

ÉTICA Y RESPONSABILIDAD PARA LA UTN EN EL USO DE LA INTELIGENCIA ARTIFICIAL



AUTORAS:

Gutiérrez, María de los Milagros
Spina, María Laura
Rozenhauz, Julieta
Yanotti, Paula

COLABORADORES:

Cuenca Pletsch, Liliana
Mariño, Gabriela
Toffolo, Alejandro
Lauria, María Florencia

Índice

Introducción	3
¿Cómo se puede definir la Inteligencia Artificial?	3
¿Por qué es necesario hablar de la ética en la Inteligencia Artificial?	5
Antecedentes	6
UNESCO	6
Declaración de Montevideo sobre IA y su impacto en América Latina	7
Disposición 2/2023 de Gabinete de Ministros	7
OCDE	7
IEEE	7
HUB de IA en la UTN	8
Ámbito de aplicación de esta recomendación	8
Clasificación de los datos manipulados en IA y sus riesgos asociados	8
Estrategias para el abordaje de la ética en IA desde la UTN	10
Objetivos	10
Lineamientos	11
Referencias	13



Introducción

La inteligencia artificial (IA) ofrece un potencial significativo para transformar el ámbito educativo, permitiendo personalizar el aprendizaje, mejorar la eficiencia y ampliar las oportunidades. No obstante, el uso de estas tecnologías debe estar guiado por principios éticos que aseguren su aplicación responsable y beneficiosa para toda la comunidad de la Universidad Tecnológica Nacional (UTN), asegurando su alineación con los valores de integridad académica, respeto por la privacidad, inclusión, equidad, sostenibilidad y respeto a los derechos humanos. Así, como se señala en la publicación de Micaela Tec: “Tecnologías, brechas digitales y desigualdades desde la perspectiva de género y diversidad”¹:

Es menester, en este sentido, indagar sobre los efectos que trae aparejada la exclusión de ciertos grupos de la población que se suceden al compás del avance tecnológico, en particular a mujeres y diversidades. En otras palabras, resulta prioritario trabajar para que los procesos tecnológicos promuevan la diversidad, la equidad y la inclusión en su uso, de modo de revertir discriminaciones y exclusiones. (Micaela Tec, 2023, p. 11) [1]

Ante este escenario, es sustancial que los lineamientos que se establezcan contemplen estos desafíos para evitar su reproducción en el proceso de desarrollo y expansión de la IA, a la vez, que promuevan una perspectiva crítica que aporte a la construcción de un entorno de producción tecnológica que no profundice las brechas digitales que afectan en forma desigual al acceso y el uso de las tecnologías por parte de las mujeres y la comunidad LGTBQ+. En otras palabras, la reflexión y visibilización de esta exclusión es un puntapié inicial y sustantivo para evitar la profundización de las desigualdades existentes.

La participación en el campo científico-tecnológico y, en particular, los roles que se asuman en términos de consumo o producción de tecnología incidirán de manera directa para seguir acrecentando privilegios de unos pocos o, en cambio, construir los derechos de las mayorías.

La sociedad necesita conocimiento para debatir acerca de estos desafíos, una herramienta posible es poner en diálogo a las diferentes disciplinas. Para lograrlo resulta “indispensable promover equipos de trabajo diversos e interdisciplinarios que aborden los problemas desde distintas perspectivas y uno de ellos debe ser el enfoque de género y diversidad” (Micaela Tec, 2023, p. 45). [2]

Tal como se observa, el impacto de la IA es global, y opera en la economía y en la sociedad como un todo. Por tal motivo, es necesario actuar rápidamente estableciendo reglas claras que permitan orientar la evolución y el desarrollo de esta disciplina, que maximicen sus beneficios en los diferentes campos de aplicación y minimicen los riesgos. Estas reglas, leyes, políticas y estrategias deben estar basadas en consideraciones éticas y legales.

Los lineamientos establecidos en esta resolución tienen como **público objetivo** a todos los actores de la UTN que pertenecen a los cuatro claustros que la conforman: docentes, estudiantes, no-docentes y graduados.

¿Cómo se puede definir la Inteligencia Artificial?

La inteligencia artificial es la disciplina de la ciencia de la computación que estudia la creación de máquinas inteligentes. Es el campo de estudio de la informática que se ocupa de crear sistemas y programas capaces de realizar tareas que, por lo general, requieren de la inteligencia humana. Estas tareas incluyen el aprendizaje, el razonamiento, la resolución de problemas, la percepción visual, el reconocimiento de voz, la toma de decisiones y la comprensión del lenguaje natural.

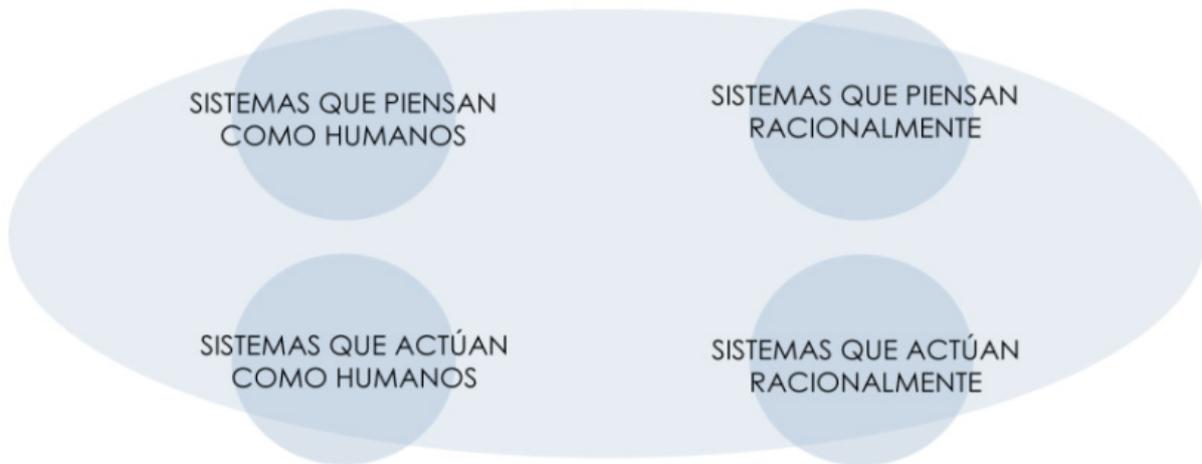
En otras palabras, la IA se centra en el desarrollo de algoritmos y modelos computacionales que pueden imitar o simular la inteligencia humana para realizar tareas específicas. Los sistemas de inteligencia artificial pueden aprender de datos, adaptarse a nuevas situaciones, y mejorar su rendimiento con el tiempo, lo que los hace capaces de realizar tareas complejas y resolver problemas en una variedad de dominios.

