



Co-funded by
the European Union

Índice de Madurez Normativa para el análisis del Marco Regulatorio de un país sobre IA en Educación

V 1.0

Español





Título del documento	Observatorio Regional de IA en Educacion - Latinoamerica y el Caribe
Work package	WP2
Autor	Jorge Maldonado Mahauad (UCuenca)
Colaboradores	Mar Pérez-Sanagustín Carlos Alario-Hoyos
Socios involucrados	Ucuenca, UT, UC3M
Versión	1.0
Estatus	Final



Construcción del Índice de Madurez Normativa-IMN

1. Propósito, alcance y definición

Este documento describe, de forma autocontenida y reproducible la construcción del Índice de Madurez Normativa (IMN). El objetivo es que cualquier lector externo pueda comprender cómo se construye el IMN, cómo se calcula, cómo se asigna, cómo se transforman en un indicador comparable entre países y, finalmente, cómo deben interpretarse sus resultados. La descripción se acompaña de las fórmulas, las escalas y las decisiones de diseño necesarias para replicar el procedimiento.

El corpus analizado está compuesto por 813 instrumentos normativos levantados por el equipo de investigadores en 33 países de Latinoamérica y el Caribe, registrados en una matriz común de análisis normativo.

Índice de Madurez Normativa (IMN). Indicador compuesto, expresado en una escala de 0 a 100, que mide el grado de desarrollo del marco regulatorio de un país sobre el uso de la inteligencia artificial en educación. Es un índice de naturaleza intensiva: en lugar de sumar el volumen de instrumentos normativos, promedia la calidad de la regulación existente, entendida como la conjunción de tres atributos —la fuerza vinculante de sus mecanismos, la orientación estratégica de su enfoque y el grado de consolidación de su ecosistema—, complementados con la amplitud o comprehensividad de su cobertura. El índice se obtiene como el promedio ponderado de seis subíndices normalizados (uno por cada dimensión del Marco Dimensional GENIAL), organizados en un bloque de calidad —mecanismo (S3), enfoque (S5) y madurez (S6)— y un bloque de amplitud —actor (S1), objeto (S2) y finalidad (S4)—. Para evitar que los países con pocos instrumentos obtengan posiciones infladas, los subíndices de calidad se ajustan por un factor de credibilidad que pondera la evidencia disponible. El IMN es intensivo y absoluto: no depende de un país de referencia y resulta comparable entre actualizaciones sucesivas del corpus. De este modo, el IMN permite situar a cada país en un continuo común de madurez normativa, identificar tipologías regionales y vacíos de gobernanza, y —al ser transparente y reponderable— operar tanto como instrumento de diagnóstico como de planificación de política pública.

2. Construcción del Índice de Madurez Normativa (IMN)

2.1 Los seis subíndices

El IMN se compone de seis subíndices, uno por dimensión del marco, cada uno normalizado en una escala absoluta de 0 a 100. Se organizan en dos bloques. El bloque de calidad —que mide la madurez de forma directa— está formado por los subíndices de las dimensiones D3, D5 y D6, calculados como la media por norma de una escala de intensidad. El bloque de amplitud —que mide el alcance— está formado por los subíndices de D1, D2 y D4, calculados como la proporción de categorías cubiertas por el conjunto de normas del país.

Bloque de calidad – intensidad por norma (media por norma × 100). A cada norma se le asigna un valor según la categoría más fuerte presente en la dimensión, y el subíndice del país es el promedio de esos valores:



Subíndice	Escala de intensidad por norma
S3 — Mecanismo (D3)	Prohibición / Obligación / Sanción = 1,0 Gobernanza institucional = 0,8 Principio ético = 0,5 Lineamiento / Formación = 0,4 No determinado = 0,2
S5 — Enfoque (D5)	Estratégico = 1,0 Preventivo = 0,85 Habilitador = 0,75 Reflexivo = 0,55 Restringido = 0,45 Reactivo = 0,25
S6 — Madurez (D6)	Ecosistema maduro = 1,0 Reg. consolidada = 0,8 Reg. inicial = 0,6 Declaración de intención = 0,4 Exploración = 0,2

Bloque de amplitud - proporción de cobertura (proporción de cobertura × 100). S1 es el número de categorías de actor cubiertas dividido entre 6; S2 es el número de objetos de IA cubiertos dividido entre 8; S4 es el número de finalidades cubiertas dividido entre 8. Estos subíndices miden cuán comprehensivo es el ecosistema, pero reciben menor peso porque son los más correlacionados con el simple volumen de normas.

