

CAPÍTULO 6

GOBERNANZA, ÉTICA Y REGULACIÓN DE LA IA EN LA EDUCACIÓN SUPERIOR

José Wilson Gómez Cumpa

Universidad Nacional Pedro Ruiz Gallo, Lambayeque. Perú.

<https://orcid.org/0000-0001-7071-6248>

correo: jgomezc@unprg.edu.pe

María Margarita Fanning Balarezo

Universidad Nacional Pedro Ruiz Gallo, Lambayeque. Perú.

<https://orcid.org/0000-0002-2666-4935>

correo: mafabbibg@unprg.edu.pe

Carol Beatriz Bao Ratzemberg

Universidad Nacional de San Martín, Tarapoto. Perú.

<https://orcid.org/0000-0003-4205-7446>

correo: cbaoratzemberg@unsm.edu.pe

6.1 INTRODUCCIÓN

La integración de la inteligencia artificial (IA) en la educación superior ha transitado, en pocos años, de la incorporación de herramientas auxiliares hacia la configuración de una infraestructura que reconfigura procesos académicos centrales: admisión, evaluación, acompañamiento estudiantil, gestión curricular y certificación de competencias. Este desplazamiento no es meramente tecnológico; es institucional. Cuando los sistemas de IA producen recomendaciones que influyen sobre trayectorias formativas, asignan alertas tempranas o retroalimentan tareas de manera automatizada, la universidad ya no administra únicamente tecnología:

administra consecuencias sobre derechos, oportunidades y legitimidad académica.

En este nuevo escenario, la pregunta decisiva deja de ser "¿conviene usar IA?" para convertirse en "¿bajo qué arquitectura de gobernanza se integra?", "¿qué evidencia respalda su uso?" y "¿cómo se protegen los derechos de quienes pueden verse afectados por decisiones algorítmicas?". La gobernanza, la ética y la regulación no son capas adicionales sobre la tecnología: son condiciones de posibilidad para que la IA sea educativamente legítima, jurídicamente defendible y socialmente confiable.

Este capítulo propone un marco analítico y operativo para que las instituciones de educación superior puedan diseñar, evaluar y gobernar sus ecosistemas de IA. Se organiza en torno a cuatro ejes articulados: la gobernanza académica en entornos inteligentes, la ética y la rendición de cuentas como ciclo continuo de legitimidad institucional, la regulación y las políticas públicas con atención especial al contexto latinoamericano, y la autonomía universitaria como condición estratégica del uso responsable de IA.

6.2 Marco teórico y fundamentación conceptual

El análisis que desarrolla este capítulo se apoya en tres marcos normativos internacionales complementarios que constituyen el estado del arte en gobernanza de sistemas de IA. El primero es el Marco de Gestión de Riesgos de Inteligencia Artificial del NIST (AI RMF 1.0), que propone un ciclo de gestión organizado en cuatro funciones interrelacionadas: gobierno, mapeo, medición y gestión, aplicables a lo largo del ciclo de vida completo del sistema (NIST, 2023). El segundo es la norma ISO/IEC 42001:2023, que establece los requisitos para un sistema de gestión de inteligencia artificial aplicable a organizaciones que desarrollan o utilizan sistemas de IA, complementada por la ISO/IEC 23894:2023, orientada específicamente a la gestión de riesgos de IA (ISO/IEC, 2023a; ISO/IEC, 2023b). El tercero es el conjunto de Principios sobre IA de la OCDE, que enfatiza la transparencia, la explicabilidad y la rendición de cuentas como condiciones de la IA confiable, y que ha sido adoptado como referencia por más de cincuenta gobiernos (OECD, 2023).

Sobre esta base normativa se articulan tres conceptos centrales que estructuran el análisis. El primero es la gobernanza por capas: concibe los

sistemas de IA como arquitecturas compuestas —datos, modelos, reglas de decisión, interfaces, auditoría— en las que cada capa tiene responsables, controles y evidencias trazables propias. Este enfoque supera la lógica de autorizar o rechazar herramientas para abrazar una lógica de gestión de portafolio institucional. El segundo concepto es la proporcionalidad del riesgo, principio según el cual las obligaciones de gobernanza se escalan en función del nivel de consecuencia potencial del sistema sobre derechos y oportunidades de las personas; en el dominio educativo, esto significa que los sistemas que determinan admisiones o certificaciones exigen un estándar de control radicalmente distinto al de un asistente de búsqueda bibliográfica. El tercer concepto es la rendición de cuentas como ciclo continuo (*accountability loop*), que convierte incidentes y resultados de auditoría en insumos para la mejora de política, cerrando el ciclo con aprendizaje institucional verificable (NIST, 2023; ISO/IEC, 2023a; OECD, 2023).

La literatura especializada en el campo advierte que estos marcos normativos son condición necesaria pero no suficiente para la gobernanza efectiva. La brecha entre principios éticos declarados y prácticas institucionales verificables constituye el principal desafío de la gobernanza ética contemporánea (Holmes et al., 2021; UNESCO, 2021). En el contexto latinoamericano e iberoamericano, esta brecha se amplifica por condiciones estructurales específicas: inversión insuficiente en capacidades de evaluación de riesgos, escasa regulación vinculante de nivel sectorial en la mayoría de los países y dependencia tecnológica respecto a proveedores externos con limitada adaptabilidad a los contextos lingüísticos y culturales de la región (Medina-Zuta et al., 2025; Levy-Yeyati, 2025).

Dos tensiones conceptuales merecen atención especial como punto de partida del análisis. La primera es la tensión entre transparencia y privacidad: abrir el libro de un sistema de IA puede revelar datos personales sensibles o estrategias de modelado que, conocidas, permiten manipulaciones adversariales. La gobernanza ética debe diseñar mecanismos de transparencia contextualmente pertinentes —suficientes para legitimar decisiones sin exponer datos protegidos— y no puede resolverse mediante fórmulas universales. La segunda tensión es entre precisión predictiva y explicabilidad: los modelos con mayor precisión estadística suelen ser los menos interpretables para actores no técnicos.

