

RIESGOS, DILEMAS ÉTICOS Y BUENAS PRÁCTICAS EN INTELIGENCIA ARTIFICIAL

Dra. Bibiana Beatriz Luz Clara; Lucía Rosario Malbernat
Grupo Ética-IA, Departamento de Sistemas, Universidad CAECE
bluzclara@ucaecemdp.edu.ar; lmalbernat@ucaecemdp.edu.ar

RESUMEN

Mientras que Europa publica su primera guía ética para el uso responsable de Inteligencia Artificial (IA), en la región hay necesidad de contar con una guía de pautas deseables y prácticas preferidas por las organizaciones que propicien la definición de buenas prácticas.

Por lo tanto, es necesario elaborar directrices éticas que se basen en el marco regulador existente y aporten a los desarrolladores, proveedores y usuarios de IA orientaciones marco.

Se comparte en este documento un Proyecto de investigación de reciente presentación que pone su foco en las prácticas que llevan a cabo las organizaciones de la región que aplican técnicas de IA en sus desarrollos o procedimientos, de modo de aportar conocimiento para minimizar sus riesgos éticos.

Palabras clave: Inteligencia Artificial, ética, Buenas prácticas

CONTEXTO

La investigación que se reporta está radicada en el Departamento de Sistemas de la Universidad CAECE, República Argentina, presentada para el período 2021-2022 con fecha de inicio 1° de abril de 2021.

Se trata de un Proyecto inter sedes ya que participarán investigadores radicados en la Sede Central y en la Subsede Mar del Plata que tendrá impacto en carreras de grado y posgrado del Departamento. En particular, investigadores del proyecto esperan poder volcar el conocimiento adquirido en las asignaturas Aspectos Profesionales para Ingeniería e Inteligencia de Negocios.

1. INTRODUCCIÓN

Si bien el concepto de Inteligencia artificial, acuñado en 1956 por John McCarthy en la que se considera primer conferencia dedicada al tema (Dartmouth College, Hanover, Nuevo Hampshire, Estados Unidos) no es nuevo y hace más de 30 años que la industria informática apuesta a la inclusión de técnicas propias de esta disciplina a sus desarrollos, en la última década empiezan la organizaciones en general a obtener provecho de sus ventajas y beneficios por lo que comienzan a impactar en la sociedad y a tomar cuerpo los dilemas éticos.

Para dar respuesta a ese contexto los países alrededor del mundo comienzan, en el último quinquenio, a definir sus estrategias nacionales o planes de desarrollo. Tal el caso de EEUU, que lo hizo en 2016, Canadá y China en 2017, Japón y Francia en 2018 [1] mientras que el gobierno de España presentó en 2020 el Consejo Asesor de Inteligencia Artificial, un órgano consultivo destinado a asesorar y a realizar recomendaciones independientes sobre las medidas adoptables para garantizar un uso seguro y ético de la IA. [2]

Para la Comisión Europea, la IA son “Sistemas de software (y posiblemente también de hardware) diseñados por humanos que, dado un objetivo complejo, actúan en la dimensión física o digital percibiendo su entorno a través de la adquisición de datos, interpretando los datos estructurados o no estructurados recolectados, argumentando sobre el conocimiento o procesando la información, derivada de estos datos y decidiendo las mejores acciones para lograr el objetivo dado. Los sistemas de IA pueden usar reglas simbólicas o aprender un modelo numérico, y también pueden adaptar su comportamiento analizando cómo el entorno es afectado por sus acciones previas”. [3]

Esta definición ha sido sugerida recientemente por el Grupo de expertos de

alto nivel independiente sobre inteligencia artificial (AI HLEG) con el fin de aclarar aspectos de la IA como disciplina científica y como tecnología con el objetivo de evitar malentendidos, para lograr un conocimiento común compartido de la IA que pudiera ser utilizado de manera fructífera también por los no expertos en el tema y para proporcionar detalles útiles que se puedan utilizar en la discusión sobre sus pautas de ética y recomendaciones de políticas. [3]

Un informe del Centro Común de Investigación (JRC) al servicio de ciencia y conocimiento de la Comisión Europea del año 2020 da cuenta de la ausencia de una definición de IA formal acordada en común pero toma la definición anterior como referencia y la compara con una colección de definiciones desarrolladas desde 1955 hasta 2019 que incluye más de 50 versiones, al tiempo que aporta una visión general completa sobre el dominio de la IA y propone una taxonomía y palabras clave representativas. [4]

Los aportes valiosos de este trabajo son: la colección de definiciones desarrolladas entre 1955 y 2019; el resumen de las principales características del concepto de inteligencia artificial tal como se refleja en la literatura relevante; y el desarrollo de un proceso replicable que pueda proporcionar una definición dinámica y taxonomía de la IA.

