Una propuesta para el trabajo responsable con los sistemas de IA generativa en ingeniería

A proposal for trustworthy generative IA system in engineering

Celso Vargas-Elizondo[1]

Vargas-Elizondo, C. Una propuesta para el trabajo responsable con los sistemas de IA generativa en ingeniería. *Tecnología en Marcha*. Vol. 39 Nº especial sobre Inteligencia Artificial. Febrero, 2026. Pág. 28-40.

https://doi.org/10.18845/tm.v39i5.8518

1 Instituto Tecnológico de Costa Rica, Costa Rica.
celvargas@itcr.ac.cr
https://orcid.org/0000-0002-1701-6186

Palabras clave

Sistemas de IA generativa; regulación de IA; IA en ingeniería; ética de la IA; modelos fundacionales.

Resumen

En este artículo realizamos una propuesta para regular el trabajo con sistemas de IA generativa en la enseñanza de ingeniería en el Instituto Tecnológico de Costa Rica. Se compone de tres secciones principales. En la primera hacemos una breve presentación de la capacidad instalada en el país tanto en tecnología móvil como en internet fijo. En la segunda presentamos las características principales de este tipo de sistemas con la finalidad de contribuir a la comprensión del potencial y las limitaciones de estos sistemas. En la tercera sección, describimos y ejemplificamos una propuesta de regulación de estos sistemas en la enseñanza de la Ingeniería. Es relevante utilizar los atributos de ingeniería, específicamente, las habilidades profesionales que propone como punto de partida para nuestra propuesta. Utilizamos las propuestas emergentes de regulación propuestas por la UNESCO en el documento consultivo del 2024 [1]. A saber, la basada en principios que, para nuestros propósitos son las habilidades propuestas en los atributos, el enfoque basado en riesgos y su relación con los tres principios de gobernanza HOTL, HITL y HIC. El tercer componente es el enfoque denominado “sandbox” que permite identificar fortalezas y amenazas de lo regulado y no regulado en el uso docente de estos sistemas.

Keywords

AI generative systems; regulation of AI; AI in engineering; ethics of AI; foundational models.

Abstract

In this paper, we propose a regulation for the use of generative IA systems for teaching engineering at the Costa Rica Institute of Technology. It is divided into three main sections. In the first, a brief account of the national installed capacity in mobile and broadband technology. In the second section, the main and general features of generative IA systems are provided, aimed at promoting a better understanding of the potentialities and limitations of these systems. In the third, we describe and exemplify our proposal for regulating the use of these systems in the teaching of Engineering. Our starting point is the professional skills derived from the corpus of engineering attributes in force in our institution. On the regulation, it is based on three of the emergent systems presented by UNESCO in 2024 [1]: principles-based, risk-based, and sandbox systems. These are correlated with the governance principles, HOTL, HITL, and HIC. Sandbox is proposed as an important tool for identifying strengths and weaknesses in the existing regulation and non-regulation of IA systems in education.

Introducción

Hablar de “trabajo responsable” con los sistemas de IA generativa, conlleva tomar en consideración al menos los tres aspectos: a) la infraestructura requerida que permita utilizar las distintas funciones del sistema de una manera rápida y eficiente, y el sistema; b) que se cumpla con la formación mínima sobre el conocimiento de las potencialidades y limitaciones de estas herramientas y c) aplicar un conjunto de criterios éticos que permitan un uso adecuado de estas tecnologías. Tanto los países, las instituciones de educación como los docentes deben participar en el logro de este uso responsable de los sistemas de IA. Dada la complejidad

de esta tarea, nos limitaremos a una de las funciones más importantes de estos sistemas, la generación de contenidos, y dejaremos de lado muchos otros aspectos relevantes del uso de estos sistemas, por ejemplo, el "fine tune", el "fine prompt", y el desarrollo de aplicaciones (API), solo para mencionar algunos ejemplos. De igual manera, la utilización de sistemas de IA en la formación es ingeniería es bastante compleja. Por ejemplo, en investigación los requerimientos son muy altos y es necesario tener acceso a bases de datos o crear nuevas bases de datos que puedan ser utilizar en procesos de investigación. Así pues, nos centramos el trabajo con los sistemas de IA generativa en los procesos de enseñanza que no demanden un componente muy fuerte de investigación, pero que lo incluyan.

Metodología

Esta investigación está basada en análisis bibliográfico, reflexión sobre la mejor forma de introducir una regulación de IA en la enseñanza de ingeniería y tomando en consideración mi experiencia como docente en la institución. Dada la complejidad del tema, nos centramos en el proceso de enseñanza-aprendizaje, excluyendo los procesos de investigación cuando estos vayan más allá de actividades propias de los cursos.

En relación con el enfoque adoptado en este trabajo, no he encontrado enfoques similares en el área de la ingeniería. Quizá el que se pueda aproximar un poco es el enfoque del Instituto Tecnológico de Monterrey (2024) [2], el cual sigue siendo bastante general, en el sentido que requiere elaboración por parte de las personas docentes. Otros enfoques como el de la National Society of Professional Engineers (2024) [3], basado en principios resulta muy general para los propósitos de este artículo. El trabajo de Shilpa Prabhdesai (2025) [4] es muy importante en cuanto analiza riesgos y habilidades requeridas para el buen uso de IA en ingeniería, entre otros, pero falta el nivel concreción que realizamos aquí. Finalmente, mencionamos el trabajo del Dr. Miguel Angel Medina-Romero (2025) [5] para regular el uso de IA en las instituciones de educación del Estados de Michoacán México que dado su objetivo resulta bastante general para una propuesta como la nuestra.

