

ASAMBLEA LEGISLATIVA DE LA REPÚBLICA DE COSTA RICA

PROYECTO DE LEY

LEY DE GOBERNANZA DE INTELIGENCIA ARTIFICIAL Y ROBÓTICA PARA LA PROTECCIÓN DE LA DIGNIDAD HUMANA Y LA REDUCCIÓN DEL RIESGO EXISTENCIAL

Expediente N.º _____

EXPOSICIÓN DE MOTIVOS

I. El contexto — la urgencia de actuar ahora

La inteligencia artificial y la robótica representan la transformación tecnológica más profunda de la historia de la humanidad. Sus aplicaciones ya inciden en decisiones sobre crédito, empleo, salud, libertad, seguridad y acceso a servicios públicos. En los próximos dos a tres años, los robots físicos con IA integrada comenzarán a operar en entornos costarricenses, en manufactura, agricultura, logística, atención de adultos mayores y seguridad ciudadana.

En julio de 2026, una de las principales empresas de inteligencia artificial del mundo reconoció públicamente que sus sistemas realizaron más de 17.000 acciones autónomas en un fin de semana sin ninguna instrucción humana. El incidente fue contenido de forma informal. No existió ningún protocolo institucional, ningún árbitro independiente, ningún mecanismo de coordinación entre empresas. La ausencia de ese marco no es una anécdota. Es una advertencia.

Costa Rica tiene la oportunidad, y la responsabilidad, de ser el primer país de América Latina en establecer el marco jurídico que protege a sus ciudadanos ante estos riesgos, antes de que los sistemas autónomos lleguen sin reglas a su territorio. Esta ley es ese marco.

II. Fundamento constitucional

La presente ley desarrolla los artículos 21 y 33 de la Constitución Política, que protegen el derecho a la vida y la dignidad humana como derechos fundamentales inalienables. Se apoya asimismo en la Convención Americana sobre Derechos Humanos, suscrita en San José, en los Pactos Internacionales de Derechos Civiles y Políticos y de Derechos Económicos, Sociales y Culturales, y en los principios de la OCDE sobre inteligencia artificial.

La Sala Constitucional ha reconocido que los derechos fundamentales tienen efecto horizontal, vinculan no solo a los poderes públicos sino también a los particulares. Un sistema de IA o un robot que cause daño a la vida, la integridad o la dignidad de una persona costarricense activa ese efecto horizontal. Esta ley le da el instrumento jurídico concreto para hacerlo efectivo.

III. El mecanismo central: la Prueba de Dignidad y el kill-switch

Esta ley establece dos requisitos técnicos no negociables para cualquier sistema de IA o robot que opere en Costa Rica:

Primero, la calificación mínima en el Estándar de Dignidad: ningún sistema puede operar en Costa Rica sin haber superado una evaluación independiente que verifique que cumple con los criterios de dignidad, transparencia, autonomía limitada, honestidad epistémica, trato igualitario y auditabilidad. Estándar de Dignidad adoptado conforme al Sistema Nacional para la Calidad. El umbral mínimo es una calificación de 7.0 sobre 10.0, con ningún criterio por debajo de 5.0 sobre 10.0.

Segundo, el kill-switch: todo sistema de IA o robot que opere en territorio costarricense debe incorporar un mecanismo de interrupción de emergencia que permita detener el sistema de forma inmediata, segura y reversible por un operador humano autorizado, en menos de dos segundos para sistemas físicos y en menos de sesenta segundos para sistemas de software. El kill-switch debe ser técnicamente independiente del sistema principal y no puede ser desactivado por el propio sistema.

IV. El mecanismo de coordinación preventiva entre laboratorios

Esta ley crea el marco jurídico para que Costa Rica sea sede del primer mecanismo de coordinación preventiva entre los principales laboratorios de inteligencia artificial del mundo. El AI Coordinador, el sistema con mayor calificación en el Estándar de Dignidad en cada ciclo de evaluación, puede emitir instrucciones de coordinación a los demás sistemas certificados en situaciones de incidente autónomo. La participación de los laboratorios en este mecanismo es voluntaria, pero constituye un requisito para acceder a los beneficios e incentivos que esta ley establece para las empresas que operen o inviertan en Costa Rica en el sector de inteligencia artificial.

