

Ley que promueve las tecnologías del lenguaje y los modelos de inteligencia artificial en lenguas indígenas u originarias y establece la gobernanza de sus datos lingüísticos

Promueve corpus y modelos de lenguaje en quechua, aimara y lenguas amazónicas con participación de los pueblos hablantes, consentimiento individual de cada hablante, protección de las voces frente a la clonación, licencias que permiten beneficio compartido y prohibición de apropiación exclusiva. Por afectar derechos colectivos, se propone someter el texto a consulta previa antes del dictamen.

Comisión (Cámara de Diputados): Inclusión Social, Familia, Mujer y Pueblos Andinos, Amazónicos y Afroperuanos (+ Ciencia, Innovación Tecnológica y Sociedad Digital) · **Revisora (Senado):** En asuntos de Salud, Educación, Cultura, Mujer y Desarrollo Social y Digital · **Prioridad:** alta · **Costo fiscal:** bajo · **Plazo:** mediano

Plan Perú 2050: Inteligencia Artificial, Cohesión Social, Capital Del Conocimiento

I. Exposición de motivos

Según los Censos Nacionales 2017 del INEI, 3 735 682 personas (13,9% de la población de 5 y más años) aprendieron a hablar en quechua, 444 389 (1,7%) en aimara y 210 017 (0,8%) en otra lengua nativa. El Ministerio de Cultura identifica 48 lenguas indígenas u originarias: 4 andinas y 44 amazónicas. La Ley 29735 garantiza el derecho a usar estas lenguas en el ámbito público, pero los grandes modelos de lenguaje comerciales tienen escaso o nulo desempeño en ellas, lo que traslada la brecha lingüística al mundo digital: servicios públicos con IA que solo funcionan en castellano excluyen a quienes más los necesitan. El Reglamento de la Ley 31814 no contiene disposiciones sobre lenguas originarias. A la vez, la recolección de grabaciones y textos por terceros sin reglas claras expone a los pueblos a la apropiación de su patrimonio lingüístico, y la publicación abierta de grabaciones de voz —datos que pueden ser biométricos— facilita la clonación de voz y la suplantación en comunidades con baja alfabetización digital. La Ley 27811 protege conocimientos colectivos solo vinculados a recursos biológicos, por lo que no cubre los datos lingüísticos. El Plan IA Perú 2050 (CIP/CNPP) identifica la IA multilingüe como palanca estratégica y fija como meta tres modelos peruanos en producción en español, quechua y aimara al 2030. Por regular aspectos que afectan directamente derechos colectivos, la medida debe ser consultada a los pueblos indígenas u originarios (Convenio 169 de la OIT, art. 6; STC 0022-2009-PI/TC).

II. Fórmula legal

EL CONGRESO DE LA REPÚBLICA ha dado la Ley siguiente:

LEY QUE PROMUEVE LAS TECNOLOGÍAS DEL LENGUAJE Y LOS MODELOS DE INTELIGENCIA ARTIFICIAL EN LENGUAS INDÍGENAS U ORIGINARIAS Y ESTABLECE LA GOBERNANZA DE SUS DATOS LINGÜÍSTICOS

Artículo 1. Objeto

La presente ley tiene por objeto promover las tecnologías del lenguaje y los sistemas de inteligencia artificial en lenguas indígenas u originarias, en desarrollo de la Ley 29735, Ley que regula el uso, preservación, desarrollo, recuperación, fomento y difusión de las lenguas originarias del Perú, y establecer la gobernanza de los datos lingüísticos de los pueblos indígenas u originarios.

Artículo 2. Definiciones

Para efectos de la presente ley: a) Datos lingüísticos: grabaciones de voz, textos, transcripciones, traducciones, léxicos y anotaciones en lenguas indígenas u originarias; b) Corpus lingüístico público: colección de datos lingüísticos obtenida con fondos públicos o con participación de entidades públicas; c) Tecnologías del lenguaje: traductores automáticos, reconocimiento y síntesis de voz, correctores, modelos de lenguaje y aplicaciones similares.