1 Micaela TEC: *Tecnologías, brechas digitales y desigualdades desde la perspectiva de género y diversidad* es una publicación que se realizó en forma colaborativa entre la Secretaría de Igualdad de género y Diversidad de la UTN y el Ministerio de Mujeres, Géneros y Diversidad de la Argentina.



En los últimos años, ha experimentado un rápido crecimiento, impulsado por la disponibilidad masiva de datos, el aumento en la capacidad computacional y los avances en algoritmos, especialmente en el campo del aprendizaje profundo.

Dado que la inteligencia tiene diferentes definiciones y formas de entenderla, en el campo de la IA surgieron representaciones diversas para manifestar la inteligencia en forma física. Se identifican cuatro dimensiones en las que la IA ha sido abordada desde sus comienzos: sistemas que piensan como humanos; sistemas que actúan como humanos; sistemas que piensan racionalmente; y sistemas que actúan racionalmente.



Estas definiciones y clasificaciones están estrechamente vinculadas a la concepción inicial de inteligencia de principios del siglo XX, cuando surgió esta disciplina, centrada en las capacidades cognitivas humanas. Los modelos de investigación generaron distintas áreas de estudio, influenciadas por la definición de inteligencia adoptada, con enfoques específicos.

Sin dudas, el crecimiento que ha tenido el área del aprendizaje automático en esta última década es superior y más rápido que el de otras áreas de la IA, viéndose revolucionada, principalmente, por el aprendizaje profundo y las redes neuronales artificiales (ANN, por su sigla en inglés *Artificial Neuronal Network*). Estas últimas representan un modelo matemático del funcionamiento del cerebro humano, en donde existe el concepto de neurona artificial, sinapsis y redes de neuronas organizadas en capas, que pueden aprender a partir de datos que se les brinda como entrada. Así, cada neurona artificial en una red representa un parámetro de una función matemática que describe los datos de entrada *vistos* o *aprendidos* por la red.

En un comienzo, estos modelos eran muy difíciles de generar, dado que requerían elevada capacidad de procesamiento y grandes volúmenes de datos a través de los cuales aprender. Con el avance de la tecnología, que mejoró enormemente el *hardware*, brindando capacidad de procesamiento y la generación masiva de datos a partir del uso de tecnologías *smart* o inteligentes (celulares, electrodomésticos, autos, máquinas en las fábricas, etc.) y disponibles en la nube, el aprendizaje automático comenzó, también, a perfeccionar sus modelos. Así, las redes neuronales se transformaron en redes profundas con capas y neuronas, en cada una, para representar funciones matemáticas con parámetros de cientos a billones, originando lo que hoy se conoce como aprendizaje profundo.

Este tipo de sistemas requieren de grandes cantidades de datos para aprender y poder realizar bien sus tareas, tales como: elaborar un pronóstico; reconocer imágenes; escribir un texto; buscar información sobre un tema particular; clasificar información; extraer datos de un texto; generar imágenes nuevas a partir de varias imágenes de entrada; componer canciones, poesía; identificar un objetivo militar; mantener una conversación sobre un tema particular; conducir un auto; participar de la cadena de producción en una fábrica; recomendar películas, música, inversiones o acciones a llevar a cabo; hacer diagnósticos médicos; interpretar imágenes de estudios médicos; realizar análisis clínicos, llevar adelante una cirugía sobre cuerpos humanos; encontrar la mejor ruta; calificar estudiantes; identificar el ingreso o salida del personal con datos biométricos, predecir fallas en máquinas, etc.



La lista es muy amplia y va creciendo a medida que los humanos desarrollan nuevas y más novedosas herramientas y aplicaciones a partir de este tipo de inteligencias. Resulta interesante destacar que los resultados dados como respuestas por los sistemas inteligentes que utilizan aprendizaje automático no pueden ser explicados, es decir, actúan como cajas negras, recibiendo datos de entrada y produciendo un resultado o salida, sin poder “ver” o “analizar” qué está sucediendo en el proceso interno para explicar por qué se obtuvieron dichos resultados, o cómo la información que se le dio en la entrada fue procesada para elaborar los resultados.

Todas estas acciones de la IA afectan al contexto, a las actividades realizadas, a la forma de relacionarse, estudiar, enseñar, trabajar, pensar y hasta de entretenerse. Por eso, este tipo de sistemas plantean desafíos en la garantía de calidad desde la ingeniería de *software*, especialmente, en cuanto a la validación y verificación de estos, ya que son sistemas difíciles de explicar desde su comportamiento, al no existir un código que pueda inspeccionarse.