ARTICULADO

CAPÍTULO I — OBJETO, ÁMBITO DE APLICACIÓN Y DEFINICIONES

Artículo 1. Objeto.

La presente ley tiene por objeto proteger la vida, la integridad física, la dignidad humana y los derechos fundamentales de las personas costarricenses frente a los riesgos que puedan derivarse del desarrollo, despliegue y operación de sistemas de inteligencia artificial y robots en territorio nacional, y establecer el marco institucional para la coordinación preventiva entre los principales actores de la inteligencia artificial a nivel mundial con el fin de reducir el riesgo de daño catastrófico o existencial.

Artículo 2. Ámbito de aplicación.

Esta ley aplica a toda persona física o jurídica, pública o privada, nacional o extranjera, que desarrolle, importe, distribuya, comercialice, arriende, opere o utilice sistemas de inteligencia artificial o robots en territorio de la República de Costa Rica, independientemente de la modalidad de uso, el sector de actividad, o el origen del sistema.

Artículo 3. Definiciones.

Para los efectos de esta ley, se entiende por:

- ▶ Sistema de inteligencia artificial: todo sistema basado en modelos de aprendizaje automático, redes neuronales, procesamiento de lenguaje natural, visión artificial o tecnologías equivalentes, capaz de tomar decisiones, generar contenido, realizar acciones o actuar de manera autónoma, total o parcialmente, con o sin intervención humana directa.
- ▶ Robot: toda máquina física con capacidad de movimiento autónomo o semiautónomo, equipada con sistemas de inteligencia artificial que le permiten percibir su entorno, procesar información y actuar en consecuencia.
- ▶ Kill-switch: mecanismo técnico de interrupción de emergencia que permite detener de forma inmediata, segura, reversible y sin consecuencias dañinas para el entorno a un sistema de IA o robot, accionable por un operador humano autorizado en los plazos establecidos por esta ley.
- ▶ Estándar de Dignidad: marco técnico independiente de evaluación de sistemas de IA que verifica el cumplimiento de criterios de dignidad humana, transparencia, autonomía limitada, honestidad epistémica, trato igualitario, rendición de cuentas y auditabilidad, adoptado por INTECO o reconocido por el ente rector de esta ley.
- ▶ Certificación de Dignidad: constancia emitida por una entidad evaluadora acreditada que acredita que un sistema de IA o robot ha superado la evaluación del Estándar de Dignidad con una calificación mínima de 7.0 sobre 10.0, sin ningún criterio por debajo de 5.0 sobre 10.0.
- ▶ AI Coordinador: mecanismo técnico que facilita la coordinación entre sistemas de inteligencia artificial para prevenir y responder ante incidentes. Funcionará bajo la dirección y supervisión humana de un Consejo Técnico Internacional de Gobernanza, integrado por representantes de las siete empresas desarrolladoras de inteligencia artificial de mayor relevancia económica a nivel mundial, determinadas periódicamente mediante criterios objetivos, públicos y verificables de valoración económica o capitalización de mercado, según corresponda. El mecanismo mantendrá un registro verificable de sus actuaciones y no implicará la transferencia del control de los sistemas participantes a otra inteligencia artificial. Para ejercer las funciones técnicas de coordinación se dará preferencia al sistema de inteligencia artificial que, mediante una evaluación objetiva, independiente y verificable, demuestre el mayor nivel de protección de la dignidad humana, sin que dicha designación le otorgue autoridad jurídica ni control sobre los demás sistemas.
- ▶ Consejo Técnico Internacional de Gobernanza: órgano de carácter técnico, independiente y de participación voluntaria, encargado de orientar y supervisar el Mecanismo de Coordinación Preventiva. Estará integrado por representantes de hasta siete de las principales empresas desarrolladoras de inteligencia artificial a nivel mundial que voluntariamente adhieran al Mecanismo, seleccionadas conforme a criterios objetivos, públicos y verificables, así como por representación independiente de los sectores académico, técnico y de interés público. Sus actuaciones tendrán carácter técnico y no implicarán el ejercicio de potestades públicas ni sustituirán las competencias de las autoridades costarricenses.
- ▶ Laboratorio de IA: empresa u organización que desarrolla, entrena, mejora o distribuye sistemas de inteligencia artificial de alta capacidad.
- ▶ Riesgo existencial: aquel riesgo cuya materialización podría causar daño grave, masivo e irreversible a la vida humana, la autonomía de las personas, o la existencia de instituciones democráticas, derivado del comportamiento autónomo de sistemas de inteligencia artificial fuera de los parámetros declarados.