Resultados

Infraestructura y el sistema

En lo referente a la infraestructura al menos dos aspectos consideramos relevantes: a) la capacidad mínima requerida y b) las características propias de los sistemas de IA generativa. En Costa Rica aspectos de cobertura, tanto en tecnología móvil como en internet fijo, hay una importante penetración de estas tecnologías (96,5% en tecnología móvil) y un 74% de los hogares cuentan con acceso a internet fijo [6]. Otro asunto tiene que ver con las velocidades tanto de subida como de descarga. El requisito mínimo es tecnología 3G que se ubica alrededor de 3Mbps, pero presenta demoras significativas en las descargas y en la subida de información. El índice Latinoamericano de Inteligencia Artificial 2025 [7] establece como indicador, dentro de los factores habilitantes, el de penetración de 5G en tecnología móvil. Y para banda ancha, el requisito mínimo es el plan económico (5GB, con velocidad de descarga de 256kb/s) (Indice…, p. 26). Según este informe, en el caso de Costa Rica, únicamente el 1,64% de los hogares cuentan con este plan económico, es decir, que la gran mayoría tienen planes que superan en mucho este nivel. En el mapa de la SUTEL 2025 se registran los operadores y las velocidades de descarga y de envío, siendo el más alto 400 Mbps. Sin embargo, no es uno de los operadores principales del país. En tecnología móvil, el país latinoamericano que tiene

la mayor velocidad de descarga es Uruguay con un promedio de velocidad de 68,1 Mbps, y en banda ancha, Costa Rica ocupa la primera posición en velocidad promedio de descarga y subida en Latinoamérica.

Para una universidad de ingeniería como el ITCR, no supone ningún inconveniente para el trabajo con los sistemas de IA generativa en cuando a la velocidad de descarga y subida. Sin embargo, hay otras consideraciones relacionadas con los sistemas de IA mismos que deben tomados en cuenta.

El primer aspecto importante se relacionada con la disponibilidad de contar con plataformas de IA local. Tener una IA local conlleva requerimientos de hardware importantes que garanticen respuestas rápidas a consultas y de buena calidad. Muchos modelos de AI como ChatGPT, MidJourney, y Google Bard (véase [8] 2025 para tendencias) todavía no cuentan con esta posibilidad, con lo cual las consultas que se hagan se hacen directamente a la nube, y esto puede aumentar el periodo de latencia necesario para tener una respuesta. Esto no supone un gran problema, pero sí hay que tener paciencia para obtener el resultado deseado. Lo que es importante indicar es que hay interesantes tendencias en el Edge computing and Storing que podrían hacer más fácil el uso de los sistemas de IA generativa en actividades que demandan mejores búsquedas y profundidad, algunas de las cuales tardarán más en llegar a nuestro país. La Organización Greeks for Greeks [9] ha enlistado 7 tendencias relevantes en este campo: 1) El aumento de Edge-IA que busca incrustar IA directamente en los dispositivos Edge que pueda hacer más eficiente la recolección de datos y la toma de decisiones inteligentes; 2) Sinergias entre 5G y Edge computing para acelerar el procesamiento y reducir los periodos de latencia; 3) el surgimiento de “Edge containers” para mejorar la “portabilidad, aislamiento y eficiencia de los recursos”; 4) EaaS (Edge-as-a-Service) una plataforma que brinda acceso a “poder de procesamiento, almacenamiento y recursos de la red localizados en la edge”; 5) ampliación de la seguridad y la privacidad; 6) convergencia IT/OT (information technolgies, operational technology) lo que favorece una mayor integración entre estos componentes haciendo más eficiencias los recursos y el flujo de información y 7) Integración de fog computing, que permite hacer más flexible y complementar los recursos de la nuble y los recursos localizados regionalmente, esto mediante capas de computación distribuida que permita administrar mejor todos estos recursos. Los sistemas de IA generativa sin duda se adaptarán a estas innovaciones para mejorar la calidad de las respuestas y la interacción con el usuario, incluyendo una mayor personalización del modelo. Finalmente, es relevante para el futuro de los sistemas de IA la consideración de las tendencias en “multicloud”. Ya algunas empresas ofrecen servicios en los que es transparente el uso de distintas nubes, y han desarrollado LLMs (Large Language Models) que permiten esta interacción de manera más efectiva y economizando recursos computacionales (véase IFX Networks [10] 2025).