Subíndice	Valor por categoría
S1 — Actor (D1)	Presente = 1, Ausente = 0
S2 — Objeto (D2)	Presente = 1, Ausente = 0
S4 — Finalidad (D4)	Presente = 1, Ausente = 0

2.2 Agregación y normalización

El índice es el promedio ponderado de los seis subíndices:

$$IMN = w1 \cdot S1 + w2 \cdot S2 + w3 \cdot S3 + w4 \cdot S4 + w5 \cdot S5 + w6 \cdot S6$$

con los pesos sumando 1. Como cada subíndice es una proporción o una media acotada en 0–100, el IMN es absoluto y comparable entre versiones del conjunto de datos: no depende, como la versión anterior, de un país-ancla. Este procedimiento —marco teórico, normalización, ponderación y agregación, y análisis de robustez— sigue las recomendaciones del manual de la OCDE y el Centro Común de Investigación para la construcción de indicadores compuestos (Nardo et al., 2008).



El bloque de calidad mide *qué tan fuerte* regula cada país, el de amplitud mide *cuántos frentes distintos* toca su regulación. Cada subíndice se expresa como una proporción (entre 0 y 1) que se multiplica por 100 para dejarlo en escala 0–100, igual que los demás.

Ejemplos de cálculo del IMN (caso Ecuador)

El Índice de Madurez Normativa (IMN) promedia seis subíndices (0–100), uno por dimensión del framework, repartidos en dos bloques que se calculan de manera distinta. El bloque de calidad mide qué tan fuerte regula cada país (se calcula norma por norma); el bloque de amplitud mide en cuántos frentes regula (se calcula por cobertura de categorías). A continuación se ilustra cada uno con datos del corpus ecuatoriano.

Bloque de calidad — escala de intensidad por norma

A cada norma se le asigna un valor de intensidad entre 0 y 1, tomando el valor de la categoría más fuerte que presente en esa dimensión (una norma puede tener varias categorías; se conserva la mayor). El subíndice del país es el promedio de esos valores multiplicado por 100. Las escalas son:

Subíndice	Escala de intensidad por norma (categoría más fuerte presente)
S3 Mecanismo	Prohibición / Obligación / Sanción = 1.0 Gobernanza institucional = 0.8 Principio ético = 0.5 Lineamiento / Formación = 0.4 No determinado = 0.2
S5 Enfoque	Estratégico = 1.0 Preventivo = 0.85 Habilitador = 0.75 Reflexivo = 0.55 Restringido = 0.45 Reactivo = 0.25
S6 Madurez	Ecosistema maduro = 1.0 Reg. consolidada = 0.8 Reg. inicial = 0.6 Declaración de intención = 0.4 Exploración = 0.2

Ejemplo ilustrativo (cinco normas reales de Ecuador). La tabla muestra la categoría más fuerte de cada norma en las tres dimensiones de calidad y el valor de intensidad asignado.

Norma (ejemplo)	D3 → valor	D5 → valor	D6 → valor
Lineamientos para uso de IA	Lineamiento → 0.4	Reflexivo → 0.55	Declaración → 0.4
Información sobre tratamiento de datos	Lineamiento → 0.4	Habilitador → 0.75	Declaración → 0.4
Política de Privacidad y Protección de Datos	Obligación → 1.0	Estratégico → 1.0	Reg. inicial → 0.6
Resolución C.U. 1.332-2024 (gobernanza IA)	Obligación+Gobernanza → 1.0	Reflexivo → 0.55	Reg. consolidada → 0.8
Resolución (obligación + sanción)	Obligación+Sanción → 1.0	Reflexivo → 0.55	Reg. consolidada → 0.8