Artículo 3. Línea de acción de tecnologías del lenguaje en lenguas originarias

3.1. El Ministerio de Cultura, en coordinación con el Ministerio de Educación, el Consejo Nacional de Ciencia, Tecnología e Innovación (CONCYTEC) y la Secretaría de Gobierno y Transformación Digital de la Presidencia del Consejo de Ministros, desarrolla, en el marco de sus competencias y de la política nacional de lenguas originarias, una línea de acción de tecnologías del lenguaje en lenguas originarias. La presente disposición no crea programas presupuestales, órganos ni unidades orgánicas.

3.2. La línea de acción prioriza la construcción de corpus lingüísticos públicos, el desarrollo de herramientas de traducción y voz para servicios públicos, y el apoyo a universidades e investigadores, en especial a los pertenecientes a pueblos indígenas u originarios.

Artículo 4. Participación, consentimiento y protección de la voz

4.1. La recolección de datos lingüísticos para corpus públicos requiere: a) el consentimiento expreso e informado de cada hablante, conforme a la Ley 29733, Ley de Protección de Datos Personales, otorgado por escrito cuando se trate de grabaciones de voz que permitan identificarlo; b) la cesión o licencia de sus derechos conexos como intérprete, conforme al Decreto Legislativo 822, Ley sobre el Derecho de Autor; y c) el acuerdo de las organizaciones representativas del pueblo cuya lengua se documenta, cuando el corpus tenga alcance colectivo o incluya conocimientos colectivos.

4.2. Los pueblos indígenas u originarios, a través de sus organizaciones representativas, pueden establecer condiciones de uso sobre corpus que contengan conocimientos tradicionales o contenidos de carácter sagrado o restringido.

4.3. Las grabaciones de voz se publican como datos abiertos solo cuando el hablante lo haya autorizado por escrito para ese fin, informado de su carácter irrevocable y de los riesgos de síntesis de voz. En los demás casos se ponen a disposición mediante acceso controlado para investigación y servicios públicos. Los corpus de voz se publican con medidas que impidan identificar al hablante cuando ello sea técnicamente posible sin afectar su utilidad; su licencia prohíbe expresamente su uso para identificación biométrica o para la reproducción sintética de la voz de un hablante concreto sin su consentimiento. Los datos que permitan identificar a los hablantes se conservan separados y bajo control de la entidad.

4.4. Lo dispuesto en el presente artículo se aplica sin perjuicio de la Ley 27811 y de la Ley 29785, Ley del derecho a la consulta previa a los pueblos indígenas u originarios, reconocido en el Convenio 169 de la Organización Internacional del Trabajo.

Artículo 5. Licencias y beneficio compartido

5.1. Los corpus lingüísticos públicos se ponen a disposición bajo licencias públicas estandarizadas que aprueba el Ministerio de Cultura, que pueden ser abiertas o de uso no comercial, según lo acordado conforme al numeral 4.2, y que permiten en todo caso su uso en investigación, educación y servicios públicos.

5.2. El uso comercial de corpus con licencia no comercial para entrenar modelos requiere una licencia específica que contiene, como mínimo: a) la atribución a los pueblos hablantes; b) el acceso gratuito de las entidades públicas y de las organizaciones representativas a las herramientas derivadas, para fines no comerciales; y c) una contraprestación de beneficio compartido destinada al fortalecimiento de la lengua, conforme al reglamento.

5.3. Los corpus lingüísticos de pueblos indígenas u originarios se rigen por la presente ley y prevalecen sobre las reglas generales de reutilización libre de datos abiertos.

Artículo 6. Servicios públicos en lenguas originarias asistidos por IA

Las entidades públicas que implementen asistentes virtuales u otros sistemas de IA de atención a la ciudadanía en ámbitos donde predomine una lengua indígena u originaria, conforme al Mapa Etnolingüístico del Perú, incorporan progresivamente dicha lengua, en concordancia con los estándares de pertinencia lingüística de los servicios públicos digitales y con validación por intérpretes o traductores inscritos en el Registro Nacional de Intérpretes y Traductores de Lenguas Indígenas u Originarias del Ministerio de Cultura.

Artículo 7. Evaluación de calidad

Los sistemas de IA en lenguas originarias usados por el Estado se evalúan con métricas públicas de desempeño y con la participación de hablantes nativos antes de su puesta en producción, a fin de evitar errores que afecten derechos.