Además de la Unión Europea, otras organizaciones internacionales como el BID, el Foro Económico Mundial (FEM), el IEEE, la OCDE, la Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura (UNESCO) o el Instituto de Ética en la IA de la Universidad de Oxford, están trabajando en la formulación de lineamientos, guías y herramientas para que los países promuevan el desarrollo de sistemas autónomos centrados en los derechos humanos. Sin embargo, América Latina se encuentra aún en una etapa temprana en el desarrollo de guías, protocolos, evaluaciones, marcos regulatorios y procesos de vigilancia social en temas de ética de IA. [5].

Un Informe de fAIR LAC, red de profesionales y expertos académicos,

gubernamentales, de la sociedad civil, la industria y el sector emprendedor cuyo objetivo es promover la aplicación ética de la IA en América Latina y el Caribe, señala en sus conclusiones que los países estudiados (Argentina, Brasil, Chile, Colombia, Costa Rica, Ecuador, México, Paraguay, Perú e Uruguay) presentan esfuerzos incipientes en cuanto a conversaciones y debates sobre privacidad, rendición de cuentas, seguridad, transparencia y explicabilidad, justicia y discriminación, responsabilidad profesional, promoción de valores humanos e impacto ambiental.[5]

La inteligencia artificial es un campo de las ciencias de la computación que investiga y desarrolla sistemas capaces de realizar tareas inteligentes, es decir, aquellas que se suelen atribuir a la inteligencia cognitiva de los seres humanos, tales como aprender, resolver problemas y tomar decisiones. En la actualidad, a través del machine learning y el deep learning, se busca, además, que los sistemas realicen dichas tareas de forma autónoma [2].

El hombre de hoy está tomando, en el entorno de la nueva economía, un papel cada vez más activo, buscando nuevos desarrollos, basados en investigación, que muchas veces integran IA con la finalidad de lograr mayor eficacia y autonomía en los procesos, así como el mejoramiento de la calidad de vida de la sociedad. En ese contexto, es preciso analizar y estudiar el marco normativo para un desarrollo adecuado, que esté fundado en la ética y valores humanos, lo que tal vez implica poner algunos límites al avance de la IA.

De hecho, los avances en Inteligencia Artificial plantean un enorme desafío ya que esta disciplina toca muchos aspectos de la ciencia, la tecnología y, especialmente, por lo dicho, de la sociedad, por lo que su estudio requiere un enfoque interdisciplinario y su aplicación una visión ética en todas sus perspectivas donde el ser humano ocupe el lugar central.

El uso responsable de la IA va más allá de no ejercer prácticas ilegales mediante su uso; se trata de usar la IA de una manera que no

vulnere a minorías, que evite violaciones a derechos humanos y que no lleve al incremento de la brecha de desigualdad existente, ya sea de manera intencional o accidental. Dentro de las consecuencias que trae consigo el uso de la IA existen riesgos intrínsecos, directamente relacionados con datos, y extrínsecos, vinculados con la adopción de la IA en la sociedad. [6]

Un reporte desarrollado por la organización mexicana C Minds y comisionado por el Centro de Estudios en Tecnología y Sociedad (CETyS) de la Universidad argentina de San Andrés identifica 4 tipos de riesgo intrínsecos (que dependen exclusivamente de los datos) en las etapas de planeación, programación e implementación para un uso responsable de la IA: justicia e inclusividad; fiabilidad y seguridad del sistema; privacidad y seguridad de los datos del usuario y transparencia y rendición de cuentas. En relación a los riesgos extrínsecos, hace referencia a la automatización como causante de cambios en los puestos de trabajos. [6]

El informe “Artificial Intelligence Index Report 2019”, del Human-Centered Artificial Intelligence Institute de la Universidad de Stanford [7] señala que el 58% de las grandes empresas encuestadas informaron haber adoptado Inteligencia Artificial en, al menos, una función o unidad de negocio en 2019, frente al 47% en 2018. El informe aborda también la percepción de la sociedad y las preocupaciones existentes con respecto al avance de la inteligencia artificial y señala que los países están empezando a introducir la regulación de la inteligencia artificial en sus agendas pero no se menciona la aprobación de leyes o regulaciones concretas. Además, en cuanto a las preocupaciones sobre principios éticos en inteligencia artificial, indica que se destacan la equidad, la interpretabilidad y la explicabilidad.

El informe reporta que se revisaron 3.600 artículos de prensa a nivel mundial y que se encontró que las preocupaciones pivotan en torno a los marcos legislativos, la privacidad de los datos, el uso de reconocimiento facial, el sesgo de los algoritmos y el papel de las grandes tecnológicas.

2. LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

Es posible que la falta de regulación de la Inteligencia Artificial provoque demoras en la investigación y desarrollo. Es fundamental determinar parámetros de actuación y fijar límites, condiciones y responsabilidades dejando en claro si las máquinas inteligentes pueden considerarse jurídicamente responsables por sus acciones ya sea de forma parcial o total.

Se ha definido, en consecuencia, como Problema a investigar cómo se están adoptando tecnologías y técnicas propias de la Inteligencia Artificial en las organizaciones de la Región, los resguardos que se toman y el marco jurídico con el que cuentan para hacerlo.

Si bien existen normas jurídicas que tipifican los delitos informáticos, protegen datos o algoritmos y múltiples regulaciones locales, es necesario conocer la normativa que regula el desarrollo y utilización de actividades propias de la IA así como también las normas técnicas que configuran estándares para el mercado.

Las normas ISO sobre inteligencia artificial [8], [9], [10] la define de forma un poco vaga, con la finalidad de no limitarse, como: “la capacidad de una unidad funcional para realizar funciones que generalmente están asociadas con la inteligencia humana, como el razonamiento y el aprendizaje.” En el ámbito de la Unión Europea, se propone el cumplimiento de estos objetivos a través de la aplicación de normas internacionales, de la creación de un código de Conducta, de la revisión de las reglas de responsabilidad vigentes y de la evaluación del impacto social que la IA pueda llegar a tener. Se propone, como una de las primeras actividades del Proyecto, relevar las normativas vigentes en la región, que se vinculen con cualquier etapa de la producción y uso de aplicaciones.

Dentro de las ciencias de la computación, la IA, como se apuntó en la Introducción, constituye una categoría cuya finalidad es diseñar sistemas inteligentes y autónomos, basados en el conocimiento humano. El

concepto de IA se aplica a los sistemas que manifiestan un comportamiento inteligente, capaces de analizar su entorno y pasar a la acción con cierto grado de autonomía, con el fin de alcanzar objetivos específicos. En la región se realizan desarrollos y se utilizan aplicaciones con tales características. Se propone como otro de los objetivos del proyecto y primeras actividades, conocer el panorama de desarrollo y aplicación de IA en las organizaciones de la región, identificando quienes llevan a cabo el desarrollo, con qué fines y contraparte, quienes son los usuarios externos o internos, etc..

La incorporación de Inteligencia Artificial a los procesos organizacionales impulsa a dotar de valores y principios a la aplicación tecnológica y a comprender las implicaciones morales y éticas que conllevan. En este sentido, se plantea, también, que es necesario, además, relevar las preocupaciones de las organizaciones sobre principios éticos aplicables a la Inteligencia Artificial.

La Enciclopedia de IA de Shapiro [11] señala que la IA se ocupa del descubrimiento de modelos computacionales que se comportan de la misma manera que lo hacen los humanos y, por lo tanto, de las representaciones y los procesos que se definen utilizando las herramientas computacionales disponibles. Los riesgos e impactos que pueden derivar de la utilización de sistemas provistos de IA llevan a preguntarse por la responsabilidad civil contractual y extracontractual que debe considerarse dentro de diversos ámbitos y, en ese sentido, se estableció otro objetivo con el fin de conocer los resguardos que toman las empresas para evitar que la IA sea utilizada con fines ilícitos o contra la ética en todas las etapas de los procesos vinculados a su utilización y de tipificar tales riesgos.

En particular, los riesgos que aparecen ante la falta de explicabilidad o transparencia ya que la IA trae consigo riesgos que pueden derivarse de defectos, errores o sesgos en la programación que atenten contra la privacidad, intimidad, dignidad, salud o patrimonio de las personas y la responsabilidad eventualmente podría estar

distribuida entre distintos actores haciendo más complejo el problema. En este aspecto surge otro objetivo tendiente a averiguar si las organizaciones que producen las aplicaciones y programan los algoritmos, toman, también, recaudos, ya sean las mismas u otras distintas a las que las aplican.

Se llevará a cabo una investigación cualitativa, con una etapa exploratoria que tomará por unidad de análisis, primero, el corpus jurídico vinculado a la IA y luego las organizaciones que desarrollan o aplican IA en la región en función de los objetivos planteados y previamente descriptos en este trabajo.

Entre las actividades propuestas en el marco del proyecto está previsto el diseño de cuestionarios para la recolección de datos en organizaciones ya mencionadas, a ser aplicados entre usuarios dentro y fuera de las organizaciones (desarrolladores, usuarios, clientes, proveedores, etc.)