Una segunda consideración tiene que ver con el tipo de arquitectura que utilizan. En general hay dos arquitecturas principales: los modelos fundaciones grandes (como ChatGPT, Claude, Gemini, entre otros), a veces llamados LLMs (modelos de lenguajes grandes), y los modelos fundaciones pequeños que se utilizar llevar a cabo tareas específicas en ámbitos determinados. DeepSeek utiliza una arquitectura de mezcla de expertos, cuyos módulos comparten similitudes con los SLMs. Sin embargo, hay discusión en denominar la arquitectura de DeepSeek como basada en SLMs, debido a que cada uno de sus módulos requieren un número de parámetros que exceden el promedio de los SLMs, el cual se sitúa entre 1 y 10 mil millones de parámetros (véase [14] para conocer algunas de las propiedades y ejemplos de SLMs y también [11], para un estudio más detallado de la arquitectura y funcionamiento de DeepSeek). Es importante señalar que el uso de LLMs como SLMs para designar estos sistemas de IA tiene algunas limitaciones, siendo la principal el hecho de que actualmente estos sistemas se utilizan para realizar tareas que van más allá del lenguaje, como para reconocimiento de imágenes (para

uso médico, por ejemplo), producción de obras de arte, música, entre otros. Varios de estos sistemas funcionan de manera multimodal (combinación de varios de estos dominios), de manera que la mejor denominación es la de modelos fundacionales grandes y pequeños, según corresponda.

Formación mínima sobre los sistemas de IA

La UNESCO [12] ha indicado como un reto importante, la formación sobre las características, potencialidades y limitaciones de los sistemas de IA en su estado actual. En relación con esto, queremos realizar una presentación general sobre estos tres aspectos para situar, nuestra propuesta de regulación que hacemos en la sección 3 de este trabajo.

Comencemos con algunas características generales de estos sistemas. Hemos diferenciado entre dos clases generales de modelos fundacionales: los grandes y los pequeños, les hemos llamado LFMs y SFMs. Ambos comparten los siguientes dos aspectos: a) la homogenización, es decir, "la consolidación de metodologías para construir sistemas de machine learning a través de un rango amplio de aplicaciones" lo cual "proporciona un fuerte apalancamiento para las tareas, pero también para crear puntos de falla específicos" ([13]: 3). Con variantes, la misma metodología, la misma forma de analizar y de responder en distintos sistemas de IA. Esto es importante para entender el comportamiento de un sistema de IA. b) La emergencia, "significa que el comportamiento de un sistema es implícitamente inducido más que explícitamente construido; esto es tanto una fuente de excitación científica y ansiedad acerca de las consecuencias no anticipadas" (Idem).

Desde el punto de vista del trabajo con estos sistemas en entornos educativos, dos aspectos son importantes en relación con las dos características indicadas: a) Debido a la forma de entrenamiento de estas redes neuronales, en el sentido de que la manera cómo responden es inducida durante el proceso de entrenamiento, no se puede afirmar de manera segura cómo se comportará la red neural en determinada circunstancia, su comportamiento es más bien probabilístico, compara un conjunto de alternativas y selecciona aquella que tiene más probabilidad de manera consistente con su proceso de aprendizaje y con lo solicitado. En este sentido, uno de estos sistemas es considerado como una caja negra. Esto significa que sus respuestas pueden desviarse de la manera cómo un experto en el campo respondería a la misma consulta. b) considerando que estos sistemas responden de manera autónoma, y como señalan Rishi Bommasani y otros [14]), " a pesar de la ubicuidad de la machine learning en IA, tareas semánticamente complejas en procesamiento de lenguaje natural (NPL) y visión computacional tales como responder a una pregunta o reconocer un objeto, donde las entradas son oraciones o imágenes, todavía ha requerido de dominio experto para realizar "feature engineering", esto es, escribir en lógica de dominio específico para convertir los datos sin procesar en rasgos de alto nivel (…) a fin de hacerlos más adecuados para los métodos de machine learning populares" (p.4). Es decir, es importante tener en cuenta las limitaciones que estos sistemas presentan en la actualidad y que generan errores, aunque se busca eliminarlos cuando sean detectados.

Un segundo aspecto que debe ser tomado en consideración es la manera como se entrenan estas redes neuronales. Utilizan grandes cantidades de datos. En general se diferencian dos aspectos aquí: las bases de datos y los llamados "datasheets" (traducido como hoja de datos o ficha técnica). El primero designa la cantidad y el tipo de datos que son utilizados para entrenar el sistema, mientras que las hojas de datos contienen valiosa información sobre el desempeño y otra información relevante sobre el sistema. El conocimiento general sobre estos dos aspectos resulta fundamental para comprender mejor las potencialidades y las limitaciones de los

sistemas de IA generativa. Una serie de temas están asociados con estos dos componentes, entre ellos, aspectos de propiedad intelectual, de la calidad de las respuestas que proporcione el sistema, de la veracidad de la información que se incluye en las respuestas.

Dentro de la regulación de la IA propuesta por la Unión Europea (2024) [15], se incluyen prohibiciones y criterios que deben ser cumplidos por las bases de datos para el entrenamiento de sistemas de IA generativa. Las prohibiciones están establecidas en el artículo 5, e incluye sistemas entrenadas con bases de datos que pueden inducir discriminación o vulneración de las personas. En general, esta prohibición se aplica a todos los sistemas que se desarrollen con esta finalidad, incluyendo claramente, las bases de datos utilizadas. En la parte regulatoria, el artículo 10 establece una serie de criterios que deben cumplir las bases de datos (20 en total) para satisfacer este requerimiento, incluyendo un sistema de tres colores para identificar y etiquetar los distintos componentes de los datos, dependiendo del tipo de información que contenga (protegida por propiedad intelectual, contiene datos sensibles, información verificada, entre otros) y, el artículo 11 regula lo referente a la información técnica que acompaña el desarrollo del sistema. Ambos aspectos mencionados tienen como finalidad garantizar que estos sistemas de desarrollen con estándares de calidad y con información verificable.