En contextos educativos donde las decisiones afectan trayectorias personales, la prioridad institucional debería inclinarse hacia sistemas interpretables, porque la legitimidad académica depende de que las decisiones puedan ser explicadas, entendidas y contestadas por quienes las reciben (Holmes et al., 2021).

6.3 METODOLOGÍA

Dado el carácter normativo y propositivo del capítulo, se adoptó una metodología de revisión comparada de marcos regulatorios internacionales y análisis documental de políticas institucionales emergentes (2023–2026). Se triangularon fuentes primarias (AI Act 2024, NIST AI RMF 1.0, ISO/IEC 42001/23894 y UNESCO 2021) con informes sectoriales y técnicos (OCDE, UNESCO IESALC, TEQSA y BID) y literatura académica sobre gobernanza ética de IA en educación superior. El procedimiento permitió construir un marco operativo de gobernanza por capas, escalable según madurez institucional y contextualizado para las condiciones estructurales de América Latina y el Sur Global. Al tratarse de una revisión documental sin intervención con personas, animales o plantas, no requirió aprobación de comité de ética; no obstante, se aplicaron criterios de trazabilidad, citación responsable y uso de fuentes verificables.

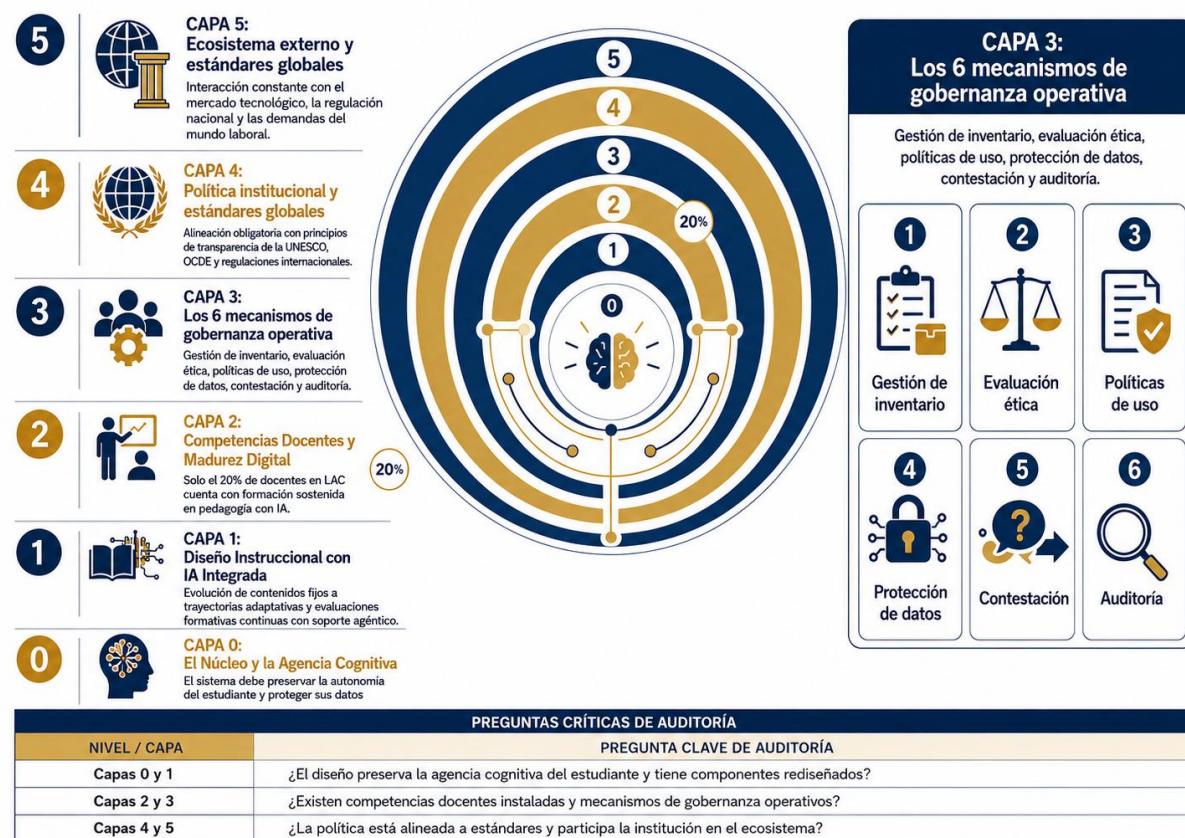
6.4 Gobernanza académica en entornos inteligentes

6.4.1 Del control de herramientas al gobierno de sistemas

La gobernanza de la IA en la universidad no puede reducirse a la autorización de aplicaciones específicas. Las arquitecturas tecnológicas actuales son sistemas compuestos: datos institucionales, modelos predictivos o generativos, reglas de decisión, interfaces de interacción y capas de automatización conviven en un mismo flujo que produce recomendaciones, alertas y decisiones con impacto sobre los estudiantes. Gobernar estos sistemas exige una lógica de capas —datos, integración, modelos, reglas de decisión, interfaces, auditoría— en la que cada capa tiene responsables, puntos de control y evidencias trazables (NIST, 2023; ISO/IEC, 2023a). La Figura 6.1 representa esta arquitectura por capas y sus controles transversales.

Figura 6.1

Mapa de gobernanza: arquitectura formativa inteligente por capas



Nota. Elaboración propia a partir de marcos de gestión y riesgo para sistemas de IA (NIST, 2023; ISO/IEC, 2023a, 2023b) y orientaciones de gobernanza de datos y tecnología en ecosistemas educativos (OECD, 2023). La figura muestra el flujo desde la capa de datos hasta la capa de auditoría, con el arco de gobernanza transversal y el piso de derechos y bienestar.