Promedios de la muestra (5 normas):

- $S3 = (0.4 + 0.4 + 1.0 + 1.0 + 1.0) / 5 = 3.8 / 5 = 0.76 \rightarrow 76$
- $S5 = (0.55 + 0.75 + 1.0 + 0.55 + 0.55) / 5 = 3.4 / 5 = 0.68 \rightarrow 68$
- $S6 = (0.4 + 0.4 + 0.6 + 0.8 + 0.8) / 5 = 3.0 / 5 = 0.60 \rightarrow 60$

Este micro-ejemplo ilustra el procedimiento. Calculado sobre las 81 normas del corpus ecuatoriano (y tras el ajuste de credibilidad de Bühlmann que encoge la media hacia la media regional), los valores reales de Ecuador son $S3 = 71.8$, $S5 = 72.4$ y $S6 = 51.8$. La madurez ($S6$) es la dimensión más baja: pese a la actividad regulatoria, la mitad de las normas no supera la “declaración de intención”.

Bloque de amplitud — proporción de cobertura

Aquí la unidad no es la norma sino la categoría. Para cada dimensión se cuenta cuántas categorías están cubiertas por al menos una norma del país y se divide entre el total de categorías de esa dimensión, multiplicando por 100. No interviene ninguna escala de intensidad: cada categoría simplemente está cubierta (1) o ausente (0).

Ejemplo (Ecuador). La tabla detalla qué categorías cubre y cuáles no.

Subíndice	Categorías cubiertas por Ecuador	Ausentes	Cálculo
S1 Actor (de 6)	Estudiantes, Docentes, Investigadores, IES, Gobierno, Todos	— (cubre las 6)	$6/6 \times 100 = 100$
S2 Objeto (de 8)	IA generativa, IA en general, Datos personales, Evaluación automatizada, Sistemas adaptativos, Algoritmos de decisión	Analíticas de aprendizaje, Reconocimiento biométrico	$6/8 \times 100 = 75$
S4 Finalidad (de 8)	Integridad, Evaluación, Personalización, Predicción, Protección de derechos, Innovación pedagógica, Gobernanza y gestión	Monitoreo estudiantil	$7/8 \times 100 = 87.5$

Así, Ecuador obtiene $S1 = 100$, $S2 = 75$ y $S4 = 87.5$. Estos valores son altos porque su corpus, amplio y variado, “toca” casi todas las categorías; por eso la amplitud correlaciona con el volumen de normas y recibe menor peso en el índice.

Combinación final — del subíndice al IMN

El IMN es el promedio ponderado de los seis subíndices. En el escenario de referencia (E2), los pesos son: $S1 = 0.10$, $S2 = 0.10$, $S3 = 0.20$, $S4 = 0.10$, $S5 = 0.20$, $S6 = 0.30$ (la calidad — mecanismo y madurez— pesa más que la amplitud).

Con los valores reales de Ecuador:

$$\text{IMN} = 0.10 \times 100 + 0.10 \times 75 + 0.20 \times 71.8 + 0.10 \times 87.5 + 0.20 \times 72.4 + 0.30 \times 51.8$$

$$\text{IMN} = 10.0 + 7.5 + 14.36 + 8.75 + 14.48 + 15.54 = 70.6$$



El resultado, 70.6, ubica a Ecuador en el nivel Muy Alto (3.º de 28). Nótese cómo la amplitud (S1, S2, S4), pese a sus valores altos, aporta en conjunto solo 26.25 puntos (10.0+7.5+8.75), mientras que el bloque de calidad aporta 44.38 (14.36+14.48+15.54), el diseño del índice garantiza que la fuerza y la madurez de la regulación pesen más que su cobertura.