Un tercer y último aspecto por mencionar se relaciona con las consultas o prompts. La UNESCO [16] ha prestado atención a éste a fin de que se aproveche mejor el potencial educativo de estos sistemas, específicamente para ChatGPT, resaltando la importancia del conocimiento para hacer buenas consultas. El principio general, como desarrollaremos más adelante, es cambiar la idea de “usar” estos sistemas a “trabajar” estos sistemas. El enfoque en trabajo colaborativo, donde el sistema complementa las capacidades, conocimientos y habilidades que se tienen o que se quieren alcanzar.

Regulación de los sistemas de IA generativa en la enseñanza de la ingeniería

Propuesta de la UNESCO

En el documento de la UNESCO [1] *Consultation paper on AI Regulation* se identifican 9 marcos emergentes a nivel mundial para la regulación de la IA, los cuales no son incompatibles entre sí, sino que pueden combinarse dependiendo del objetivo y ámbito de la regulación. De acuerdo con esta propuesta, “el orden en el que se presentan los nueve enfoques regulatorios de IA es deliberadamente estructurada para guiar a los lectores partiendo de medidas menos intervencionistas, medidas regulatorias más suaves hasta las más coercitivas y demandantes” ([1]: 19).

Haremos una breve referencia a estos enfoques con el objetivo de situar la aproximación que haremos en este trabajo, de manera que lector interesado pueda consultar la fuente. Estos son los siguientes: a) el enfoque basado en principios que proporciona algunas guías para el desarrollo y el trabajo con IA que se incluyan de manera explícita durante todo el proceso (diseño, desarrollo y puesta en funcionamiento), es decir, criterios éticos, responsabilidad, centralidad en el ser humano y el respeto de los derechos humanos. Estos principios son segregados en 4 grupos de valores fundamentales y 10 principios ([1]: 22-23); b) Enfoque basado en estándares como aquellos que han desarrollado o que puedan desarrollar organizaciones internacionales que proponen estándares, como normas ISO o SME los cuales serán obligatorios y aplicables durante todo el ciclo del sistema, incluyendo consideraciones sobre el uso responsable. Se han propuesto bastantes estándares para diferentes ámbitos de aplicación de la IA, incluyendo bases de datos y preparación de la documentación, entre otros; c) enfoque basado en experimentos ágiles que busca el establecimiento de esquemas de regulación flexibles adaptables a diferentes áreas y usuarios, conocidas como “sandboxes”, que permitan a las autoridades regulatorias acceder a sistemas desarrollados (antes de su

puesta en funcionamiento) para determinar si cumplen o no con lo regulado por ley, por estándares o por otros medios e investigar otros aspectos de interés como nuevos modelos de negocios; d) enfoque facilitador y habilitador, se utilizan para identificar fortalezas y vacíos en la regulación y cuán preparado se está para desarrollar una IA responsable y transparente. Un ejemplo de esto es RAM (Readiness Assessment Methology) que identifica fortalezas y vacíos regulatorios en cinco dimensiones: "1) Legal, 2) Social y cultural, 3) Científica y educativa, 4) Económica y 5) Tecnológica e infraestructura" ([1]: 28). e) Adaptación de leyes existentes. Este enfoque prefiere adaptar las leyes existentes en diferentes ámbitos (por ejemplo, protección de datos personales) con nuevos requerimientos y modificando aquellos que se desactualicen de manera que las leyes presenten un carácter más dinámico y evolutivo lo que las hace relevantes en los nuevos contextos; f) Enfoque de acceso de información y mandatos de transparencia. Este enfoque puede ilustrarse de la siguiente manera: "La transparencia algorítmica está entre los más comunes principios de los marcos éticos de IA … Cumplir con este principio conlleva la implementación de instrumentos transparentes que permitan a la población el acceso a información básicas sobre los sistemas de IA adoptados por organizaciones públicas y privadas" ([1]: 31). g) enfoque basado en riesgos. Dado un conjunto de dimensiones que se consideran valiosos, este enfoque identifica tan exhaustivamente como sea posible, aquellas variables que pueden comprometer o afectar estas dimensiones, determina el impacto que su ocurrencia puede tener y propone un modelo de clasificación de riesgos tomando tanto el riesgo como su impacto y adapta las medidas que se consideran relevantes para evitar que estos eventos ocurran. Volveremos como este enfoque. h) Enfoque basado en derechos. El marco predilecto de este enfoque es los derechos humanos tal y como los encontramos en las proclamas universales y regionales, en agendas, como la agenda 2030 y otras normas que garanticen principios como la autonomía de las personas, la beneficencia, la justicia distributiva y la no-maleficencia. i) Enfoque basado en lesividad. Este enfoque centra su atención en la responsabilidad y las sanciones, y la forma de hacerlo puede ser muy diversa, por ejemplo, proponiendo códigos de ética para desarrolladores y responsables del funcionamiento de los sistemas de IA, o introduciendo éstas en distintos cuerpos normativos o en una ley general para regular la IA.