La transición desde una lógica de "proyectos" aislados hacia una gestión de "portafolio" institucional resulta indispensable. Una universidad no adopta IA una única vez; construye, a lo largo del tiempo, un conjunto creciente de casos de uso que incluyen herramientas propias, plataformas de terceros, integraciones con sistemas de gestión académica y aplicaciones de uso informal. Sin un inventario institucional activo, clasificado por criticidad y con responsables asignados, la gobernanza se vuelve ilusoria: los riesgos se multiplican sin control y la rendición de cuentas queda fragmentada (OECD, 2023).

6.4.2 La irrupción de la IA agéntica y sus nuevas exigencias

El ecosistema universitario ha comenzado a incorporar sistemas de IA agéntica: modelos capaces de razonar, planificar subobjetivos y ejecutar

acciones con relativa autonomía, como gestionar alertas de deserción, responder consultas complejas de estudiantes o participar en la revisión preliminar de trabajos. Esta autonomía extendida intensifica las demandas de gobernanza en tres dimensiones críticas.

Trazabilidad de acciones agénticas. Cuando un agente de IA opera en nombre de la institución, sus acciones deben ser atribuibles: cada intervención —un cambio de registro, una notificación, una recomendación— debe quedar vinculada a un agente identificado, con registros de auditoría que permitan reconstruir decisiones ante cualquier reclamación o revisión.

Delimitación del espacio de acción. Antes del despliegue, la institución debe definir explícitamente qué puede y qué no puede hacer el sistema, restringiendo su acceso a recursos y datos según el principio de privilegio mínimo. Esta delimitación no es solo técnica; es una decisión de política institucional con consecuencias sobre derechos.

Puntos de control humano obligatorio. En acciones irreversibles —emisión de calificaciones definitivas, actuaciones disciplinarias, comunicaciones con consecuencias legales o académicas— la aprobación humana debe ser condición sine qua non, no una opción configurable que el proveedor puede desactivar.

6.4.3 Estructuras institucionales para la gobernanza operativa

La gobernanza se materializa en estructuras organizacionales con capacidad real de decisión. La experiencia comparada sugiere un diseño mínimo con separación de funciones para evitar conflictos entre quienes promueven el despliegue y quienes verifican el riesgo:

1. Un comité de IA educativa con mandato para autorizar, condicionar, restringir o suspender casos de uso según criticidad y riesgo, con criterios explícitos de decisión.
2. Roles diferenciados: patrocinio rectoral para coherencia estratégica; responsables de caso de uso en unidades académicas; gestores de datos para calidad, acceso y retención; responsables técnicos para desempeño y seguridad; asesoría jurídica; y auditoría interna con función independiente.
3. Un liderazgo pedagógico con capacidad efectiva de decisión sobre criterios de uso, evaluación y salvaguardas de integridad académica, porque la principal amenaza en arquitecturas formativas no siempre

es técnica: frecuentemente es la degradación silenciosa de la calidad evaluativa (UNESCO, 2021).

Las políticas institucionales deben operar como sistema integrado, no como colección de documentos sin articulación: política de uso de IA en docencia y evaluación; gobernanza de datos educativos; gestión de proveedores y tecnología; seguridad informática; y rendición de cuentas con garantías institucionales explícitas (ISO/IEC, 2023a; ISO/IEC, 2023b).

6.5 Ética, transparencia y rendición de cuentas

6.5.1 Principios éticos en tensión: del catálogo a la operación

La literatura especializada converge en un núcleo de principios para la IA educativa: justicia y no discriminación, transparencia y explicabilidad, privacidad y protección de datos, responsabilidad con control humano significativo, y protección especial de grupos vulnerables (Holmes et al., 2021; UNESCO, 2021). Sin embargo, la experiencia institucional advierte que este catálogo resulta insuficiente cuando permanece "solo en papel". La brecha de implementación entre principios declarados y prácticas verificables constituye el principal desafío de la gobernanza ética contemporánea.

6.5.2 El ciclo de rendición de cuentas: de la declaración a la práctica

La rendición de cuentas no es un acto puntual de transparencia, sino una capacidad organizacional sostenida en el tiempo. La Tabla 6.1 sintetiza seis etapas de un ciclo operativo de rendición de cuentas para la IA educativa y las evidencias mínimas requeridas en cada una.

Tabla 6.1

Ciclo de rendición de cuentas para IA educativa

Estación del ciclo	Pregunta de control	Evidencia mínima	Responsable principal
1. Inventario y propósito	¿Qué sistema de IA se usa y para qué finalidad institucional?	Ficha de caso de uso, finalidad, responsable asignado e inventario actualizado.	Unidad académica y comité de IA educativa
2. Evaluación previa	¿Qué derechos, datos o trayectorias pueden verse afectados?	Matriz de criticidad, evaluación de impacto y criterios de proporcionalidad.	Comité de IA, asesoría jurídica y responsable técnico

Estación del ciclo	Pregunta de control	Evidencia mínima	Responsable principal
3. Documentación y transparencia	¿Cómo se informa el uso y cómo se registra el funcionamiento?	Política de uso, registro de datos/modelo, aviso a usuarios y trazabilidad básica.	Responsable del sistema y gestor de datos
4. Supervisión y monitoreo	¿Cómo se verifica desempeño, sesgos e incidentes durante el uso?	Indicadores, bitácoras, reportes de incidentes y auditorías periódicas.	Equipo técnico y auditoría interna
5. Contestación y reparación	¿Cómo puede una persona afectada impugnar una salida algorítmica?	Procedimiento de revisión, instancia humana, plazos y mecanismo de reparación.	Unidad académica, defensoría o asesoría jurídica
6. Mejora continua	¿Cómo se actualizan las políticas a partir de incidentes y auditorías?	Informe de mejora, actualización normativa y ajuste de modelos o procesos.	Comité de IA educativa y alta dirección

Nota. Elaboración propia con base en NIST (2023), ISO/IEC (2023a), OECD (2023), UNESCO (2021) y Holmes et al. (2021).

