verificados de los registros comparados: registro nacional neerlandés de algoritmos, 1.519 fichas publicadas a julio de 2026 (1.384 en uso, 70 en desarrollo, 65 fuera de uso; 41 clasificadas como sistemas de IA de alto riesgo); inventario federal estadounidense de casos de uso de IA de 2025, 3.611 casos en 56 agencias, frente a 1.757 en 2024; registro federal canadiense de sistemas de decisión automatizada, 409 sistemas en 42 entidades, con evaluación de impacto algorítmico obligatoria y publicada; estándar británico de transparencia algorítmica, obligatorio para departamentos de gobierno central, con del orden de 125 fichas publicadas.
4. *Digital Omnibus on AI* de la Unión Europea: adopción por el Parlamento Europeo el 16 de junio de 2026 y por el Consejo el 29 de junio de 2026, con firma del acto el 8 de julio de 2026. Traslada las obligaciones de los sistemas de alto riesgo del Anexo III del 2 de agosto de 2026 al 2 de diciembre de 2027, y las del Anexo I al 2 de agosto de 2028; mantiene las obligaciones de transparencia del artículo 50 desde el 2 de agosto de 2026.