En este trabajo nos centramos en tres de estos enfoques considerándolos como complementarios: el enfoque basado en principios, el basado en riesgos y el de sandboxes. En cuanto al enfoque de principios diferenciaremos dos niveles: los principios que regulan los atributos de ingeniería y tres principios de gobernanza. Así pues, procederemos de la siguiente manera: primero introduciremos los principios generales que orientan la formación en ingeniería, segundo, introduciremos una perspectiva general sobre riesgos, tercero, los tres principios de gobernanza y su aplicación en ingeniería. En todos estos momentos, la utilización de sandboxes es importante para aumentar la comprensión, potencialidades, riesgos de la IA generativa y propiciar un trabajo responsable con estos sistemas.

Principios de la Ingeniería

Saint-Simon (1760-1825) es considerado como el padre de la ingeniería, su proyecto de "Internacionalismo tecnocrático" ([17]) contiene algunos elementos iniciales relevantes para entender el papel de la ingeniería en el mundo moderno. Saint-Simon pensó un nuevo orden social basado en la razón en la que los científicos, los ingenieros, los artistas y los administradores jugarían un papel fundamental en el desarrollo y progreso del nuevo orden. Los ingenieros juegan un papel social fundamental y consiste en mediar entre la ciencia y las necesidades sociales de manera que presenten soluciones a distintos problemas sociales mediante el desarrollo de tecnologías, y agregaríamos hoy, la adaptación tecnológica, la integración de distintas tecnologías y contribuir a la creación de políticas científico-tecnológicas, entre otros aspectos relevantes que beneficien a la sociedad. Los ingenieros deben tener una muy buena

formación científica y matemática, así como social, humana y ambiental para llevar a cabo mejor su labor de mediación y de transformación social. Debe, diríamos, tomar decisiones considerando como de igual peso (salvo razonamiento fundado) la salud, la seguridad y el ambiente, así como aquellos aspectos relativos a la excelencia y la responsabilidad profesional.

En este contexto y situándonos en los procesos de formación ingenieril, los principios que rigen esta formación están fuertemente basados en las capacidades o habilidades que los y las futuras ingenieras deben mostrar cuando concluyan su plan de estudios, así como en los ámbitos o áreas en las que esta formación es atinente. Los atributos de ingeniería son un buen punto de partida. Son 11 los atributos de ingeniería, los cuales mencionamos aquí en el siguiente cuadro 1.

Cuadro 1. Enumeración de los atributos de ingeniería.

- Conocimiento de ingeniería
- Análisis de problemas
- Diseño
- Investigación
- Herramientas de ingeniería
- Administración de proyectos y finanzas
- Trabajo individual y en equipo
- Ética y equidad
- Aprendizaje continuo
- Habilidades de comunicación
- Persona Ingeniera y el mundo

De un análisis de estos atributos obtenemos en la columna 1 las principales habilidades requeridas y en la columna 2 indicamos las áreas formativas de la ingeniería.

Cuadro 2. Habilidades y áreas de la ingeniería.

- Desarrollo sostenible
- Visión holística (ética. Social, ciencia, matemática, ambiental)
- Seguridad humana y sus dimensiones
- Recursos culturales, sociales y ambientales
- Obtención, uso, análisis e interpretación de datos
- Modelación de problemas y su análisis
- Análisis y decisiones económicas
- Trabajo en equipo e individual
- Diversidad e inclusión
- Evolución tecnológica
- Legislación y ambiente

Cada una de estas habilidades puede mapearse de varias maneras con los temas propuestos, por ejemplo, el pensamiento crítico y solución de problemas complejos está conectado con prácticamente todos temas. Por ello, podemos asumir un tipo de relación todos-todos para captar esta relación. Nuestro punto de partida en este artículo son los atributos de ingeniería. Aunque es importante tener presente que los atributos de egreso de ingeniería, en su formulación actual, relaciona solo algunas habilidades con algunos temas.

Estos atributos pueden ser jerarquizados de distintas maneras poniendo de manifiesto distintos matices y prioridades formativas. Una de estas jerarquizaciones ubica el atributo 11 "persona ingeniera y el mundo" como el atributo más importante y a partir de aquí podemos ubicar todos los demás atributos. De esta manera, por ejemplo, la capacidad de análisis y solución de problemas complejos implica todas las otras habilidades y considerando el área de desarrollo sostenible, también podemos ver las demás áreas como derivadas del desarrollo sostenible.