6.5.3 Integridad académica en la era de la IA generativa

La IA generativa ha alterado de forma estructural el equilibrio entre lo que se evalúa (aprendizajes), lo que se produce (evidencias) y lo que se infiere (logro y autoría). Cuando un sistema puede redactar, resumir o resolver problemas con resultados plausibles, la integridad académica deja de ser un asunto periférico de disciplina estudiantil para convertirse en un componente de calidad institucional: determina si la universidad puede sostener credenciales confiables y trayectorias formativas legítimas (Miao & Holmes, 2023).

La respuesta institucional más sólida no radica en la vigilancia reactiva sino en el "diseño de confianza": rediseñar la evaluación de modo que el aprendizaje sea observable independientemente del uso de IA. Esto implica avanzar hacia la evaluación auténtica y por proceso —hitos intermedios, versiones, defensas orales breves, bitácoras reflexivas, prototipos— donde el estudiante demuestra decisiones, justificaciones e iteraciones, no solo entrega un resultado final. Implica también distinguir entre tareas donde la IA puede ser un colaborador formativo legítimo y tareas donde la institución necesita evidencia directa de competencia, por su criticidad en términos de progresión, licenciamiento o acreditación.

El modelo de las "dos vías" (ver Figura 6.2) de evaluación propone combinar evaluaciones seguras (presenciales, supervisadas, sin mediación tecnológica no autorizada) con evaluaciones abiertas y auténticas que integran la IA como herramienta de ideación, con declaración obligatoria de uso y reflexión sobre el proceso de co-creación. Cada titulación necesita un plan de evaluación que equilibre ambas vías, actuando como el artefacto de gobernanza principal para acreditación y aseguramiento de calidad (NIST, 2023; TEQSA, 2024).

Figura 6.2

Modelo de las dos vías de evaluación en la era de la IA generativa



Nota. Elaboración propia a partir de marcos de integridad académica y gobernanza de IA (NIST, 2023; TEQSA, 2024). La figura ilustra la complementariedad entre evaluaciones seguras y evaluaciones abiertas con IA, con criterios de decisión para cada vía.

Respecto al uso de detectores automáticos de texto generado por IA como base de sanción, la evidencia disponible indica que estos sistemas presentan limitaciones severas: pueden producir falsos positivos que afectan de forma desigual a ciertos grupos, particularmente a hablantes no nativos de inglés o a estudiantes con estilos de escritura más directos.

Su uso como único fundamento sancionador es jurídicamente insostenible. Deberían emplearse, en todo caso, únicamente como señal preliminar dentro de un debido proceso que incluya evidencia adicional, oportunidad de explicación, revisión humana y mecanismos de apelación (Scarfe et al., 2024).

6.6 Regulación institucional y políticas públicas

6.6.1 El paisaje regulatorio internacional

El entorno normativo global sobre IA ha evolucionado con rapidez. El Reglamento de Inteligencia Artificial de la Unión Europea (AI Act, 2024) constituye el instrumento vinculante más comprehensivo: clasifica los usos de IA según niveles de riesgo e impone obligaciones diferenciadas. En el ámbito educativo, clasifica como de alto riesgo los sistemas que determinan el acceso a instituciones o programas, evalúan resultados de aprendizaje, estiman niveles educativos y detectan conductas prohibidas durante pruebas. Esta clasificación implica obligaciones sustantivas de documentación técnica, supervisión humana, trazabilidad, evaluación de conformidad y mecanismos de recurso para las personas afectadas (European Union, 2024).

Otros marcos de referencia relevantes incluyen el AI Risk Management Framework del NIST (2023), los Principios de la OCDE sobre IA y las recomendaciones de la UNESCO, que consideran la educación un ámbito en el que la IA puede afectar derechos fundamentales y, por ello, requiere salvaguardas reforzadas. La Tabla 6.2 compara las implicaciones de estos marcos para la educación superior (UNESCO, 2021; OECD, 2023).

Tabla 6.2

Comparación del marco regulatorio de IA con implicaciones para la educación superior: Unión Europea, Estados Unidos y América Latina y el Caribe

Marco o jurisdicción	Enfoque dominante	Implicaciones para IES	Tipo de exigencia
Unión Europea: AI Act 2024	Regulación vinculante basada en riesgo; identifica usos educativos de alto riesgo.	Obliga a documentar, supervisar, auditar y habilitar mecanismos de recurso cuando la IA afecta acceso, evaluación o trayectoria.	Alto nivel de obligatoriedad

Marco o jurisdicción	Enfoque dominante	Implicaciones para IES	Tipo de exigencia
Estados Unidos: NIST AI RMF y marcos sectoriales	Gestión de riesgos, buenas prácticas y responsabilidad organizacional a lo largo del ciclo de vida.	Ofrece un esquema operativo para gobernar, mapear, medir y gestionar sistemas de IA aun sin una ley federal única equivalente al AI Act.	Guía técnica y estándar de referencia
América Latina y el Caribe	Entorno heterogéneo y dinámico; predominan iniciativas preliminares, lineamientos generales y regulación parcial de datos.	Requiere registros de sistemas, cláusulas de transparencia, capacidades de auditoría, protección de datos y cooperación interuniversitaria.	Implementación escalonada y contextual

Nota. Elaboración propia con base en European Union (2024), NIST (2023), OECD (2023), UNESCO (2021), Medina-Zuta et al. (2025) y Levy-Yeyati (2025).

6.6.2 La brecha de implementación: entre el principio y la práctica

La comparación internacional revela un patrón sistemático: la proliferación de principios declarados no se acompaña de mecanismos de implementación y verificación equivalentes. Una encuesta de 2025 reportó que solo el 19% de las instituciones de educación superior participantes tenía políticas formales de IA, mientras que el 42% las estaba elaborando. En América Latina y el Caribe, únicamente el 45% de las instituciones declaraba contar con orientaciones formales sobre IA, frente al 70% en Europa y América del Norte (UNESCO IESALC, 2025).