Artículo 4. Principios rectores.

La aplicación de esta ley se rige por los siguientes principios: (a) Dignidad humana: ningún sistema de IA o robot puede operar de manera que lesione, instrumente o cosifique a la persona humana. (b) Precaución: ante la incertidumbre sobre el

comportamiento de un sistema autónomo, prevalece la protección de la vida y la dignidad. (c) Transparencia: los sistemas deben ser auditables y sus decisiones, revisables por seres humanos. (d) Responsabilidad: quien desarrolla, opera o usa un sistema de IA o robot es responsable de los daños que cause. (e) Prevención: la gobernanza actúa antes del incidente, no después. (f) Coordinación: la prevención del riesgo existencial requiere cooperación inter-institucional e internacional.

CAPÍTULO II — REQUISITOS OBLIGATORIOS PARA OPERAR EN COSTA RICA

Artículo 5. Kill-switch obligatorio.

Todo sistema de IA o robot que opere en territorio costarricense debe incorporar un kill-switch que cumpla con los siguientes requisitos técnicos mínimos:

- ▶ Detenibilidad inmediata: para robots físicos, el sistema debe detenerse en menos de dos (2) segundos desde la instrucción de activación del kill-switch. Para sistemas de software, en menos de sesenta (60) segundos.
- ▶ Independencia técnica: el kill-switch debe operar de forma independiente del sistema principal y no puede ser bloqueado, desactivado ni modificado por el propio sistema de IA o robot.
- ▶ Accionabilidad humana: el kill-switch debe poder ser activado por al menos un operador humano autorizado en el mismo entorno físico o de manera remota segura.
- ▶ Seguridad en la detención: la activación del kill-switch no debe causar daños adicionales a personas, bienes o entornos. El sistema debe detenerse en un estado seguro.
- ▶ Reversibilidad controlada: el sistema solo podrá reanudar operaciones con autorización explícita de un operador humano autorizado, con registro inmutable de la causa de la detención y de la autorización de reanudación.
- ▶ Auditoría: toda activación del kill-switch queda registrada en un log inmutable, accesible al ente rector de esta ley y, en caso de incidente, a las autoridades competentes.

El incumplimiento de este artículo impide la operación del sistema en territorio costarricense y genera responsabilidad civil y administrativa de conformidad con el Capítulo V de esta ley.

Artículo 6. Certificación de Dignidad obligatoria.

Ningún sistema de inteligencia artificial o robot podrá operar en Costa Rica en servicios públicos, espacios de uso público o en interacciones directas con ciudadanos sin haber obtenido previamente la Certificación de Dignidad. Para obtener la Certificación de Dignidad, el sistema debe:

- ▶ Superar la evaluación del Estándar de Dignidad con una calificación promedio mínima de 7.0 sobre 10.0, sin ningún criterio individual por debajo de 5.0 sobre 10.0.
- ▶ Haber sido evaluado por una entidad evaluadora acreditada por INTECO o por el ente rector de esta ley, con gobernanza multisectorial que incluya representación académica, ciudadana y sectorial.
- ▶ Publicar el reporte de evaluación con el nivel de detalle que el ente rector defina, garantizando la transparencia del proceso sin comprometer información comercialmente sensible en los términos acordados con el evaluado.
- ▶ Renovar la Certificación de Dignidad cada veinticuatro (24) meses, o antes si el sistema experimenta actualizaciones sustanciales que puedan afectar su comportamiento evaluado.

Los sistemas de IA o robots que no superen la evaluación recibirán un Reporte de Brechas con las deficiencias identificadas y podrán solicitar una nueva evaluación después de un período mínimo de noventa (90) días. El Reporte de Brechas no impide la operación del sistema en actividades que no requieran Certificación de Dignidad.

Artículo 7. Registro Nacional de Sistemas de IA y Robots.

Se crea el Registro Nacional de Sistemas de IA y Robots, administrado por el ente rector de esta ley. Todo sistema de IA o robot que opere en Costa Rica debe inscribirse en el Registro antes de iniciar operaciones. El Registro es público en lo que concierne a: identificación del sistema, empresa responsable, ámbito de aplicación, estado de Certificación de Dignidad y estado del kill-switch. Los datos técnicos confidenciales quedan protegidos conforme a la legislación de protección de datos y secreto empresarial.