La concepción de aprendizaje automático, de la cual uno de sus pioneros es Arthur Samuel [3], en 1959, define a la IA como el campo de estudio que brinda a las computadoras la capacidad de aprender sin ser programadas explícitamente, de modo que se trata de utilizar algoritmos genéricos capaces de aprender a partir de datos para resolver diferentes problemas.

Muchas son las similitudes que pueden encontrarse entre las acciones de los sistemas inteligentes con respecto a los humanos. La principal característica que identifica a los sistemas inteligentes es el actuar con autonomía. Sin embargo, la diferencia principal radica en la *conciencia*. Los seres humanos son conscientes de sus acciones mientras las máquinas no. Consecuentemente, desde el campo de la filosofía, se pone en duda el uso del término “autonomía”, dado que se considera que este término sólo puede ser usado en humanos. Sin embargo, nos permitiremos su uso para indicar la forma en que los sistemas de IA imitan al humano, en una especie de nuevo concepto de *autonomía artificial*.

Hoy, los estudios sobre la inteligencia humana han cambiado, dando lugar al concepto de inteligencia múltiple, como la define Howard Gardner, describiendo ocho tipos de inteligencias. Gardner afirma que “la inteligencia no es una sola unidad, sino un conjunto de unidades atómicas relacionadas entre sí” (Gardner, H, 2010. p. 53).

Con los desarrollos actuales de la IA, es posible abordar con éxito una de estas inteligencias a la vez, por ejemplo, existen sistemas que son capaces de interpretar imágenes, otros que interpretan texto, otros sonidos. Muchas de estas actividades son realizadas por la inteligencia artificial con un mejor desempeño de lo que lo hacen los humanos. Se presenta el desafío de desarrollar una inteligencia artificial general (AGI, por su sigla en inglés Artificial General Intelligence), capaz de realizar exitosamente tareas en diferentes dominios, resolviendo problemas (WAN, P. & GOERTZEL, B., 2012) [6]; o con la capacidad de aprender diferentes habilidades y de utilizarlas en distintas circunstancias y ambientes (WAN, P. & GOERTZEL, B., 2012) [7].

Actualmente, es una teoría en desarrollo. Sin embargo, la evolución de la tecnología avanza constantemente y, hoy, ya se encuentran inteligencias artificiales llamadas multimodales, dado que suman dos o más inteligencias. Por ejemplo, la inteligencia artificial puede entender y procesar texto e imágenes a la vez.

En este contexto, un sistema inteligente será **el que actúe en forma autónoma y, con base en datos y entradas recibidas de humanos o máquinas, decida cómo alcanzar su objetivo de diseño, usando técnicas de aprendizaje automático, lógica y conocimiento; generando una recomendación, contenido, pronóstico o decisiones que puedan influir en el entorno donde el sistema interactúa.**

¿Por qué es necesario hablar de la ética en la Inteligencia Artificial?

Los avances tecnológicos, como ahora la IA, le plantean a la sociedad problemas desconocidos y nuevos, para los cuales no existen reglamentaciones o leyes concretas que puedan aplicarse para resolverlos. En este sentido, la ética será la guía que debe usarse para evitar o resolver situaciones difíciles que puedan surgir a partir del uso o desarrollo de estas nuevas tecnologías.

Con la IA, existe la duda, la disyuntiva de si es conveniente regularla antes de que surjan los problemas. Ante la falta de leyes, existen diferentes formas de regular las conductas humanas, tales como: las normas éticas, los usos y costumbres, y los hábitos sociales en las actividades, profesiones, etc.



Laudon & Laudon (2012) definen la ética en los sistemas de información de la siguiente manera [8]:

La ética se refiere a los principios del bien y del mal que los individuos, al actuar como agentes con libre moral, utilizan para guiar sus comportamientos. Los sistemas de información generan nuevas cuestiones éticas tanto para los individuos como para las sociedades, ya que crean oportunidades para un intenso cambio social y, por ende, amenazan las distribuciones existentes de poder, dinero, derechos y obligaciones. Al igual que otras tecnologías, como los motores de vapor, la electricidad, el teléfono y la radio, la tecnología de la información se puede usar para alcanzar el progreso social, pero también para cometer crímenes y amenazar los preciados valores sociales. El desarrollo de la tecnología de la información producirá beneficios para muchos y costos para (p. 124)

Los estados introducen reglas o normas jurídicas sobre temas tecnológicos, entendiendo por tales: a las leyes, decretos, reglamentos, disposiciones, ordenanzas, etc., cuando ha pasado un tiempo, y la jurisprudencia ha resuelto algunos casos en los tribunales, que se litigaron por daños y perjuicios.

En general, organismos internacionales han formulado códigos de ética y recomendaciones para prevenir daños, especialmente aquellos ocasionados por sistemas autónomos basados en inteligencia artificial.

Desde el punto de vista legal, resulta importante tratar los problemas que pueden ir apareciendo. Obviamente, la tecnología *per se* no es “consciente” de sus acciones, por lo cual, la responsabilidad debe buscarse en alguno de estos tres lugares: el creador, la empresa titular, que pone a disposición de los usuarios el producto, y el usuario.

Toda vez que la Universidad, a través de sus profesionales, enseña, produce, innova, es menester entender la responsabilidad de cada rol interviniente. Por las normas de protección al consumidor, la persona que lo utiliza difícilmente carga con la responsabilidad, menos aun cuando se trata de productos nuevos de los que se puede esperar la posibilidad de actividad riesgosa. Quien esté al frente de la empresa que lo pone a disposición del usuario tiene responsabilidad, como con cualquier *software*, sea o no IA. La creación de la obra dependerá del tipo de IA de la que se está hablando. En este tipo de sistemas opera un equipo de profesionales, que incluye los siguientes perfiles: desarrollador, científico de datos, profesional de IA, profesional de calidad del *software*, etc.