Clasificación de los riesgos

En esta propuesta seguimos la clasificación de riesgos propuesta por la Unión Europea en el documento "EU Artificial Intelligence Act"([14]) en cuatro categorías: a) riesgos inaceptables, b) de riesgo alto, c) de riesgo limitado y d) riesgo mínimo. Esta clasificación tiene como fondo cuatro principios éticos fundamentales y siete requerimientos generales claves que deben garantizarse. Los principios éticos son: a) respeto a la autonomía humana; b) prevención del daño; c) justicia y d) explicabilidad (explicability). De estos cuatro principios, obtenemos los siguientes 7 requerimientos claves: 1) Agencia humana y vigilancia, 2) Robustez técnica y seguridad, 3) Privacidad y gobernanza de datos, 4) Transparencia, 5) Diversidad, no discriminación y justicia, 6) Bienestar social y ambiental y 7) Responsabilidad ([14]: 2). Los riesgos inaceptables son aquellos para los cuales hay una gran certeza que incumplirán algunos de estos requerimientos o los valores fundamentales que dan sentido a estos. Ejemplos, incluyen la aplicación de IA en los procesos electorales donde se puede recurrir a DeepFakes y otros algoritmos con el propósito de obtener beneficios políticos por medios no transparentes. Los sistemas y aplicaciones de IA que presentan riesgos inaceptables son prohibidos dentro de la UE, salvo casos muy calificados como realización de investigaciones para mejorar los sistemas de IA. Los riesgos altos aplican a aquellos sistemas de IA a los cuales si no se les aplica un control estricto en todo el ciclo del sistema tienen una alta probabilidad de comprometer valores fundamentales como la diversidad, autonomía, justicia, la salud o el ambiente. Se incluyen dentro de esta categoría, por ejemplo, los sistemas que son diseñados e implementados utilizando bases de datos indiscriminadas que puede comprometer la propiedad intelectual y promover el plagio, así como el engaño. Los sistemas clasificados bajo riesgo limitado son aquellos que presentan un nivel importante de probabilidad de que puedan ser utilizados para manipular o para engañar. De manera general, los sistemas de IA generativos son clasificados como sistemas de riesgo limitado. Sin embargo, en contextos específicos pueden ser clasificados de manera diferente, como veremos en el caso de la docencia, cuando hay razones para hacerlo. Finalmente, los sistemas clasificados bajo la categoría de riesgo mínimo son aquellos que no presentan mayor afectación a valores fundamentales, tales como los filtros de spam o el uso del sistema de IA sobre definiciones de conceptos, como en Google, en los cuales la responsabilidad sobre el uso adecuado recae sobre el usuario.

Principios de gobernanza

El Grupo de Expertos de Alto nivel de la Unión Europea sobre IA ([18]) en el apartado sobre Agencia humana y vigilancia, introduce tres principios de gobernanza que resultan muy útiles para proponer una regulación de la IA en la institución:1) Human-on-the-loop (HOTL); 2) Human-in-the-loop (HITL) y 3) Human-in-Command (HIC). Estos principios aplican a todo el ciclo del

sistema de IA, desde su diseño, las bases de datos y datasheets que son utilizados para su entrenamiento, su puesta en funcionamiento y lo que se debe hacer con la información cuando el sistema deje de ser funcional. Como veremos, estos principios nos permiten también entender mejor la naturaleza de las interacciones con los sistemas de IA. Proporcionan un marco flexible para la toma de decisiones relacionadas con este tema. El primer principio (HOTL) es el menos estricto de todos y requiere únicamente "intervención humana durante el ciclo del diseño del sistema y el monitoreo del funcionamiento del sistema" a fin de garantizar que funciona de manera adecuada. El segundo principio (HITL) es mucho más estricto y "se refiere a la capacidad del ser humano de intervenir en cada ciclo de decisión del sistema, que en muchos casos no es posible ni deseable". Dos circunstancias limitan la intervención humana, según este principio: en aquellos casos donde no es posible, y en aquellos casos en los que no es deseable la intervención humana. Por ejemplo, debido a la gran cantidad de fuentes (base de datos) con las que es entrenada una red neuronal, en una consulta bien realizada, la respuesta del sistema puede incluir información que, para obtenerla de otra forma, se requiere un mayor esfuerzo investigativo y en ocasiones no se tiene la formación para realizarlo. Este caso es un tipo de intervención que no es deseable realizar, pero sí verificar. Este principio de asocia con el trabajo colaborativo humano-tecnología y ha recibido nombres como "cobótica" cuando se trata de trabajo colaborativo con robots.

Finalmente, el principio HIC, el más estricto de todos y "se refiere a la capacidad de supervisar toda la actividad del sistema de IA (incluyendo su impacto, en sentido amplio, económico, social, legal y ético) y la capacidad de decidir cuándo y cómo [actúa] el sistema en una situación particular. Esto incluye la decisión de no usar un sistema de IA en una situación particular, establecer los niveles de discreción durante el uso del sistema, o asegurar la capacidad de modificar una decisión tomada por el sistema" ([18]: 16)

Esbozo de una propuesta de regulación

Nuestro punto de partida son los distintos cursos que integran la malla curricular. Esto debido a que en la estructura curricular nuestra, los cursos son las unidades básicas. Claramente puede pensarse de manera diferente, con una aproximación más holística que parte de la malla curricular y termine en los cursos, con lo cual pueden establecerse relaciones entre los cursos de manera más articulada.

El primer paso es determinar el potencial uso de los sistemas de IA generativa para el curso que se va a impartir. Esto se puede hacer de dos maneras complementarias. Primero mediante un "sandbox" en el que se permita a estudiantes, docentes y otros hacer contribuciones de manera que se logre un nivel razonable de comprensión del potencial positivo y de los posibles impactos negativos de la utilización no regulada enumerando sus riesgos. La segunda forma es por el conocimiento y experiencia que tiene el o la docente en relación con estos aspectos. Sin embargo, lo mejor es proponer un "sandbox" que permita una construcción colectiva más sólida.