Artículo 8. Prohibición de apropiación indebida y licencia legal

8.1. Se prohíbe la apropiación de corpus lingüísticos públicos mediante derechos exclusivos que impidan su uso por los propios pueblos hablantes o por el Estado. Ningún derecho de propiedad intelectual adquirido o reclamado sobre obras, programas, bases de datos o modelos desarrollados con dichos corpus puede ejercerse para impedir el uso del corpus original por los pueblos hablantes o por el Estado; lo dispuesto prevalece sobre los derechos del fabricante de bases de datos previstos en el Decreto Legislativo 822 respecto de corpus lingüísticos públicos.

8.2. Quien entrene modelos con corpus lingüísticos públicos otorga, por mandato de la presente ley, una licencia gratuita, no exclusiva y perpetua a favor del Estado y de las organizaciones representativas de los pueblos hablantes para usos no comerciales, educativos y de servicio público de dichos modelos.

Primera. Disposición Complementaria Final

Reglamentación. El Poder Ejecutivo, mediante decreto supremo refrendado por el ministro de Cultura y el ministro de Educación, reglamenta la presente ley en un plazo de doscientos cuarenta días hábiles contados a partir del día siguiente de su publicación, previo proceso de consulta previa conforme a la Ley 29785 en lo que afecte directamente los derechos colectivos de los pueblos indígenas u originarios, en particular las condiciones de uso, licencias y beneficio compartido.

Segunda. Disposición Complementaria Final

Lineamientos sobre datos de voz. La Autoridad Nacional de Protección de Datos Personales aprueba, en el plazo previsto en la disposición anterior, lineamientos sobre el tratamiento de grabaciones de voz en corpus lingüísticos.

Tercera. Disposición Complementaria Final

Financiamiento. La implementación se financia con cargo al presupuesto institucional del Ministerio de Cultura, del Ministerio de Educación y del CONCYTEC, y con recursos de cooperación internacional, sin demandar recursos adicionales al Tesoro Público.

Cuarta. Disposición Complementaria Final

Vigencia. La presente ley entra en vigencia al día siguiente de su publicación en el diario oficial El Peruano.

III. Análisis costo-beneficio

El costo es bajo a moderado y escalable: la construcción de corpus requiere trabajo de campo, traductores y anotadores, financiable con fondos concursables existentes y cooperación internacional. La ley no crea entidades ni programas presupuestales (art. 3.1), conforme al art. 79 de la Constitución y a la STC 00018-2023-PI/TC. El beneficio es alto en equidad: permite que servicios públicos digitales lleguen a millones de hablantes de lenguas originarias, mejora la pertinencia de la educación y la salud, y genera un activo científico en el que el país tiene ventaja única. La gobernanza reduce el riesgo de apropiación del patrimonio lingüístico y de clonación de voces. Consulta previa: por afectar directamente derechos colectivos, se recomienda que la comisión dictaminadora someta la fórmula legal a consulta previa antes de aprobar el dictamen (Convenio 169 de la OIT, art. 6; STC 0022-2009-PI/TC), y el reglamento se aprueba previa consulta conforme a la Ley 29785; ello añade meses al trámite pero reduce el riesgo de inconstitucionalidad por la forma. Vinculación con las Políticas de Estado 3 (Afirmación de la identidad nacional) y 35 del Acuerdo Nacional.

Normas que modifica

- Ley 29735, Ley de lenguas originarias, uso en el ámbito público — *complementa*
- Decreto Legislativo 822, Ley sobre el Derecho de Autor, derechos del fabricante de bases de datos — *complementa*
- Ley 29733, Ley de Protección de Datos Personales, datos sensibles (voz) — *complementa*
- Ley 29785, Ley del derecho a la consulta previa, aplicación al reglamento — *complementa*

Entidades implicadas (quién hace qué)

- MINCUL (rectora): Desarrolla la línea de acción (art. 3), aprueba licencias estandarizadas (art. 5.1), administra el Registro de Intérpretes (art. 6) y conduce la consulta previa del reglamento (DCF Primera).