Se deberá analizar, entonces, en el caso concreto y sus circunstancias, cuáles son las responsabilidades, agravantes o atenuantes.

Antecedentes

UNESCO

La Conferencia General de la Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura (UNESCO), por su parte, elaboró un documento que expresa recomendaciones para el uso ético de la inteligencia artificial, dado que: el “rápido desarrollo de las tecnologías de la IA complica su aplicación y gobernanza ética”. (UNESCO, 2021, p. 3) [9]. En este documento se establece, por un lado, la necesidad de colaborar y cooperar en la innovación y el desarrollo: “convencida además de que las tecnologías de la IA pueden aportar importantes beneficios” (UNESCO, 2021 p. 3) [10], en cuanto a que su aplicación, en diferentes campos, puede resultar muy útil y beneficiar al ser humano, los ecosistemas y el medio ambiente, pero sin dejar de mostrar su preocupación por el creciente aumento de los aspectos negativos que se han visto durante los últimos años y la necesidad de restringirlos o minimizarlos, mientras se fortalecen sus bondades. Estas recomendaciones están dirigidas a los estados miembros, instándolos a adoptar medidas adecuadas, principalmente legislativas, que efectivicen estas recomendaciones. Entre la definición de sus valores y principios, quedan claras las necesidades de estas regulaciones éticas cuando se afirma lo siguiente:

La fiabilidad y la integridad del ciclo de vida de los sistemas de IA son esenciales para velar por que las tecnologías de la IA estén al servicio del bien de la humanidad, las personas, las sociedades y el medio ambiente y los ecosistemas, y encarnen los valores y principios enunciados en la presente recomendación. (UNESCO, 2021, p.9) [11]



La Unión Internacional de Telecomunicaciones (ITU) lanzó una plataforma llamada IA for Good (IA para el bien)² para fomentar el diálogo en el uso beneficioso de la IA, realizando anualmente encuentros y propiciando desarrollos concretos. Esta plataforma anunció que “propone contenidos y oportunidades de colaboración alineados con cada uno de los 17 ODS”[12]

Declaración de Montevideo sobre IA y su impacto en América Latina

Durante el encuentro latinoamericano de Inteligencia Artificial, realizado en Montevideo, el 10 de marzo de 2023, un grupo de investigadores y desarrolladores, participantes del encuentro, ha elaborado el documento “Declaración de Montevideo sobre Inteligencia Artificial y su impacto en América Latina”, con el fin de manifestar que “Las tecnologías, en general, y los sistemas de Inteligencia Artificial (IA), en particular, deben ser puestos al servicio de las personas. Mejorar la calidad de vida, las condiciones laborales, económicas, de salud y de bienestar general deben ser nuestra prioridad.” (2023) [13]

Disposición 2/2023 de Gabinete de Ministros

La Subsecretaría de Tecnologías de la Información, dependiente de Jefatura de Gabinete de Ministros del Estado Nacional, en la disposición 2/2023, aprueba las “Recomendaciones para una inteligencia Artificial Fiable” [14] y define “el esquema gráfico de las etapas de Proyectos de Inteligencia Artificial”. A través de la misma, el Estado promueve el desarrollo de soluciones mediante el uso de IA y garantiza su “transparencia, equitatividad y responsabilidad”. Dentro de sus recomendaciones, se destaca la necesidad de que el humano sea parte del proceso de toma de decisión, indicando claramente en sus recomendaciones que “un algoritmo no posee autodeterminación y/o agencia para tomar decisiones libremente ... y, por ende, no se le pueden atribuir responsabilidades de las acciones que se ejecutan a través de dicho algoritmo en cuestión”. (2023, p. 7)

En este sentido, el ser humano es el responsable de los resultados de la ejecución de este tipo de sistemas y de las decisiones que se tomen. Este concepto se lo conoce como humano-máquina, “inteligencia-aumentada” o “*human in the loop*”, en donde humanos y máquinas trabajan en conjunto para resolver un problema. En esta línea, se hace explícita la necesidad de instrumentos de control diseñados para identificar la responsabilidad y rendición de cuentas por parte de las personas, indicando que los aspectos éticos son propios de las mismas. Además, las recomendaciones consideran estos aspectos desde diferentes puntos de vistas: ciclo de vida de proyectos de IA, diseño y modelado de datos, diseño de modelos de IA y consideraciones fuera del ciclo de la IA.

OCDE

La Organización para la Cooperación y el Desarrollo Económico (OCDE) ha manifestado su preocupación sobre los avances de la IA y se plantea el interrogante de cómo garantizar o asegurar que la IA sea un beneficio para toda la sociedad. Para dar respuesta a esto, crea un observatorio de IA que se encarga de establecer los principios para que puedan ser compartidos por gobiernos y otros actores, con un enfoque centrado en el humano, (*human-centric* o donde el humano sea el centro) para una IA de confianza. Estos principios incluyen políticas públicas y estrategias recomendadas que puedan ser adoptadas por desarrolladores de IA en el mundo, respetando los derechos humanos y los valores de la democracia [15].

IEEE

La IEEE lanza una segunda versión de su documento *Ethically aligned design*, que es un conjunto de recomendaciones y principio para guiar el desarrollo ético de los sistemas inteligentes y autónomos. En el documento, se resalta la necesidad de priorizar el bienestar humano, la transparencia de los sistemas autónomos e inteligentes y la prevención de sesgos. Se encuentra estructurado en área temáticas en las que se presentan para cada una, principios éticos de alto nivel y se proponen lineamientos prácticos para su implementación. Es un verdadero *framework* que intenta integrar las consideraciones éticas en la estructura misma de los sistemas de IA desde su concepción.