3. Ponderación y escenarios E1 a E5

La elección de los pesos es, inevitablemente, un juicio de valor, no existe un peso “verdadero” que el dato imponga. Por ello, en lugar de fijar un único conjunto de pesos de forma arbitraria, se definieron cinco escenarios de ponderación, cada uno representando una postura teórica distinta sobre qué significa “madurez normativa”. Calcular el índice bajo los cinco y comparar los resultados es un análisis de sensibilidad, y permite distinguir las conclusiones que son robustas —se sostienen sea cual sea el peso— de las que dependen de una decisión particular. Esta es precisamente la prueba de robustez que recomienda el manual de indicadores compuestos.

¿Qué representa cada escenario? Los cinco vectores de pesos (sobre D1, D2, D3, D4, D5, D6) son:

Escenario	Pesos D1·D2·D3·D4·D5·D6 (%)	Postura que encarna
E1 Equiponderado	16,7 · 16,7 · 16,7 · 16,7 · 16,7 · 16,7	Línea base: sin evidencia para diferenciar, todo pesa igual
E2 Calidad	10 · 10 · 20 · 10 · 20 · 30	Madurez = profundidad + vinculatoriedad + estrategia (referencia)
E3 Vinculatoriedad	8 · 8 · 35 · 8 · 15 · 26	Madurez = capacidad de obligar, sancionar y controlar
E4 Cobertura	20 · 20 · 13 · 20 · 13 · 14	Premia el alcance; en línea con la lógica de volumen
E5 Madurez pura	10 · 10 · 15 · 10 · 15 · 40	Domina la escalera de madurez del ecosistema (D6)

3.1 Resultados de los escenarios

La tabla siguiente muestra el IMN (ya ajustado por credibilidad, sección 4) de una selección de países bajo los cinco escenarios. La lectura no debe centrarse en el valor exacto sino en el comportamiento conjunto: **qué se mantiene y qué se mueve al cambiar los pesos.**



País	n	E1	E2	E3	E4	E5
Brasil	81	77,6	72,2	72,5	79,4	70,1
Chile	105	75,2	72,0	74,4	75,8	69,6
Bahamas	3	62,7	63,8	66,3	61,1	61,3
Perú	29	74,0	70,4	71,2	74,9	68,2
Ecuador	81	76,4	70,6	70,4	78,5	68,6
Belice	3	55,0	59,2	62,6	51,9	56,7
México	179	70,3	65,5	65,5	71,9	63,7
Argentina	100	63,9	61,7	61,9	64,0	59,2
Colombia	36	70,6	67,4	68,4	71,4	65,1
Uruguay	56	58,8	57,3	55,7	58,9	55,4

Lo que es robusto. Brasil y Chile encabezan la región en los cinco escenarios, y México permanece en la zona media en todos ellos (entre 63,7 y 71,9), sin alcanzar nunca la categoría de “faro regional” (si se decantara por el E4, México sería faro regional). Que esta conclusión —la madurez no es lo mismo que la cantidad de normas— se sostenga independientemente de los pesos es la evidencia más fuerte de su validez.

Lo que es sensible. El escenario E4 (cobertura) eleva de forma general los puntajes y favorece a los países grandes y de amplio catálogo —Brasil sube a 79,4; Ecuador a 78,5; México a 71,9— porque los subíndices de amplitud recompensan el alcance, que vuelve a correlacionar con el volumen. Los escenarios E2 y E5, centrados en la calidad y la madurez, comprimen y bajan los puntajes. En otras conclusión, E4 reintroduce parcialmente el sesgo de volumen que el rediseño buscaba eliminar, mientras que E2 y E5 son los más fieles a la idea de madurez. ***Por eso se adopta E2 como escenario de referencia, prioriza la calidad sin descartar la amplitud, y sus resultados son intermedios y estables.***

4. Ajuste de credibilidad e índice de confianza

Los subíndices de calidad son medias por norma. En países con muy pocas normas, una sola norma fuerte puede inflar la media y producir un puesto alto que no refleja madurez sino falta de evidencia. Para corregirlo sin excluir a ningún país —algo especialmente importante en un observatorio que busca visibilizar a los Estados pequeños del Caribe— se aplicó un ajuste de credibilidad: la media de cada país se “encoge” hacia la media regional con un peso proporcional a su evidencia disponible.