Esta brecha tiene consecuencias concretas: cuando los sistemas de IA operan sin marcos institucionales claros, la gestión de riesgos se delega de facto a docentes y estudiantes individuales, generando lo que la literatura denomina "trabajo invisible" y "responsabilidad individualizada del fracaso". Las políticas ambiguas o ausentes desplazan la carga de la gobernanza hacia los eslabones más débiles de la cadena institucional.

Un problema emergente y creciente es el de la "IA en la sombra" (shadow AI): el uso no sancionado de herramientas de IA por parte de docentes,

estudiantes y personal administrativo, que se expande con mayor rapidez que los marcos de gobernanza formales. Esta adopción informal crea riesgos en protección de datos, integridad académica y responsabilidad institucional que las estructuras de gobernanza deben abordar mediante la provisión de alternativas sancionadas, privadas por diseño, y mediante políticas claras de uso aceptable (Autenrieth & Schluchter, 2025).

6.6.3 Regulación basada en riesgo: criterios de gradación institucional

La regulación basada en riesgo —el enfoque que escala las obligaciones según el nivel de consecuencia potencial— ofrece el marco más útil para las universidades, porque permite priorizar recursos de gobernanza en los usos de mayor impacto sobre derechos y oportunidades. La Tabla 6.3 sistematiza las tres zonas de exigencia diferenciada.

Tabla 6.3

Clasificación de casos de uso de LA en educación superior por criticidad y exigencias de gobernanza

Zona de riesgo	Ejemplos de casos de uso	Exigencias mínimas de gobernanza	Marco de referencia
Máxima exigencia (alto riesgo)	Admisión/selección, evaluación certificatoria, decisiones de trayectoria, detección de conducta prohibida	Evaluación de impacto previa, documentación completa, auditoría independiente, supervisión humana reforzada, mecanismos de contestación y reparación	AI Act UE Art. 6; NIST AI RMF; ISO/IEC 42001
Exigencia estándar (riesgo moderado)	Tutorías adaptativas, retroalimentación formativa, alertas de riesgo académico no vinculantes, sistemas de recomendación	Inventario documentado, evaluación previa básica, supervisión humana, monitoreo de sesgos y efectos pedagógicos	NIST AI RMF; OECD 2023; UNESCO 2021
Bajo riesgo (control básico)	Asistentes administrativos de consulta, búsqueda semántica en repositorios, corrección ortográfica/gramatical asistida	Inventario activo, acceso autorizado, registro de uso, política de uso aceptable	ISO/IEC 23894; buenas prácticas institucionales

Nota. Elaboración propia con base en European Union (2024), NIST (2023), ISO/IEC (2023a) y OECD (2023). La clasificación es orientativa; cada institución debe adaptarla a su contexto normativo nacional y capacidad institucional.

6.6.4 Contexto latinoamericano e iberoamericano

El paisaje regulatorio en América Latina y el Caribe es heterogéneo y dinámico. Un análisis comparativo iberoamericano señala que Chile y España muestran avances más robustos, mientras que Argentina y Perú permanecen en fases preliminares con iniciativas poco articuladas (Medina-Zuta et al., 2025). En ninguno de los casos la regulación específica sobre IA en educación superior alcanza el nivel de concreción y obligatoriedad del AI Act europeo.

Para el contexto latinoamericano, la literatura propone un enfoque regulatorio escalonado: como medidas inmediatas, el establecimiento de registros de sistemas de IA en el sector público, cláusulas de transparencia en contratos de adquisición y mecanismos iniciales de reporte de incidentes; como medidas de mediano plazo, regímenes de certificación, requisitos de auditoría algorítmica y unidades especializadas de evaluación con capacidades técnicas instaladas (Levy-Yeyati, 2025).

6.7 Autonomía universitaria y uso responsable de IA

6.7.1 La autonomía en riesgo: dependencia tecnológica y soberanía de datos

La autonomía universitaria —históricamente entendida como independencia respecto al poder político— enfrenta hoy un nuevo vector de erosión: la dependencia tecnológica respecto a proveedores privados de sistemas de IA. Cuando una universidad adopta plataformas cerradas sin cláusulas de portabilidad, auditoría ni control de versiones, delega implícitamente decisiones sobre su propia arquitectura formativa. Los cambios de modelo, las actualizaciones de política del proveedor y los ajustes en los términos de servicio pueden alterar el comportamiento del sistema sin que la institución tenga control efectivo ni conocimiento oportuno.

La soberanía de datos es la dimensión más concreta de esta autonomía. Los datos de aprendizaje, desempeño, progresión e investigación son activos estratégicos: su valor depende de reglas e instituciones que permitan su uso para bienes públicos educativos y eviten concentraciones asimétricas de poder informacional. En sistemas de IA generativa, el riesgo es particular: las interacciones cotidianas de docentes y estudiantes con herramientas externas pueden alimentar el entrenamiento de modelos de terceros o filtrar información sensible a

través de prompts, comprometiendo privacidad e integridad académica (UNCTAD, 2021).

6.7.2 Criterios de adopción responsable

Una adopción responsable exige que la institución pueda responder afirmativamente a cinco preguntas antes de desplegar cualquier sistema de IA con impacto en decisiones educativas relevantes:

1. ¿Propósito legítimo? ¿Existe un propósito pedagógico o institucional explícito, delimitado y verificable?
2. ¿Evaluación previa? ¿Se ha realizado una evaluación de riesgos proporcional a la criticidad del caso de uso?
3. ¿Datos mínimos? ¿Los datos utilizados son los mínimos necesarios, gestionados con controles de privacidad y seguridad adecuados?
4. ¿Supervisión y contestación? ¿Existen mecanismos de supervisión humana, trazabilidad y contestación para las personas afectadas?
5. ¿Capacidad de control? ¿Puede la institución auditar, corregir y, si es necesario, suspender el uso del sistema?

Si alguna de estas preguntas no tiene respuesta afirmativa demostrable, la adopción debería condicionarse o diferirse hasta resolver el déficit de gobernanza correspondiente (NIST, 2023; OECD, 2023).