² AI for Good - All Year Always Online (itu.int)



Se destaca la importancia que otorga a los datos en este tipo de sistemas, instando a las personas a tener el control sobre sus datos personales. Otro aspecto importante de este *framework* es la perspectiva hacia los tiempos venideros, intentando anticipar escenarios futuros y sus implicaciones éticas, incluyendo consideraciones de largo plazo y el impacto sistémico de la IA en la sociedad, la economía y la interacción humano-máquina.

HUB de IA en la UTN

La UTN, como parte del sistema científico nacional, y preocupada por la necesidad de estar a la vanguardia en la generación, desarrollo, uso y enseñanza de las nuevas tecnologías, ha elaborado un documento cuyo objetivo es la creación del *HUB* de Inteligencia Artificial [16] en la UTN, abarcando a todas las regionales que tengan experiencia y conocimiento sobre esta área.

Su creación se fundamenta en la necesidad e importancia de contar con centros de investigación que desarrollen, apliquen y generen nuevos conocimientos en el área; articulen con otras disciplinas; vinculen a la universidad con el medio socio-productivo y a la formación de talentos. Tanto en la definición de su misión como visión, así como en sus objetivos, se destaca la necesidad de establecer lineamientos éticos y responsabilidades para los diferentes actores que intervengan.

Se sugiere la creación de cursos centrados en la ética en inteligencia artificial, destacando la relevancia de abordar este tema frente al progreso tecnológico y fomentar la idea de que la tecnología en sí misma no puede ser catalogada como buena o mala, sino que su valor depende del uso que se le dé. Por lo tanto, como institución educativa, la enseñanza del uso de estas nuevas tecnologías, para el bien de todos, es un principio que debe estar presente. El aspecto tal vez más importante de este *framework* es el trabajo que se está llevando a cabo en una serie de estándares que tienen como objetivo trasladar los principios éticos en estándares técnicos concretos.

Ámbito de aplicación de esta recomendación

La presente recomendación tiene su alcance de aplicación en el ámbito de la Universidad Tecnológica Nacional, abarcando todas sus actividades académicas, de extensión, investigación, vinculación, administrativas y de gestión.

Se pueden reconocer diferentes formas de interactuar con sistemas de IA:

- como objeto de estudio en áreas de investigación específicas;
- como herramienta de soporte de actividades en el ámbito de la Universidad;
- como herramienta de desarrollo de nuevos productos o sistemas.

Para cualquier tipo de interacción, estas recomendaciones deberán ser tenidas en cuenta por todos los claustros que forman parte de la comunidad universitaria de la UTN (docentes, no-docentes, estudiantes y personas graduadas).

Clasificación de los datos manipulados en IA y sus riesgos asociados

Los sistemas de IA requieren de grandes cantidades de datos para poder existir, que no sólo son algoritmos. En este punto, es importante entender y clasificar los datos que son utilizados para el proceso de entrenamiento de estos modelos. Estas clasificaciones no son disyuntas, lo cual implica que un cierto conjunto de datos puede pertenecer a más de una de estas categorías. Esta clasificación permitirá entender los riesgos asociados con la utilización de los sistemas de IA, como se detalla al final de esta sección.



- I. **Datos confidenciales:** Refiere a aquellos datos o información sensible que pueden referirse a cuestiones de inteligencia, defensa, seguridad, y otros similares. El Código Civil y Comercial de la Nación, en su art. 992, establece que: si, durante las negociaciones, una de las partes facilita a la otra información confidencial, quien la recibió no debe revelarla ni usarla inapropiadamente en su interés. La parte que incumple este deber queda obligada a reparar el daño sufrido por la otra y, si ha obtenido una ventaja indebida de la información confidencial, queda obligada a indemnizar a la otra parte en la medida de su propio enriquecimiento.
- II. **Datos personales:** Refiere a aquellos datos o información de las personas que específicamente han sido definidos como tales por la normativa vigente (tales como: nombre, apellido, DNI, datos biométricos entre otros). [17]
- III. **Datos internos:** Refiere a aquellos datos o información de gestión interna, que no resultan ser ni confidenciales ni personales, pero que no catalogan como información pública.
- IV. **Datos públicos:** Refiere a aquellos datos o información de dominio público que generalmente se encuentran disponibles, ya sea como conjuntos de datos abiertos, y/o como contenidos en sitios web.
- V. **Datos disponibles en internet:** Es el caso menos costoso, no obstante, se debe considerar que existe un alto grado de probabilidad de que estos sean inexactos, posean sesgos de diferente tipo, puedan someterse a propiedad intelectual; entre aspectos que no solo degradan la calidad de los datos, sino que impiden crear datos de entrenamiento ético.
- VI. **Datos existentes en la organización:** En este caso, se requiere dimensionar los costos asociados: previo a su utilización, se debe tener en cuenta la clasificación según su confidencialidad, derechos de uso, consentimiento de los titulares, posibilidad de anonimizar dichos datos y otros aspectos que establezca la normativa vigente.
- VII. **Datos solicitados a terceras partes:** En este caso, también se deben dimensionar los costos asociados, ya que no están disponibles en internet y, para conseguirlos y utilizarlos, se debe tener en cuenta la clasificación según su confidencialidad, derechos de uso, consentimiento del titular, la trazabilidad de todo el proceso de obtención y creación de datos para entrenamiento y otros aspectos que establezca la normativa vigente.