Fórmula. Para cada subíndice de calidad,

$$S_{ajustado} = Z \cdot S_{país} + (1 - Z) \cdot S_{regional}$$

con $Z = n / (n + k)$, donde n es el número de normas del país y k una constante de credibilidad. Cuando n es grande, Z se acerca a 1 y prevalece el dato propio; cuando n es pequeño, el puntaje se apoya en la media regional.

Estimación de k . La constante no se fijó a mano, se estimó por credibilidad de Bühlmann (1967).

$$k = EPV/VHM$$

donde EPV es la varianza de las normas dentro de los países (varianza de proceso) y VHM es la varianza entre las medias verdaderas de los países. Los valores obtenidos por dimensión son:

Subíndice	k	Media regional	Lectura
S3 — Mecanismo	7,04	72,4	≈7 normas para credibilidad del 50%
S5 — Enfoque	14,06	73,2	≈14 normas para credibilidad del 50%
S6 — Madurez	13,25	51,1	≈13 normas para credibilidad del 50%

El efecto del ajuste es el esperado, los casos inflados por muestras diminutas descienden de forma proporcional a su falta de evidencia, mientras que los países con muchas normas apenas se mueven en puntaje pero ascienden en el ordenamiento al desinflarse los pequeños. Los subíndices de amplitud no se ajustan, pues no se inflan con n pequeño sino que, al contrario, penalizan a los países con pocas normas.

Índice de confianza. De forma complementaria, cada país recibe una etiqueta de confianza según su número de normas:

Alta ($n \geq 40$)
Media ($15 \leq n < 40$)
Provisional ($n < 15$)

De esta forma, el lector ve simultáneamente el puntaje robusto y cuánta evidencia lo respalda. El resultado es un ranking que mide madurez, justifica cada peso, descansa sobre datos verificados y es robusto frente a muestras pequeñas.

5. Limitaciones y reproducibilidad

La clasificación depende de la calidad y exhaustividad del levantamiento, una norma cuyo texto en la matriz es escueto puede subclasificarse, motivo por el cual se mantiene la etiqueta de confianza y la hoja de trazabilidad. Las escalas de intensidad y los vectores de pesos son decisiones de diseño explícitas y, por tanto, discutibles; el análisis de escenarios existe



precisamente para acotar su influencia. Finalmente, el ajuste de credibilidad encoge hacia la media regional, lo que puede atenuar la excelencia genuina de un país pequeño; la etiqueta de confianza señala estos casos para su lectura cuidadosa. Todos los parámetros —escalas, pesos, constantes k y umbrales— están documentados en este texto y en las hojas de cálculo del Observatorio, de modo que el procedimiento es íntegramente reproducible.

6. Referencias

- Bühlmann, H. (1967). Experience rating and credibility. *ASTIN Bulletin: The Journal of the IAA*, 4(3), 199-207.
- Chatti, M. A., Dyckhoff, A. L., Schroeder, U., & Thüs, H. (2012). A reference model for learning analytics. *International Journal of Technology Enhanced Learning*, 4(5/6), 318–331.
- Nardo, M., Saisana, M., Saltelli, A., Tarantola, S., Hoffman, A., & Giovannini, E. (2008). *Handbook on Constructing Composite Indicators: Methodology and User Guide*. OECD Publishing, Paris.
- Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education — where are the educators? *International Journal of Educational Technology in Higher Education*, 16(1), 39.
- Selwyn, N. (2022). The future of AI and education: Some cautionary notes. *European Journal of Education*, 57(4), 620–631.



Co-funded by
the European Union

genial

Université
de Toulouse

uc3m

Universidad
Carlos III
de Madrid



Universidad
del Cauca

UNIVERSIDAD
EAFIT

 *Galileo*
UNIVERSIDAD

 **USAC**
TRICENTENARIA
Universidad de San Carlos de Guatemala

UCUENCA



UNA E

intec



PUCMM
Pontificia Universidad Católica
México y Maestros