6.7.3 Gestión de proveedores como componente de gobernanza

La mayoría de las universidades incorporará IA mediante proveedores comerciales y plataformas de terceros. Por ello, la gobernanza operativa debe traducirse en debida diligencia y cláusulas contractuales robustas. La Tabla 6.4 presenta los requisitos mínimos para la contratación responsable de proveedores de IA educativa.

Tabla 6.4

Matriz de requisitos mínimos para la contratación responsable de proveedores de IA educativa

Dimensión contractual	Cláusulas y requisitos mínimos	Riesgo que mitiga
Documentación técnica y de uso	Descripción de capacidades y limitaciones; control de versiones; notificación anticipada de cambios sustantivos al comportamiento del sistema	Pérdida de trazabilidad; decisiones incontroladas por cambios externos

Dimensión contractual	Cláusulas y requisitos mínimos	Riesgo que mitiga
Tratamiento de datos	Finalidad limitada; prohibición de usos secundarios no autorizados; retención acotada; medidas de seguridad verificables; identificación de subprocesadores	Filtración de datos académicos; uso de datos para entrenamiento sin consentimiento
Derechos de auditoría	Acceso a registros y evidencias suficientes para demostrar cumplimiento normativo e institucional	Opacidad operativa; imposibilidad de rendir cuentas ante auditores o reguladores
Notificación de incidentes	Plazos máximos de notificación (24-72 h); canales y responsables definidos; categorización de severidad	Exposición institucional tardía ante brechas de seguridad o fallos sistémicos
Portabilidad y reversibilidad	Condiciones para recuperar datos y migrar a sistemas alternativos sin penalizaciones desproporcionadas; formatos abiertos o estándar	Lock-in tecnológico; pérdida de soberanía sobre datos académicos estratégicos

Nota. Elaboración propia con base en NIST (2023, 2024), OECD (2023), ISO/IEC (2023a) y Miao y Holmes (2023).

6.7.4 Modelos de pesos abiertos y cooperación interuniversitaria

Una estrategia de autonomía emergente es la preferencia por modelos de "pesos abiertos" (*open weights*): arquitecturas de IA cuyos parámetros pueden ser inspeccionados, ajustados localmente y desplegados sin dependencia de servicios en la nube de terceros. Este enfoque permite mantener control sobre los datos de entrenamiento, evitar el bloqueo tecnológico (lock-in) y adaptar el comportamiento del modelo a contextos lingüísticos, disciplinarios y culturales propios.

Complementariamente, la cooperación interuniversitaria ofrece una vía para construir escala y capacidad sin centralizar decisiones pedagógicas. Los consorcios regionales pueden compartir infraestructura de cumplimiento normativo, plantillas de evaluación previa, protocolos de incidentes, repositorios de casos de uso y capacidades de negociación colectiva frente a proveedores. Esta lógica de "economías de escala en gobernanza" resulta especialmente pertinente para sistemas universitarios del Sur Global, donde los recursos individuales de cada institución son insuficientes para sostener estructuras completas de gestión de riesgos (UNCTAD, 2024).

6.8 Implicaciones estratégicas para América Latina y el Sur Global

6.8.1 Capacidades institucionales como cuello de botella crítico

En el Sur Global, la principal brecha para una gobernanza efectiva de la IA educativa no es únicamente de infraestructura digital, sino de capacidad institucional para operar sistemas sociotécnicos con evidencia. Esto incluye habilidades para inventariar casos de uso, gobernar datos con criterios de calidad y privacidad, evaluar riesgos antes del despliegue, negociar contratos con cláusulas de auditoría, supervisar el desempeño de modelos y sostener mecanismos de contestación y reparación. Sin estas capacidades, la adopción se vuelve reactiva: las herramientas se incorporan por presión competitiva, sin trazabilidad que permita distinguir valor pedagógico de efectos adversos.

Los datos disponibles ilustran la magnitud del desafío. En Ecuador, uno de los países con evidencia institucional más detallada en el contexto andino, solo el 10% de las instituciones de educación superior encuestadas disponía de una directiva escrita o política formal que regulara el uso de IA en actividades académicas, y ninguna contaba con un código de ética de IA independiente. La gobernanza era, en la mayoría de los casos, más reactiva que proactiva, con consideraciones de IA incluidas tangencialmente en políticas de TIC o códigos éticos generales (Kosbar, 2026; UNESCO IESALC, 2025). La Tabla 6.5 propone indicadores para evaluar y comparar la madurez institucional.

6.8.2 El riesgo del colonialismo digital y las alternativas cooperativas

La dependencia de modelos y plataformas desarrollados principalmente en el Norte Global entraña riesgos específicos para contextos latinoamericanos: sesgos lingüísticos en modelos entrenados con corpus mayoritariamente anglosajones, representaciones culturales inadecuadas para la realidad regional, ausencia de datos históricos relevantes para los contextos locales y asimetrías en los flujos transfronterizos de datos académicos. Sin gobernanza deliberada, la adopción de IA puede reproducir y amplificar desigualdades estructurales preexistentes.

La Figura 6.3 sitúa estas capacidades en una hoja de ruta gradual, con puertas de control verificables entre fases. Una respuesta estratégica implica tres líneas complementarias: la inversión en capacidades técnicas locales (formación de equipos con competencias en evaluación de riesgos, auditoría de algoritmos y gobernanza de datos), la preferencia por

soluciones con mayor transparencia y portabilidad —incluidos modelos de código abierto adaptables a lenguas y contextos regionales— y la construcción de bienes públicos digitales compartidos: infraestructuras de cómputo, corpus de entrenamiento en español y lenguas originarias, y marcos de evaluación pedagógica contextualmente validados (UNESCO, 2021; UNCTAD, 2021).

Figura 6.3

Hoja de ruta de madurez institucional para la gobernanza de IA en educación superior



Nota. Elaboración propia a partir de NIST (2023), ISO/IEC (2023a, 2023b) y OECD (2023). La hoja de ruta propone cinco fases secuenciales y puertas de control institucionales verificables.