Habiendo obtenido los datos anteriores, podemos proceder de la siguiente manera: Dado un curso X, enlistamos las habilidades digamos a,b,c… involucradas. Teniendo en cuenta las distintas funciones de los sistemas de IA generativa (textos, simulaciones, imágenes, arte, análisis de imágenes, audios, videos entre las principales y sus interacciones entre otros), determinamos cuáles de estas funciones mantienen relación con las habilidades y con los contenidos del curso, y cuán fuerte es esa relación. Para ello utilizamos los resultados del sandbox. Seguidamente, determinamos los riesgos asociados con estos sistemas, ordenándolos, cuando se trata de más de un uso, según el grado de afectación negativo sobre el logro de los objetivos del curso (es decir, obstaculizan el logro de un determinado grupo de habilidades). Finalmente, asociamos el análisis anterior con los principios de gobernanza

indicados, es decir, con el nivel de intervención en el control del proceso involucrado. Para ello, consideramos el nivel de correspondencia siguiente, regulando adecuadamente por medio de controles aquellas situaciones en las que debe prohibirse la utilización de estos sistemas o limitarse de manera significativa. Por ejemplo, un curso que tenga como finalidad contribuir con la formación en habilidades de comunicación escrita, no tener ningún control sobre el uso de estos sistemas puede comprometer seriamente tanto esta habilidad como la de razonamiento crítico, tal y como lo puso de manifiesto el estudio preliminar realizado por el MIT en el 2025 ([19]). Claramente, la recomendación aquí no es prohibir su uso, sino regularlo más estrictamente. Para correlacionar riesgos con los principios de gobernanza utilizamos la siguiente correspondencia (excluyendo riesgos inaceptables).

Cuadro 3. Tipo de Riesgo-principio de gobernanza.

Tipo de riesgo	Nivel de intervención del docente según principio de Gobernanza
Alto riesgo	HIC (control del docente durante todo el proceso de uso de sistemas de IA generativa)
Riesgo limitado	HITL (trabajo colaborativo sistema IA-Docente-Estudiante)
Riesgo mínimo	HOTL (supervisión general del proceso de uso del sistema de IA generativa)

Estos principios de gobernanza aplicados a la docencia nos proporcionan importantes criterios sobre la forma de entender el papel de los sistemas de IA generativa en un determinado curso. Por ejemplo, en el curso anteriormente mencionado de comunicación escrita, el o la docente debe asegurarse que la utilización de estos sistemas de tal manera que ayuden a profundizar los temas a desarrollar y al mismo tiempo permitan que logre los objetivos pedagógicos planteados. La supervisión docente constante es requerida aquí. Por tanto, se aplica el principio de gobernanza HIC. Otros cursos, por ejemplo, que involucren habilidades de comunicación oral pueden ser mejor regulados utilizando el principio HITL de manera que el docente se asegure tanto el dominio de los temas del curso como el logro de los objetivos de la comunicación oral.

Quisiera ilustrar con un curso que imparto sobre ética en la ingeniería. El objetivo del curso es formar en análisis ético y aplicarlo a varias de las tecnologías de punta en este momento. Tradicionalmente, en este curso se realizaba una investigación bibliográfica sobre los temas asignados por el o la docente o en acuerdo con los y las estudiantes. Con el advenimiento de ChatGPT-3 la situación cambio, y ahora se realizan investigaciones cortas (3 en total durante el semestre), se permite trabajar con estos sistemas de manera regulada, y se enfatiza en habilidades como trabajo individual y en equipo, presentaciones grupales orales, la inclusión del tema de la equidad. El docente tiene el control sobre este proceso, pero se trabaja de manera colaborativa con estos sistemas de IA (HITL) para optimizar su uso.

El grupo de organiza en cinco grupos pequeños de cinco integrantes. Para cada uno de los grupos se les asigna, desde el inicio del curso, los temas que deben investigar de acuerdo con el cronograma propuesto. Cada uno de los temas se divide en subtemas de manera que cada uno o una de las estudiantes pueda desarrollar un tema. El coordinador o coordinadora es la encargada de reunir la información y dirigir el proceso de investigación y de presentación oral. Para cada investigación se han establecido los elementos y criterios que deben cumplirse: el enunciado del prompt para cada uno de los y las estudiantes, y el resultado obtenido. Como se trata de temas generales, el prompt debe tener los siguientes elementos: a) debe solicitar que responda como experto en el campo con amplia experiencia, b) que realice un desarrollo

temático de cada uno de los subtemas, c) que el documento tenga una extensión mínima de 12 páginas, d) en caso de que la información sea insuficiente complementarlo con otras búsquedas como ejemplos, biografías relevantes, aplicaciones, etc.

El coordinador o coordinadora debe remitir los resultados para que sean analizados por el docente para que haga las recomendaciones correspondientes. Como se encuentran redundancias en las consultas realizadas, el grupo debe identificar, complementar la información cuando se requiera, reasignar los temas de manera que no se presenten temas repetidos y organizar la presentación oral de los resultados. Finalmente, se solicita que se haga una revisión general para identificar sesgos en las respuestas a partir de una lista de temas preseleccionados por el docente.

Los resultados de la investigación deben ser presentados de manera oral y aplica la siguiente restricción: no está permitido que los y las estudiantes recurran a la lectura de su subtema, sino que tienen que mostrar dominio fluido del tema. Los resultados hasta el momento han sido muy satisfactorios y una característica que sobresale es que, bajo esta forma de investigación, se generan contenidos que analizan los temas de manera mucho más amplia que una investigación bibliográfica tradicional. El control del proceso por parte del docente es fundamental para garantizar que se logran los objetivos de aprendizaje tanto en los contenidos del curso como en el desarrollo de las habilidades consideradas en el curso. Los temas del curso se complementan con presentaciones por parte del docente y con tres trabajos asincrónicos en momentos específicos del semestre.