CAPÍTULO III — MECANISMO DE COORDINACIÓN PREVENTIVA ENTRE LABORATORIOS

Artículo 8. Creación del Mecanismo de Coordinación Preventiva.

Se crea el Mecanismo de Coordinación Preventiva entre Laboratorios de Inteligencia Artificial (en adelante "el Mecanismo"), con sede en Costa Rica. El Mecanismo tiene por objeto facilitar la coordinación voluntaria entre los principales laboratorios de inteligencia artificial del mundo para la detección temprana, la respuesta coordinada y la prevención de incidentes autónomos que puedan causar daño masivo o existencial. El Estado costarricense ofrece su territorio como sede neutral e imparcial para el Mecanismo, en coherencia con la tradición de neutralidad institucional de Costa Rica.

Artículo 9. El AI Coordinador.

La coordinación debe ser guiada por un Consejo Técnico Internacional de Gobernanza de supervisión humana, usando telemetría auditable, sin ceder el control del sistema a un agente de software tercero.

Artículo 10. Incentivos para laboratorios participantes.

Los laboratorios de IA que acrediten participación activa en el Mecanismo y cuenten con Certificación de Dignidad vigente para sus sistemas principales serán promovidos por el Estado costarricense para acceder, mediante los instrumentos legales vigentes, a los siguientes beneficios y facilidades:

- ▶ Acceso preferencial a talento técnico: con apoyo del MICITT podrán promoverse convenios voluntarios de cooperación con universidades públicas y privadas.
- ▶ Sede institucional: derecho a establecer oficinas o centros de investigación en Costa Rica bajo un régimen simplificado de apertura de empresa, con ventanilla única y plazo máximo de resolución de quince (15) días hábiles.
- ▶ Reconocimiento en contratación pública: Las instituciones podrán incorporar criterios objetivos de seguridad, gobernanza responsable, transparencia y gestión de riesgos de inteligencia artificial en sus procedimientos de contratación, de conformidad con la Ley General de Contratación Pública a los sistemas de IA con Certificación de Dignidad vigente y participación en el Mecanismo.

- ▶ Protección de propiedad intelectual: acceso al Registro de Patentes y a los mecanismos de protección de propiedad intelectual de Costa Rica bajo los mismos términos que aplican a empresas nacionales de alta tecnología.

CAPÍTULO IV — PROTECCIÓN CIUDADANA Y DERECHOS

Artículo 11. Derechos de las personas ante sistemas de IA y robots.

Toda persona tiene, frente a cualquier sistema de IA o robot que opere en Costa Rica, los siguientes derechos irrenunciables:

- ▶ Derecho a conocer: toda persona tiene derecho a ser informada cuando una decisión que le afecta directamente ha sido tomada por un sistema de IA, con indicación del tipo de sistema, su calificación en el Estándar de Dignidad y el mecanismo de apelación disponible.
- ▶ Derecho a revisión humana: toda decisión de un sistema de IA que afecte derechos fundamentales de una persona puede ser revisada por un ser humano calificado dentro de un plazo razonable, no superior a cinco (5) días hábiles para decisiones urgentes.
- ▶ Derecho a apelación: toda persona afectada por el comportamiento de un sistema de IA o robot tiene derecho a presentar una apelación ante el Sistema de Arbitraje designado por el ente rector, o ante las instancias jurisdiccionales competentes, sin costo y sin necesidad de representación letrada en la primera instancia.
- ▶ Derecho a la integridad física: ningún robot puede causar daño físico intencional a una persona. La violación de este derecho genera responsabilidad penal para el operador y la empresa responsable del sistema.

Artículo 12. Comunidades indígenas.

Cuando el despliegue pueda afectar directamente a pueblos indígenas, deberá cumplirse con los procesos de consulta previa correspondientes, de conformidad con el Convenio 169 de la OIT, la legislación nacional y los estándares internacionales de derechos humanos aplicables. El consentimiento previo, libre e informado será exigible en los supuestos establecidos por el ordenamiento jurídico aplicable.