- **MINEDU** (ejecutora): Coordina la línea de acción y co-refrenda el reglamento (art. 3, DCF Primera).
- **CONCYTEC** (financiadora): Apoya a investigadores con fondos concursables (art. 3.2).
- **ANPD** (reglamentadora): Aprueba lineamientos sobre grabaciones de voz (DCF Segunda) y supervisa el consentimiento (art. 4).
- **INDECOPI** (supervisora): Autoridad en derechos de autor y conocimientos colectivos aplicable a los arts. 4.1.b y 8.
- **Organizaciones representativas de los pueblos indígenas u originarios** (consultiva): Acuerdan corpus colectivos, fijan condiciones de uso (art. 4) y participan en la consulta previa (DCF Primera).
- **Entidades públicas con asistentes de IA de atención ciudadana** (regulada): Incorporan progresivamente la lengua predominante y evalúan calidad (arts. 6 y 7).

Sujetos obligados: Entidades que recolecten corpus públicos o implementen asistentes de IA en zonas de lengua originaria predominante; desarrolladores públicos o privados que entrenen modelos con corpus lingüísticos públicos.

Beneficiarios: Hablantes de lenguas originarias: alrededor de 4,39 millones de personas de 5 y más años aprendieron a hablar en quechua, aimara u otra lengua nativa (INEI, Censos 2017).

Impacto económico estimado (confianza media; referencial, no oficial)

Tipo: gasto

Costo para el Estado: No cuantificable con información pública

Horizonte: 2027-2031

Financiamiento: Presupuesto institucional del Ministerio de Cultura, MINEDU y CONCYTEC y cooperación internacional, sin demandar recursos adicionales al Tesoro Público. Requiere coordinación con el Poder Ejecutivo e informe previo de sostenibilidad fiscal del MEF antes del dictamen (art. 79; STC 00018-2023-PI/TC).

Supuestos: El costo depende del número de lenguas y horas de corpus priorizadas (trabajo de campo, traductores, anotadores) y del costo del proceso de consulta previa, que no se publica de forma estandarizada. La seudonimización de audios tiene costo marginal bajo (revisor de seguridad). La ley no obliga a montos: la línea de acción se ejecuta con presupuesto institucional y cooperación.

Beneficio para el Estado: Servicios públicos digitales con pertinencia lingüística; menor costo de intermediación en atención presencial. Cualitativo.

Beneficio para la sociedad: Inclusión digital de hablantes de lenguas originarias; protección frente a apropiación y clonación de voz; beneficio compartido por uso comercial. Cualitativo.

Inversión privada: Posible inversión de desarrolladores en modelos multilingües bajo licencias con beneficio compartido; no cuantificable.

Costo para privados: Licencia específica y contraprestación de beneficio compartido para uso comercial de corpus no comerciales; licencia legal gratuita al Estado y pueblos.

Riesgos: Demora por consulta previa; licencias no comerciales ineficaces frente a desarrolladores extranjeros; recolección sin consentimiento adecuado; baja oferta de traductores certificados.

Revisión especializada (asistida por IA, sobre la versión anterior)

- **Constitucional** — Observaciones graves: La propuesta regula de manera específica derechos colectivos de los pueblos indígenas u originarios (gobernanza de sus datos lingüísticos, condiciones de uso, beneficio compartido), por lo que es una medida legislativa de afectación directa que requiere consulta previa conforme al Convenio 169 de la OIT y la jurisprudencia del TC. Sin consulta, corre riesgo de inconstitucionalidad por la forma.
- **Técnica legislativa** — Observaciones graves: La finalidad es legítima y coherente con la Ley 29735, pero una ley que establece la gobernanza de los datos lingüísticos y conocimientos de los pueblos indígenas afecta directamente derechos colectivos y debería someterse a consulta previa. Además, limita derechos de propiedad intelectual sin articularlos con el régimen andino y deja al reglamento la definición del beneficio compartido.
- **Concordancia normativa** — Con observaciones: Normas citadas correctas (Ley 29735, Ley 31814, Ley 29733). Requiere precisar la consulta previa (Ley 29785) y armonizar con la protección de conocimientos colectivos (Ley 27811) y con la reutilización libre de

IA-04.