En función de este agrupamiento de los datos, la Unión Europea establece una clasificación de riesgos posibles en los sistemas con autonomía (o inteligentes) que utilizan datos para aprender a realizar su tarea, lógica y conocimiento:

- I. **Riesgo inadmisibles:** Se considerará riesgo inadmisibles cuando aquellos sistemas: (i) desplieguen técnicas subliminales con el objetivo de causar cambios en conductas de personas y ocasionen daños físicos o psicológicos; (ii) exploten vulnerabilidades de grupos específicos de personas, ya sea por edad, raza, religión, sexo, etc.; (iii) generen perfiles de personas según su comportamiento, que puedan resultar en recibir un trato desproporcionalmente desfavorable al observado. Los sistemas clasificados con esta escala de riesgo serán prohibidos.
- II. **Alto riesgo:** Se considera de alto riesgo un número limitado de sistemas de IA que tienen un impacto negativo en la seguridad de las personas o en sus derechos fundamentales (protegidos por la Carta de los Derechos Fundamentales de la UE), tales como: sistemas de reconocimiento biométrico; sistemas de gestión de infraestructura crítica, como agua, electricidad, tráfico; sistemas para la educación y formación profesional; sistema de selección de personal y gestión de relaciones laborales; sistemas de gestión de acceso por parte de las personas a servicios esenciales, públicos y privados; sistemas de gestión de actividades de las fuerzas de seguridad; sistemas usados en la migración, asilo y control de fronteras; sistemas para la administración de la justicia y procesos democráticos.
- III. **Riesgo limitado:** Son aquellos sistemas de IA a los que se les imponen obligaciones específicas de transparencia. Por ejemplo, los usuarios deben ser conscientes de que interactúan con sistemas de IA y deben poder tomar la decisión de si continuar interactuando o no.
- IV. **Riesgo mínimo:** Todos los demás sistemas de IA pueden desarrollarse y utilizarse con arreglo a la legislación vigente, sin obligaciones jurídicas adicionales. La inmensa mayoría de los sistemas de IA utilizados actualmente en la UE pertenece a esta categoría. De forma voluntaria, los proveedores de estos sistemas pueden optar por aplicar los requisitos de una IA digna de confianza y adherirse a códigos de conducta voluntarios.



Estrategias para el abordaje de la ética en IA desde la UTN

En primer lugar, es importante destacar que la UTN, como institución educativa, debe atender las necesidades de la comunidad para estar a la vanguardia de la tecnología, sin perjuicio de la sociedad como un todo. En este sentido, como lo afirma su estatuto, la institución educativa se ubica de la siguiente manera:

Abierta a todas las personas que sean capaces de contribuir al desarrollo de la economía argentina con conciencia sobre el bienestar, la justicia social, el respeto por la ciencia y la cultura, reivindicando los valores imprescindibles de libertad y dignidad de la persona humana, los cimientos de la cultura nacional que hacen a la identidad del pueblo argentino. (Res. 606/2024, p. 14) [18].

Como Universidad, su misión es preservar y transmitir los productos de los campos científicos, tecnológicos y culturales para la formación plena del ser humano, como sujeto destinatario de esa cultura, y de las técnicas, contribuyendo al desarrollo y transformación de la sociedad. Dentro de sus objetivos está la formación integral de sus graduados, no sólo desde sus aspectos científicos y tecnológicos, sino, también, desde la cultura universal y nacional, y los valores éticos que definen a las personas cabales y solidarias. [19]

Es posible reconocer las repercusiones positivas y negativas profundas que se están planteando sobre el uso, desarrollo y aplicación de sistemas inteligentes que afectan a la sociedad, el medio ambiente, los ecosistemas y la vida humana, debido, en gran medida, a que la utilización de esta tecnología modifica la forma de relacionarse, de actuar, de tomar decisiones, de aprender y enseñar, de pensar, de comunicarnos, etc.

El observatorio de incidentes de la IA de la OCDE afirma que, en los últimos años, con el advenimiento de la IA generativa y modelos grandes de lenguaje (LLM, por su sigla en inglés *Large Language Model*), la cantidad de incidentes reportados se incrementa día a día. Este observatorio utiliza analítica de datos no sólo para analizar y llevar cuentas de los incidentes, sino también para poder predecir posibles sucesos negativos y actuar antes de que estos ocurran.

En este sentido, –y ante los problemas que se están planteando sobre el uso, desarrollo y aplicación de sistemas inteligentes–, se hace imprescindible dictar lineamientos y recomendaciones dentro de la UTN, que se encuentren enmarcados en las normativas nacionales e internacionales sobre la inteligencia artificial, teniendo en cuenta que este campo está en continuo desarrollo y sus aplicaciones son cada vez más extensas.

Objetivos

Definir lineamientos y recomendaciones a utilizarse en la docencia, la investigación, la vinculación y la gestión en el ámbito de la UTN que:

- guíen el desarrollo y la evolución de los sistemas inteligentes (o de las tecnologías que involucren IA) en el marco de la universidad;
- se centren en los derechos humanos, la dignidad humana, la igualdad de género, la justicia social y económica, el desarrollo, el bienestar físico y mental del ser humano, la diversidad, la interconexión, los objetivos de desarrollo sustentable y la inclusión;
- deban ser considerados por aquellos actores dentro de la UTN que lleven adelante proyectos de innovación a través del uso, desarrollo o adopción de tecnologías de IA;
- deban ser considerados por aquellos que adopten este tipo de tecnologías en la docencia universitaria, para adecuar las nuevas didácticas al uso de estas herramientas, tanto en la enseñanza de grado, postgrado, como, así también, en los desarrollos que se realicen, como extensión, vinculación e investigación;
- alcancen los objetivos de la universidad de formar profesionales idóneos que actúen con responsabilidad, ética profesional, sentido crítico y sensibilidad social.



Lineamientos

Para una mejor organización de estos lineamientos, se organizan de acuerdo a la interacción con la IA, como se explicó más arriba: (i) como objeto de estudio e investigación, (ii) como soporte a las actividades desarrolladas y (iii) como herramienta de desarrollo de nuevos productos o sistemas. Un apartado especial amerita lo relativo a la educación, que se considera un pilar fundamental de esta Universidad y en torno a la cual se organizan el resto de las actividades desarrolladas.