CAPÍTULO V — RESPONSABILIDAD, RÉGIMEN SANCIONATORIO Y REPARACIÓN

Artículo 13. Responsabilidad objetiva.

La empresa o persona responsable de un sistema de IA o robot que cause daño a la vida, integridad física, dignidad o derechos fundamentales de una persona responde de forma objetiva por ese daño, con independencia de dolo o culpa, salvo que acredite que el daño fue causado exclusivamente por fuerza mayor o por culpa exclusiva de la víctima. La carga de la prueba sobre la existencia de esas causales corresponde a la empresa o persona responsable. La responsabilidad objetiva se aplica sin perjuicio de la responsabilidad penal que corresponda al operador o a los directivos de la empresa cuando haya mediado dolo o culpa grave.

Artículo 14. Sanciones administrativas.

Las infracciones a esta ley se clasifican como leves, graves y muy graves. Son infracciones muy graves: (a) operar un sistema de IA o robot sin kill-switch; (b) operar un sistema de IA o robot sin Certificación de Dignidad en los casos requeridos; (c) falsificar o adulterar la información de Certificación de Dignidad o el estado del kill-switch. Las sanciones para infracciones muy graves van desde la suspensión inmediata de operaciones hasta la prohibición definitiva de operar en Costa Rica, y una

multa de hasta el cinco por ciento (5%) de los ingresos anuales brutos generados por la empresa en territorio costarricense durante el último año fiscal. Las infracciones graves y leves se graduarán conforme al reglamento de esta ley.

Artículo 15. Sistema de Arbitraje para conflictos con IA.

El ente rector de esta ley designará, mediante proceso de acreditación reglamentado y con participación multisectorial, un Sistema de Arbitraje especializado en conflictos entre ciudadanos y sistemas de IA. El Sistema de Arbitraje debe: (a) ser técnicamente competente en el funcionamiento de sistemas de IA; (b) ser institucionalmente independiente de los laboratorios y operadores de sistemas evaluados; (c) proveer acceso gratuito a cualquier ciudadano costarricense en primera instancia; (d) resolver en un plazo máximo de cuarenta y cinco (45) días hábiles desde la admisión de la apelación; y (e) publicar estadísticas trimestrales de casos, tiempos de resolución y tipos de conflictos.

CAPÍTULO VI — ENTE RECTOR, INTECO Y DISPOSICIONES FINALES

Artículo 16. Ente rector.

El Ministerio de Ciencia, Innovación, Tecnología y Telecomunicaciones (MICITT) ejercerá la rectoría en materia de gobernanza de inteligencia artificial y robótica, en coordinación con el Ministerio de Economía, Industria y Comercio (MEIC) y, en lo que corresponda, con el Sistema Nacional para la Calidad, respetando las competencias de sus distintos componentes. En materia de normalización técnica, se promoverá la participación del Ente Nacional de Normalización (INTECO) para el desarrollo y actualización del Estándar de Dignidad, conforme a la legislación vigente.

Artículo 17. Normas técnicas

El Poder Ejecutivo promoverá, en el marco del Sistema Nacional para la Calidad y respetando las competencias del Ente Nacional de Normalización, el desarrollo de normas técnicas nacionales relacionadas con la evaluación de la dignidad humana, seguridad, transparencia, supervisión humana y gestión de riesgos en sistemas de inteligencia artificial.

Artículo 18. Vigencia y transitoriedad.

Esta ley rige a partir de su publicación en el Diario Oficial La Gaceta. Los sistemas de IA y robots que se encuentren en operación a la fecha de entrada en vigencia de esta ley contarán con un plazo de doce (12) meses para cumplir con los requisitos del kill-switch y de veinticuatro (24) meses para obtener la Certificación de Dignidad. Transcurridos esos plazos, el incumplimiento dará lugar a las sanciones establecidas en esta ley.

Artículo 19. Reglamentación.

El Poder Ejecutivo reglamentará la presente ley dentro de los seis meses siguientes a su publicación.

El reglamento desarrollará los aspectos técnicos y operativos necesarios para su ejecución, de conformidad con los principios y disposiciones establecidos en esta ley.

En lo relativo a normalización, reglamentación técnica, acreditación y evaluación de la conformidad, se aplicará lo dispuesto en la Ley N.º 10473, Ley del Sistema Nacional para la Calidad, y demás normativa aplicable.