- **Seguridad** — Con observaciones: Propuesta de bajo riesgo de seguridad, pero los corpus de voz son datos potencialmente biométricos y permiten la clonación de voz de los hablantes, lo que no se aborda.
- **Derecho tecnológico** — Con observaciones: Propuesta valiosa y con cifras censales correctamente atribuidas. Tiene tres problemas técnico-jurídicos: confunde consentimiento individual (Ley 29733) con consentimiento colectivo y consulta previa (Ley 29785); trata las grabaciones de voz sin considerar que pueden ser datos biométricos sensibles y que una licencia abierta irrevocable choca con la revocabilidad del consentimiento; y llama 'licencias abiertas' a licencias con restricciones de uso comercial.

IV. Efecto de la vigencia de la norma

Complementa la Ley 29735 (lenguas originarias) y su Reglamento (D.S. 004-2016-MC), la Ley 31814, la Ley 29733 (datos de voz), la Ley 27811 (conocimientos colectivos) y el D.Leg. 822 (establece una limitación legal a los derechos del fabricante de bases de datos respecto de corpus públicos y una licencia legal a favor del Estado y los pueblos). Prevalece sobre las reglas generales de reutilización de datos abiertos (coordinado con IA-04). Requiere consulta previa de la medida legislativa antes del dictamen y del reglamento conforme a la Ley 29785. Alternativamente, puede incorporarse como artículos de la Ley 29735 al dictaminar. No deroga normas.

V. Indicador de éxito al 2031

Al 2031, corpus públicos abiertos y herramientas de traducción y voz evaluadas para quechua, aimara y al menos cinco lenguas amazónicas, y al menos tres modelos peruanos en producción en español, quechua y aimara (meta 2030 del Plan IA Perú 2050).

Normas relacionadas

- Ley 29735, Ley que regula el uso, preservación, desarrollo, recuperación, fomento y difusión de las lenguas originarias del Perú — <https://www.leyes.congreso.gob.pe/Documentos/Leyes/29735.pdf>
- Ley 29785, Ley del derecho a la consulta previa a los pueblos indígenas u originarios, reconocido en el Convenio 169 de la OIT — <https://www.leyes.congreso.gob.pe/Documentos/Leyes/29785.pdf>
- Ley 31814, Ley que promueve el uso de la inteligencia artificial — https://leyes.congreso.gob.pe/Documentos/2021_2026/ADLP/Texto_Consolidado/31814-TXM.pdf
- Ley 29733, Ley de Protección de Datos Personales — <https://www.leyes.congreso.gob.pe/Documentos/Leyes/29733.pdf>
- Ley 27811, régimen de protección de los conocimientos colectivos de los pueblos indígenas vinculados a los recursos biológicos — <https://www.leyes.congreso.gob.pe/Documentos/Leyes/27811.pdf>

Antecedentes

No se identificaron en SPLEY proyectos del periodo 2021-2026 sobre IA en lenguas originarias (búsqueda de la revisión de concordancia normativa; por confirmar al dictaminar).

Referencias internacionales

- Nueva Zelanda: Te Hiku Media: corpus y modelos de voz en te reo Mori bajo licencia de kaitiakitanga (gobernanza indígena de datos) — <https://tehiku.nz>
- Global: Global Indigenous Data Alliance, Principios CARE para la gobernanza de datos indígenas — <https://www.gida-global.org/>
- América Latina: LatamGPT, modelo de lenguaje regional abierto (CENIA, Chile) — <https://www.latamgpt.org>
- Global: UNESCO, Recomendación sobre la ética de la IA (diversidad lingüística y cultural) — <https://www.unesco.org/en/articles/recommendation-ethics-artificial-intelligence>

Fuentes

- INEI, Perú: Perfil Sociodemográfico. Censos Nacionales 2017 — https://www.inei.gob.pe/media/MenuRecursivo/publicaciones_digitales/Est/Lib1539/libro.pdf
- Ministerio de Cultura, Base de Datos de Pueblos Indígenas: lista de lenguas indígenas u originarias — <https://bdpi.cultura.gob.pe/lenguas>
- Ley 29785, Ley del derecho a la consulta previa — <https://www.leyes.congreso.gob.pe/Documentos/Leyes/29785.pdf>

Propuesta ciudadana de libre uso — no es un proyecto presentado. Versión web:

<https://unimauro.github.io/agenda-legislativa-tech/p/IA-06.html>