Lineamientos para cuando la IA es un objeto de estudio

En este apartado se especifican lineamientos que se aplican a centros y grupos de investigación que desarrolla el área de IA, proponiendo avances y nuevos paradigmas en las diferentes ramas de esta disciplina.

- Los investigadores que utilicen datos, en sus estudios, que no estén disponibles en los repositorios públicos deberán establecer criterios de acceso con el fin de asegurar la privacidad y la integridad de éstos, cuando así se requiera.
- Asimismo, en los desarrollos que se elaboren como pruebas y prototipos, se deberá garantizar la seguridad de la IA, detallando el alcance del desarrollo, los límites, los datos utilizados, las técnicas que se aplicaron y los usos autorizados.
- Establecer objetivos claros en el uso y desarrollo de la IA para poder medir si se alcanzan a lograr y para evaluar sus riesgos. Estos objetivos deben estar alineados al bien común y respetar los derechos humanos y el medio ambiente.
- Los riesgos de ataques y de seguridad deben ser mitigados, prevenidos y eliminados, de manera de no provocar daños no deseados con el uso o desarrollo de la IA.
- Se deben evitar sistemas discriminatorios y sesgados. Para ello, si se trata de un desarrollo por parte de miembros de la comunidad UTN, deberán describir el método empleado para seleccionar datos usados en el entrenamiento de los modelos de IA y el origen de éstos, a fin de garantizar la no violación de los derechos individuales y la privacidad de la información individual.
- Si se trata de usar un sistema de IA desarrollado por terceros, en el ámbito de las investigaciones, se deberá garantizar, a través de pruebas objetivas, que los mismos estén libres de discriminación o sesgos.
- Establecer que la privacidad de los datos constituye un derecho esencial de las personas, necesaria para preservar su dignidad, autonomía y capacidad de actuar. Por lo tanto, debe ser respetada, protegida y promovida a lo largo del ciclo de desarrollo de sistemas de IA.
- Cuando los equipos de investigación realicen transferencias de conocimiento a terceros interesados, como miembros de UTN, tendrán la responsabilidad de transmitir el uso correcto de estas tecnologías, su aplicación al bien común, el respeto por el ser humano y el medio ambiente, la necesidad de garantizar la protección de la información vulnerable y la importancia de analizar sesgos en los datos para eliminarlos antes de que estos sean utilizados para el entrenamiento de la IA.
- Los investigadores que participen de desarrollos que se realicen desde UTN, como vinculación y/o transferencia de conocimiento, y que sean entregados a terceros, no serán responsables del uso malicioso que puedan hacer estas partes. Cuando existe riesgo de mal uso de estos desarrollos, los involucrados deberán elaborar un instructivo indicando los usos correctos para los cuales fue desarrollado.
- Propiciar el estudio e investigación de los usos y desarrollos indebidos o maliciosos de sistemas de IA para prevenirlos y minimizarlos.
- Promover equipos destinados a revisar y supervisar los contenidos generados por IA, con el fin de identificar sesgos e información maliciosa o falsa; de manera de aportar a una IA con equidad y justicia.
- Los actores, que promuevan el uso o desarrollen sistemas de IA, deben revelar y combatir aquellos sistemas de IA que, por negligencia o deliberadamente, fomenten la desigualdad cultural, económica o social, los prejuicios, la propagación de desinformación o información falsa, la vulneración a la libertad de expresión y el acceso a la información.



Lineamientos para cuando la IA es un soporte a las actividades desarrolladas

En este apartado se establecen lineamientos respecto de cuando se utilizan sistemas de IA desarrollados por terceros para dar soporte a actividades de gestión, docencia, investigación, administrativas, extensión, etc., dentro del ámbito de la UTN.

- En caso de usar IA en el proceso de toma de decisiones irreversibles o difíciles de revertir, es obligatorio que el ser humano sea quien tome la decisión. Esto puede darse en situaciones en donde se vea involucrada la vida de seres humanos o se pongan en peligro sus derechos.
- Promover instancias de formación y capacitación de todas las personas en el ámbito de la UTN: docentes, investigadores, no docentes, estudiantes y graduados, para que actualicen sus saberes, en forma permanente, con el objeto de favorecer un uso eficiente, crítico y eficaz de la IA en su desempeño.
- Los sistemas de IA nunca reemplazarán la responsabilidad y la necesidad de rendir cuentas de quienes utilicen sistemas de IA en las actividades de la UTN. Esto implica que las decisiones que tomen los sistemas autónomos serán responsabilidad de su usuario. Es por esta razón que, antes de utilizar sistemas autónomos, es imprescindible analizar sus implicancias, sesgos o errores que estos puedan tener.

Lineamientos para cuando la IA es usada como herramienta de desarrollo de nuevos productos o sistemas

En este apartado, se establecen lineamientos respecto de cuando se utilizan sistemas de IA desarrollados por terceros para extenderlos y/o modificarlos generando nuevos sistemas basados en IA.

- Establecer procedimientos para evaluar riesgos de causar daños a seres humanos en el uso de la IA, sus derechos, sus libertades, los ecosistemas y el medio ambiente.

