



JOURNAL
ETHICS,
ECONOMICS
AND COMMON GOODS

N° 21 (2), JULY - DECEMBER 2024.

EE & CG
ETHICS,
ECONOMICS
COMMON
GOODS

JOURNAL
ETHICS,
ECONOMICS
AND COMMON GOODS

N° 21 (2), JULY-DECEMBER 2024.

EE & **CG**
ETHICS,
ECONOMICS
COMMON
GOODS

Journal Ethics, Economics & Common Goods, Vol.21, No. 2 July- December 2024 biannual publication edited by the Universidad Popular Autónoma del Estado de Puebla A. C, calle 21 sur 1103, Col. Santiago, C.P 72410, Puebla, Puebla. Tel. (222) 2299400, <https://journal.upaep.mx/index.php/EthicsEconomicsandCommonGoods>. Director: María Teresa Herrera Rendón-Nebel. Co-Editor: Shashi Motilal. Exclusive use rights reserved No. 04-2022-071213543400-102, ISSN 2954 - 4254, both granted by the Instituto Nacional del Derecho de Autor. Technical responsible: Ana Xóchitl Martínez Díaz.

Date of last modification: Marzo 5, 2025.

ISSN: 2954-4254

ESSENTIAL IDENTIFICATION

Title: Journal Ethics, Economics and Common Goods

Frequency: Bi-annual

Dissemination: International

ISSN online: 2954 - 4254

Place of edition: Mexico

Year founded: 2003

DIRECTORY

Editor

María Teresa Herrera Rendón Nebel
Universidad Popular Autónoma del Estado de Puebla. *México*
Facultad de Contaduría y Finanzas

Design

Ana Xóchitl Martínez Díaz

Editorial board

Jérôme Ballet. Université de Bordeaux.
France

Shashi Motilal. Université of Delhi. *India*

Mathias Nebel. Universidad Popular Autónoma del Estado de Puebla. *México*

Patrizio Piraino. University of Notre Dame. *United States of America*

GENERAL INFORMATION

The Journal Ethics, Economics and Common Goods aims to be a space for debate and discussion on issues of social and economic ethics. Topics and issues range from theory to practical ethical questions affecting our contemporary societies. The journal is especially, but not exclusively, concerned with the relationship between ethics, economics and the different aspects of common goods perspective in social ethics.

Social and economic ethics is a rapidly changing field. The systems of thought and ideologies inherited from the 20th century seem to be exhausted and prove incapable of responding to the challenges posed by, among others, artificial intelligence, the transformation of labor and capital, the financialization of the economy, the stagnation of middle-class wages, and the growing ideological polarization of our societies.

The Journal Ethics, Economics and the Common Goods promotes contributions to scientific debates that combine high academic rigor with originality of thought. In the face of the return of ideologies and the rise of moral neopharisaisms in the Anglo-Saxon world, the journal aims to be a space for rational, free, serious and open dialogue. All articles in the journal undergo a process of double anonymous peer review. In addition, it guarantees authors a rapid review of the articles submitted to it. It is an electronic journal that publishes its articles under a creative commons license and is therefore open access.

Research articles, research reports, essays and responses are double-blind refereed. The journal is bi-annual and publishes two issues per year, in July and December. At least one of these two issues is thematic. The journal is pleased to publish articles in French, English and Spanish.