3. RESULTADOS ESPERADOS

En conformidad con lo planteado en los objetivos específicos del proyecto se espera:

- a) Identificar normas, protocolos, guías, etc que se estén aplicando en la región en relación a las prácticas vinculadas con el desarrollo y aplicación de IA
- b) Identificar el contexto
 - a. quiénes desarrollan IA,
 - b. qué técnicas aplican
 - c. con qué fines,
 - d. con qué contrapartes,
 - e. usuarios internos y externos alcanzados, etc.,
- c) recoger valores y principios de responsabilidad corporativa,
- d) reconocer normas de conducta que se aplican en torno al diseño, desarrollo y aplicación de las tecnologías IA
- e) e identificar dilemas éticos y prejuicios.

La información que se releve y el conocimiento que se genere podrá ser un insumo para el desarrollo de una guía de pautas deseables y prácticas preferidas por las organizaciones que permitan la definición de buenas prácticas orientadoras

Se espera que los logros alcanzados por este Proyecto, además de generar conocimiento empresarial útil para las organizaciones que se estudien y el mercado informático, impacten positivamente en la carrera de grado Ingeniería en Sistemas y demás carreras del Departamento, en especial en las asignaturas Inteligencia de Negocios y Aspectos profesionales para Ingeniería.

4. FORMACIÓN DE RECURSOS HUMANOS

El grupo de investigación Etica-IA está integrado por profesores del Departamento de Sistemas, docentes de la asignatura Inteligencia de Negocios y de Aspectos Profesionales para Ingeniería de la carrera de grado Ingeniería en Sistemas, asignaturas que toman como insumo académico lo producido en el marco de los Proyectos de Investigación que desarrollan.

Las investigaciones en curso del Departamento de Sistemas son propuestas cada año a los estudiantes como espacios de desarrollo de sus trabajos finales. Con este Proyecto en particular, se espera, además, aportar organizaciones dónde los estudiantes avanzados puedan realizar sus prácticas profesionales.

Las tareas propuestas en el proyecto para el primer trimestre incluyen la incentivación de estudiantes a integrar el Proyecto de investigación en carácter de investigador inicial.

5. BIBLIOGRAFÍA

[1] Mantilla, Sofia E. “Hacia una Estrategia Nacional de Inteligencia Artificial. Instituto de Estrategia Internacional, 2018.

[2] Ministerio de Asuntos Económicos y Transformación Digital, “ENIA, Estrategia Nacional de Inteligencia Artificial” Versión 1.0, 2020

[3] High-Level Expert Group on Artificial Intelligence “A Definition of AI: Main Capabilities and Disciplines” European Commission, 2019

[4] Samoili, S., López Cobo, M., Gómez, E., De Prato, G., Martínez-Plumed, F., and

Delipetrev, B., AI Watch. Defining Artificial Intelligence. Towards an operational definition and taxonomy of artificial intelligence, EUR 30117 EN, Publications Office of the European Union, Luxembourg, 2020, ISBN 978-92-76-17045-7, doi:10.2760/382730, JRC118163.

[5] Constanza Gómez Mont; Claudia May Del Pozo; Cristina Martínez Pinto; Ana Victoria Martín del Campo Alcocer. “La Inteligencia Artificial al Servicio del Bien Social en América Latina y el Caribe: Panorámica Regional e Instantáneas de doce Países”, 2020.

[6] Gómez Mont, C.; Del Pozo, C.; Martín del Campo, A. V. (2020). Economía de datos e inteligencia artificial en América Latina. Oportunidades y riesgos para un aprovechamiento responsable. En C. Aguerre, (Ed.). Inteligencia Artificial en América Latina y el Caribe. Ética, Gobernanza y Políticas. Buenos Aires: CETyS Universidad de San Andrés.

[7] Raymond Perrault, Yoav Shoham, Erik Brynjolfsson, Jack Clark, John Etchemendy, Barbara Grosz, Terah Lyons, James Manyika, Saurabh Mishra, and Juan Carlos Niebles, “The AI Index 2019 Annual Report”, AI Index Steering Committee, Human-Centered AI Institute, Stanford University, Stanford, CA, December 2019.

[8] ISO 10218–1:2011, Robots and robotic devices — Safety requirements for industrial robots.

[9] ISO 13855, Safety of machinery, Positioning of safeguards with respect to the approach speeds of parts of the human body.

[10] ISO/TS 15066:2016 Robots and robotic devices — Collaborative robots.

[11] Shapiro, Stuart C.: Encyclopedia of artificial intelligence (2nd Ed.). John Wiley & Sons, Inc., New York, 1992.