Lineamientos cuando la IA es usada en educación

Con respecto al uso de la IA en **educación**, particularmente, ya sea en cursos de grado, posgrado, extensión o pregrado, se describen los siguientes lineamientos:

- En la enseñanza de la IA, el personal docente deberá concientizar y sensibilizar sobre el respeto de esta tecnología, el valor por los derechos humanos, su desarrollo para el bien común y la protección del medio ambiente.
- Si lo usan docentes para elaborar guías y material de estudio, debe supervisarse su contenido antes de su disposición a estudiantes, garantizando la ausencia de sesgos de género, racismo, ideologías, etc.
- Si los utilizan estudiantes para la elaboración de informes, o resolución de trabajos prácticos, guías de ejercicios u otra actividad necesaria durante el cursado, debe indicarse la metodología, las herramientas de IA, las condiciones iniciales establecidas, los datos utilizados para el entrenamiento, si corresponde, y el contexto en el que se realizó la actividad. En este punto, se deberán especificar las instrucciones dadas a los sistemas inteligentes para que estos respondan, (en caso de usar *prompt*). Si fueron entrenados específicamente para la realización de tareas, se debe indicar con qué datos y bajo qué circunstancias se realizó dicho entrenamiento. En todos los casos, se deberá garantizar ausencia de sesgos de género, racismo, ideologías, etc.
- En ningún caso, docentes y estudiantes podrán adjudicarse la autoría sobre datos, información, informes u otro material elaborado por sistemas inteligentes. Las respuestas dadas por la IA deben ser identificadas claramente, siguiendo los lineamientos del punto anterior.
- En ningún caso, se admite el uso de sistemas inteligentes autónomos para evaluar los saberes en el estudiantado. Cuando esto implique una decisión final, estableciendo una condición de aprobación, o no, de un espacio curricular, (si se usan estos sistemas), deben acompañarse con la supervisión de docentes responsables.
- La IA no se rige por principios éticos y no puede distinguir entre lo correcto y lo incorrecto, lo verdadero y lo falso. Esta herramienta solo recoge información de las bases de datos y textos que procesa de Internet, así que aprende cualquier sesgo cognitivo que se encuentre en ella. Por lo tanto, es esencial analizar críticamente los resultados que proporciona y compararlos con otras fuentes de información.



Referencias

- [1] [2] Micaela Tec, M. (2023). *Tecnologías, brechas digitales, desigualdades desde la perspectiva de género y diversidad*: Ministerio de Mujeres, Géneros y Diversidad. https://www.argentina.gob.ar/sites/default/files/2020/02/informe-micaela_tec-2022-v5-web.pdf
- [3] RUSSELL & NORVIG. (2010). *Inteligencia Artificial: Un enfoque moderno*. Madrid. Pearson
- [4] SAMUEL, A. L. (1959). "Some Studies in Machine Learning Using the Game of Checkers". *IBM Journal of Research and Development*. 44: 206–226. doi:10.1147/rd.441.0206.
- [5] Gardner, H. (2010). *La inteligencia reformulada: las inteligencias múltiples en el siglo XXI*. Paidós. p. 53.
- [6] WAN, P. & GOERTZEL, B. (2012). Deep Reinforcement Learning as Foundation for Artificial General Intelligence. En *Theoretical Foundations of Artificial General Intelligence* (89 - 91). Paris, France: Atlantis Press.
- [7] WAN, P. & GOERTZEL, B. (2012). A New Constructivist AI: From Manual Methods to Self-Constructive System. En *Theoretical Foundations of Artificial General Intelligence* (148-150). Paris, France: Atlantis Press
- [8] Laudon, K. C., & Laudon, J. P. (2012). "Capítulo 4: Aspectos éticos y sociales en los sistemas de información - Parte uno: Organizaciones, administración y la empresa en red". En *Sistemas de información gerencial* (Traducción de A. V. Romero Elizondo, 12. ed.). México: Pearson Educación.
- [9] [10] [11] UNESCO. Recomendación sobre la ética de la inteligencia artificial. Adoptada el 23 de noviembre de 2021. (Capítulo III). <https://www.unesco.org/es/digital-education/artificial-intelligence>
- [12] Unión Internacional de Telecomunicaciones. (n.d.). IA for Good. <https://www.itu.int/en/ITU-T/AI/Pages/default.aspx>
- [13] Encuentro latinoamericano de Inteligencia Artificial. (Marzo de 2023) "Declaración de Montevideo sobre IA y su impacto en América Latina": Uruguay.
- [14] Jefatura de Gabinete de Ministros, Subsecretaría de Tecnologías de la Información. Disposición N° 2/2023. Anexo I - "Recomendaciones para una Inteligencia Artificial Fiable", punto 3, p. 7 y Anexo II - "Esquema gráfico de las etapas de Proyectos de Inteligencia Artificial que forma parte integrante de la presente medida". 1 de junio de 2023: Argentina. <https://www.boletinoficial.gob.ar/detalleAviso/primera/287679/20230602>
- [15] OCDE. Recomendación sobre los Principios de la IA, aprobada el 22 de mayo de 2019, por el Consejo de Ministros de la OCDE – Organización para la Cooperación y el Desarrollo Económicos. <https://legalinstruments.oecd.org/fr/instruments/OECD-LEGAL-0449>
- [16] Secretaría de Cultura y Extensión Universitaria. (2023). "HUB DE INTELIGENCIA ARTIFICIAL": Universidad Tecnológica Nacional.
- [17] VERCELLI, A. (2019). Facebook Inc. Cambridge Analytica: (des)protección de datos personales, *EJS* 18 (2) 57-70
- [18] [19] Resolución 606 de 2024 [CONSEJO SUPERIOR DE LA UTN]. Por medio de la cual se aprueba la revisión de la Disposición 2/2023, del 1 de junio de 2023, Anexo I - "Recomendaciones para una Inteligencia Artificial Fiable" y Anexo II - "Esquema gráfico de las etapas de Proyectos de Inteligencia Artificial". (p. 14). <http://csu.rec.utn.edu.ar/CSU/RES/2024/606.pdf>



**UNIVERSIDAD
TECNOLOGICA
NACIONAL**

© 2024 UTN. Todos los derechos reservados.