6.8.3 Una adopción secuenciada por madurez institucional

La propuesta para el contexto latinoamericano no es ralentizar la adopción, sino secuenciarla con criterios de madurez institucional: comenzar donde la institución puede controlar evidencia y aprender, y escalar solo cuando existan capacidades para gobernar el sistema, sostener auditorías y responder incidentes. La Tabla 6.5 operacionaliza esta progresión mediante indicadores de madurez. En términos prácticos, esto implica:

1. Priorizar usos de alto valor pedagógico y bajo riesgo —tutorías de apoyo, retroalimentación formativa, acceso a recursos— antes de introducir IA en decisiones de alta consecuencia.
2. Construir el inventario de casos de uso y la estructura de gobernanza mínima antes de escalar, evitando la expansión funcional sin controles.
3. Incorporar la evaluación de impacto en equidad como requisito antes de desplegar cualquier sistema que afecte acceso o certificación.
4. Diseñar mecanismos de contestación accesibles desde el inicio del despliegue, no como corrección posterior una vez que se han producido daños.

Tabla 6.5

Indicadores de madurez institucional para gobernanza de IA educativa

Dimensión	Fórmula / Indicador	Fuente de evidencia	Periodicidad	Nivel de madurez
Gobernanza y trazabilidad	% de sistemas IA con documentación de propósito, responsable asignado y registro de decisiones automatizadas	Inventario institucional de sistemas IA; actas de comité de ética digital	Anual	● Básico: Existe registro parcial de herramientas sin responsable asignado ● Intermedio: Inventario completo; política aprobada; responsable designado por unidad ● Avanzado: Trazabilidad total; auditorías periódicas; publicación de registros
Valor pedagógico	% de sistemas IA evaluados con evidencia de impacto en resultados de aprendizaje antes del despliegue institucional	Informes de evaluación piloto; analítica de aprendizaje; encuestas docentes y estudiantiles	Semestra 1	● Básico: Adopción sin evaluación previa; decisiones basadas en preferencia individual ● Intermedio: Protocolos de evaluación piloto definidos; al menos 50% de sistemas evaluados ● Avanzado: Evaluación sistemática con evidencia de impacto; ciclo continuo de mejora

Dimensión	Fórmula / Indicador	Fuente de evidencia	Periodicidad	Nivel de madurez
Justicia y equidad	Índice de disparidad en tasas de acceso, rendimiento y experiencia entre subgrupos estudiantiles en entornos con IA	Datos de analítica desagregados por género, origen socioeconómico, condición de discapacidad y zona geográfica	Anual	● Básico: Sin desagregación de datos; no se monitorea equidad de resultados ● Intermedio: Desagregación disponible; al menos una auditoría de sesgo realizada ● Avanzado: Monitoreo continuo; planes de acción correctiva; resultados publicados
Integridad y robustez evaluativa	% de programas con arquitectura evaluativa rediseñada para contextos con IA generativa; tasa de incidentes de integridad académica asociados a IA	Registros de integridad académica; informes de rediseño curricular; encuestas de percepción estudiantil	Semestral	● Básico: Sin política de integridad académica actualizada; evaluaciones sin rediseño ● Intermedio: Política actualizada; al menos 30% de programas con rediseño evaluativo ● Avanzado: Arquitectura evaluativa robusta en todos los programas; incidentes < umbral definido
Seguridad y soberanía de datos	% de sistemas IA con análisis de privacidad completado; % de datos estudiantiles procesados en infraestructura con acuerdos de protección verificados	Análisis de impacto de privacidad (DPIA); contratos con proveedores; informes de ciberseguridad	Anual	● Básico: Sin análisis de privacidad; contratos sin cláusulas de protección de datos educativos ● Intermedio: DPIA realizadas para sistemas críticos; contratos revisados; política de datos vigente ● Avanzado: Soberanía de datos garantizada; preferencia por soluciones auditables; plan de contingencia activo

Nota. Elaboración propia a partir de los marcos de gobernanza y riesgo citados en este capítulo.

6.9 CONCLUSIONES

La gobernanza, la ética y la regulación de la IA en la educación superior no son obstáculos a la innovación: son las condiciones que la hacen sostenible y legítima. Una universidad que integra IA sin estructuras de control verificables no innova; asume riesgos que otros —estudiantes, docentes, la sociedad— terminarán pagando.

Este capítulo ha articulado cinco criterios de legitimidad que constituyen un estándar mínimo, no un horizonte ideal. El primero es la gobernanza transversal por capas: los sistemas de IA deben describirse por capas, con responsables, controles y trazabilidad en cada una de ellas. El segundo es la proporcionalidad del riesgo: no todos los usos tienen el mismo peso institucional y en los casos de alta consecuencia —admisión, evaluación certificatoria, decisiones de trayectoria— incluso riesgos de probabilidad moderada pueden ser inaceptables. El tercero es la auditabilidad: la capacidad institucional de reconstruir decisiones y demostrar cumplimiento, que no se reduce a "explicar el modelo" sino que abarca documentación, registros de cambios, monitoreo e incidentes. El cuarto es la contestabilidad: el derecho práctico, no solo teórico, de impugnar, revisar y obtener reparación, diseñado en el sistema desde el inicio. El quinto es la mejora continua: convertir incidentes y auditorías en insumos para actualizar políticas y elevar el estándar de gobernanza (NIST, 2023; ISO/IEC, 2023a; OECD, 2023).

El desafío para América Latina y el Sur Global es particularmente estratégico. La regulación internacional avanza con mayor rapidez que la capacidad institucional de la región para implementarla, lo que crea el riesgo de una doble dependencia: tecnológica respecto a proveedores del Norte Global y normativa respecto a marcos diseñados sin participación de los sistemas universitarios del Sur. La respuesta no puede ser solo reactiva; debe incluir inversión deliberada en capacidades de gobernanza, cooperación interuniversitaria para construir escala, y participación activa en los espacios internacionales donde se definen los estándares de IA en educación.