Recomendaciones

La propuesta presentada anteriormente es provisional, pero proporciona una metodología para la regulación de los sistemas de IA en la enseñanza de la ingeniería en el Instituto Tecnológico de Costa Rica. Realizamos tres recomendaciones para cada una de las principales secciones de este artículo.

Recomendación sección 1: En relación con la situación actual en cuanto a capacidad instalada (velocidad de subida y bajada, acceso, entre otros factores) la institución debe determinar la conveniencia de establecer una plataforma institucional que permita un acceso más rápido y apropiado de aquellos sistemas de IA generativa que potencien su utilización responsable en la institución, y analizar la posibilidad del uso de la infraestructura con la cuenta el CENAT (CONARE) que permita utilizar la supercomputadora disponible para la realización de simulaciones y otros procesos que permitan una mejor formación en la habilidad de soluciones a problemas complejos en ingeniería

Recomendación sección 2: El Instituto Tecnológico de Costa Rica debe contar con guías y otros recursos educativos para facilitar el trabajo adecuado con estas herramientas, incluyendo recomendaciones sobre como elaborar correctamente prompts, al mismo tiempo que incluye orientaciones generales sobre las potencialidades y limitaciones de estos sistemas a fin de obtener información confiable y veraz.

Recomendación sección 3: El Instituto Tecnológico de Costa Rica debe elaborar propuestas piloto orientadas a analizar los riesgos de no regular los sistemas de IA generativa en la institución, proponer sandboxes en los que participen estudiantes, docentes y personal de apoyo con el objetivo de proponer y probar regulaciones flexibles para trabajar con sistemas de IA generativa en docencia, y también en los procesos de investigación con el objetivo de obtener el mayor beneficio de estos sistemas tecnológicos.

Referencias

[1] UNESCO Consultation paper on AI regulation: emerging approaches across the world, 2024, https://unesdoc.unesco.org/ark:/48223/pf0000390979.

[2] Instituto Tecnológico de Monterrey, Lineamientos para el Uso Ético de Inteligencia Artificial, 2024. https://recursos.tecmilenio.mx/sites/default/files/2024-08/Lineamientos%20para%20Estudiantes.pdf

[3] National Society of Professional Engineers, Use of Artificial Intelligence in Engineering Practice, 2024. https://www.nspe.org/career-growth/ethics/board-ethical-review-cases/use-artificial-intelligence-engineering-practice

[4] Shilpa Prabhdesai, AI in Engineering – Ethical Implications, 2025. https://testrigor.com/blog/ai-in-engineering/#section role_of_stakeholders_in_ai_ethics

[5] Medina-Romero, Towards A Regulatory Framework For The Use Of Artificial Intelligence In Educational Institutions Of Michoacan, Mexico: Challenges, Principles, and Perspectives, 2025. https://www.revista-trans-digital.org/index.php/transdigital/article/view/491/824

[6] SUTEL, Mapa de Banda Ancha en Costa Rica, 2025. https://sutel.go.cr/pagina/mapa-de-banda-ancha-en-costa-rica

[7] CEPAL, Indice Latinoamericano de Inteligencia Artificial (ILIA) 2025. https://www.cepal.org/es/publicaciones/82514-indice-latinoamericano-inteligencia-artificial-ilia-2025

[8] Ranjan, The State of Local AI Model Deployment in 2025: Tech Landscape and Performance, 2025. https://arkwaysolutions.com/blog?slug=local-ai-model-deployment-2025-tech-landscape-performance

[9] Greeks for Greeks, Top 7 Trends in Edge Computing, 2025.

[10] IFX Networks Tendencias de multicloud para 2025: Transformando la infraestructura digital empresarial, 2025 . https://ifxnetworks.com

[11] Jphn Johnson, Small Language Models (SLM): A Comprehensive Overview 2025. https://huggingface.co/blog/jjokah/small-language-.

[12] Aixin Liu et al. DeepSeek-V3 Technical Report, 2024. https://arxiv.org/abs/2412.19437

[13] UNESCO ChatGPT and artificial intelligence in higher education: quick start guide, 2023. https://unesdoc.unesco.org/ark:/48223/pf0000385146

[14] Rishi Bommasani y otros, On the Opportunities and Risks of Foundation Models, 2023. https://crfm.stanford.edu/assets/report.pdf

[15] European Union EU Artificial Intelligence Act 2024. https://artificialintelligenceact.eu/es/ai-act-explorer/

[16] UNESCO, Guidance for generative AI in education and research, 2023, https://unesdoc.unesco.org/ark:/48223/pf0000386693

[17] Jan Eijking, Jan, Saint-Simon's Technocratic Internationalism, 1802–1825, 2021. https://intellectualhistory.web.ox.ac.uk/article/saint-simons-technocratic-internationalism-1802-1825

[18] HLEG-AI, Ethics Guidelines for Trustworthy AI, 2019. https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai

[19] Nataliya Kosmyna y otros, Your Brain on ChatGPT: Accumulation of Cognitive Debt when Using an AI Assistant for Essay Writing Task, 2025. https://arxiv.org/abs/2506.08872

Declaración sobre uso de Inteligencia Artificial (IA)

Los autores aquí firmantes declaramos que no se utilizó ninguna herramienta de IA para la conceptualización, traducción o redacción de este artículo.