La universidad inteligente del siglo XXI no se define por la cantidad de herramientas de IA que adopta, sino por su capacidad de demostrar —con evidencia verificable— que cada algoritmo que influye en su misión formativa responde a un propósito pedagógico legítimo, está bajo supervisión efectiva y protege los derechos de quienes dependen de ella

para aprender, certificarse y desarrollarse. Esa capacidad de demostración es, en sí misma, la forma más profunda de autonomía universitaria en la era de la inteligencia artificial y el requisito indispensable para que la gobernanza ética sostenga la arquitectura de la inteligencia formativa desarrollada en los capítulos precedentes.

6.10 REFERENCIAS

- Autenrieth, D., & Schluchter, J.-R. (2025). Partners or tools? Anticipatory governance for human-AI complementarity in higher education. *Zeitschrift für Hochschulentwicklung*, 20(4). <https://doi.org/10.21240/zfhe/20-4/10>
- European Union. (2024). Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). Official Journal of the European Union. <https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng>
- Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., Schafer, B., Valcke, P., & Vayena, E. (2018). An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. *Minds and Machines*, 28(4), 689–707. <https://doi.org/10.1007/s11023-018-9482-5>
- Holmes, W., Bialik, M., & Fadel, C. (2019). Artificial intelligence in education: Promises and implications for teaching and learning. Center for Curriculum Redesign.
- Holmes, W., Porayska-Pomsta, K., Holstein, K., Sutherland, E., Baker, T., Shum, S. B., Santos, O. C., Rodrigo, M. T., Cukurova, M., Bittencourt, I. I., & Koedinger, K. R. (2021). Ethics of AI in education: Towards a community-wide framework. *International Journal of Artificial Intelligence in Education*, 32(3), 504–526. <https://doi.org/10.1007/s40593-021-00239-1>
- ISACA. (2018). COBIT 2019 framework: Governance and management objectives. ISACA.
- International Organization for Standardization & International Electrotechnical Commission. (2023a). ISO/IEC 42001:2023 – Artificial intelligence – Management system. ISO. <https://www.iso.org/standard/42001>

- International Organization for Standardization & International Electrotechnical Commission. (2023b). ISO/IEC 23894:2023 – Information technology – Artificial intelligence – Guidance on risk management. ISO. <https://www.iso.org/standard/77304.html>
- Kosbar, Y. (2026). TransformED futures: Towards human-centred, ethical and inclusive use and governance of AI in higher education. *World Future Review*. <https://doi.org/10.1177/19467567251406292>
- Levy-Yeyati, E. (2025). An enabling regulatory framework for artificial intelligence in Latin America and the Caribbean. Inter-American Development Bank. <https://doi.org/10.18235/0013801>
- Medina-Zuta, P., Sotomayor-Soloaga, P., Becerra, G., Bobadilla-Pérez, M., & Yáñez-Guajardo, A. (2025). Perspectivas iberoamericanas sobre la regulación de inteligencia artificial en educación superior: Un análisis comparativo. *Revista de Estudios y Experiencias en Educación*, 24(56). <https://doi.org/10.21703/rexe.v24i56.3159>
- Miao, F., & Holmes, W. (2023). Guidance for generative AI in education and research. UNESCO. <https://doi.org/10.54675/EWZM9535>
- National Institute of Standards and Technology. (2023). Artificial intelligence risk management framework (AI RMF 1.0) (NIST AI 100-1). <https://doi.org/10.6028/NIST.AI.100-1>
- National Institute of Standards and Technology. (2024). Artificial intelligence risk management framework: Generative artificial intelligence profile (NIST AI 600-1). <https://doi.org/10.6028/NIST.AI.600-1>
- OECD. (2023). Advancing accountability in AI: Governing and managing risks throughout the lifecycle for trustworthy AI (OECD Digital Economy Papers, No. 349). OECD Publishing. <https://doi.org/10.1787/2448f04b-en>
- OECD. (2024). Defining AI incidents and related terms (OECD Artificial Intelligence Papers, No. 16). OECD Publishing. <https://doi.org/10.1787/d1a8d965-en>
- Scarfe, P., Watcham, K., Clarke, A., & Roesch, E. (2024). A real-world test of artificial intelligence infiltration of a university examinations system: A "Turing Test" case study. *PLOS ONE*, 19(6), e0305354. <https://doi.org/10.1371/journal.pone.0305354>

- Tertiary Education Quality and Standards Agency. (2023). Assessment reform for the age of artificial intelligence (Discussion paper). <https://www.teqsa.gov.au/sites/default/files/2023-09/assessment-reform-age-artificial-intelligence-discussion-paper.pdf>
- Tertiary Education Quality and Standards Agency. (2024). The evolving risk to academic integrity posed by generative artificial intelligence: Options for immediate action. <https://www.teqsa.gov.au/sites/default/files/2024-08/evolving-risk-to-academic-integrity-posed-by-generative-artificial-intelligence.pdf>
- UNCTAD. (2021). Digital economy report 2021: Cross-border data flows and development – For whom the data flow. United Nations Conference on Trade and Development. https://unctad.org/system/files/official-document/der2021_en.pdf
- UNCTAD. (2024). Digital economy report 2024: Shaping an environmentally sustainable and inclusive digital future. United Nations Conference on Trade and Development. https://unctad.org/system/files/official-document/der2024_en.pdf
- UNESCO. (2021). Recommendation on the ethics of artificial intelligence. UNESCO. <https://unesdoc.unesco.org/ark:/48223/pf0000381137>
- UNESCO IESALC. (2025). Accelerating digital transformation in higher education: UNESCO IESALC marks International Day for Digital Learning. <https://www.iesalc.unesco.org/en/articles/accelerating-digital-transformation-higher-education>
- 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. <https://doi.org/10.1186/s41239-019-0171-0>