SCIENTIFIC BOARD

Alain Anquetil. ESSCA. France
Alejandra Boni. Universitat Politècnica de València. España
Andrew Crabtree. Copenhagen Business School. Denmark
Byaruhanga Rukooko Archangel. Makerere University. Uganda
Clemens Sedmak. University of Notre Dame. United States of America
David Robichaud. Université d'Ottawa. Canada
Demuijnck Geert. EDHEC Business School. France
Des Gasper. International Institute of Social Studies. Netherlands
Flavio Commin. IQS School of Management. España
François- Régis Mahieu. Fonds pour la recherche en éthique économique. France
Felipe Adrián Vásquez Gálvez. Universidad Autónoma de Ciudad Juárez. México
Javier María Iguñiz Echevarría. Universidad Pontificia de Lima. Perú
Jay Drydyk. Carleton University. Canada
Jean Marcel Koffi. Université de Bouaké. Côte d'Ivoire
Jean-Luc Dubois. Institute de recherche sur le Développement. France
John Francis Díaz. Chung Yuan Christian University. Taiwan
Luigino Bruni. Università Lumen y Sophia. Italia
Mahefasoa Randrianalijaona. Université d'Antananarivo. Madagascar
Marianne Camerer. University of Capetown. South Africa
Mario Biggeri. Università di Firenze. Italia
Mario Maggioni. Università Cattolica del Sacro Cuore. Italia
Mario Solis. Universidad de Costa Rica. Costa Rica
Michel Dion. Université de Sherbrooke. Canada
Mladjo Ivanovic. Northern Michigan University. United States of America
Óscar Garza Vázquez. Universidad de las Américas Puebla. México
Óscar Ibáñez. Universidad Autónoma de Ciudad Juárez. México
Patrick Riordan. University of Oxford. United Kingdom
Pawel Dembinski. Université de Fribourg. Switzerland
Pedro Flores Crespo. Universidad Autónoma de Querétaro. México
Prodipto Ghosh. The Energy and Resources Institute. India
Rebecca Gutwald. Ludwig-Maximilians Universität. Deutschland
Sandra Regina Martini. Universidade Ritter. Brasil
Sara Balestri. Università Cattolica del Sacro Cuore. Italia
Simona Beretta. Università Cattolica del Sacro Cuore. Italia
Stacy Kosko. University of Maryland. United States of America
Steve Viner. Middlebury College. United States of America
Volkert Jürgen. Hochschule Pforzheim. Deutschland

INDEX

RESEARCH ARTICLES

- p. 8-22 Decisiones en conflicto con la Inteligencia Artificial
Damian Emilio Gibaja Romero
- p. 23-46 Entrepreneurial action for the common good
Guillermo J. Larios-Hernandez
- p. 47-61 Thomas Piketty and the Natural Rights Argument for Equality
John D. Feldmann

ESSAYS

- p.63-71 Juvenile Crime, Juvenile Justice and the Collective Social Good
Shashi Motilal, Tarang Kapoor
- p. 72-78 Repensar la IA a través de la mirada humana
Laura Trujillo Liñán

RESEARCH ARTICLES

Decisiones en conflicto con la Inteligencia Artificial

Damian Emilio Gibaja Romero. Director Académico del Área de Matemáticas de la Universidad Popular Autónoma del Estado de Puebla.

Orcid: 0000-0002-3536-4117

Resumen

Los avances tecnológicos han permitido usar inteligencia artificial para analizar y resolver problemas antes considerados complejos. Gracias al incremento del poder computacional, actualmente es posible acceder a herramientas de inteligencia artificial para desarrollar actividades artísticas, productivas, económicas o recreativas de manera más eficiente. Por ello, las inteligencias artificiales son cada vez más utilizadas en procesos de toma de decisiones. Aunque el avance en las inteligencias artificiales ha permitido que tomen decisiones certeras, la automatización en la toma de decisiones ha generado cuestionamientos sobre las implicaciones éticas y sociales de su despersonalización. El presente ensayo analiza los conflictos que pueden surgir cuando tomadores de decisiones son sustituidos total o parcialmente por inteligencias artificiales para resolver un problema común. Por medio de la teoría de juegos se muestra que la automatización de las decisiones puede llevar al Dilema del Prisionero donde el bienestar social no sea el máximo posible. También, el uso asimétrico de estas herramientas pone en riesgo a quienes no las utilizan ya que las inteligencias artificiales pueden tomar ventaja de la información que se les comparte. Entonces, es necesario diseñar algoritmos que internalicen el impacto social, y mecanismos que regulen la forma en qué se usan.

Palabras clave: Juegos no Cooperativos, Elección Racional, Egoísmo.

Abstract

Technological advances have made it possible to use artificial intelligence to analyze and solve problems that were previously considered complex. Thanks to the increased computing power, it is now possible to access artificial intelligence tools to develop artistic, productive, economic or recreational activities more efficiently. Therefore, artificial intelligence is increasingly used in decision-making processes. Although advances in artificial intelligence have allowed them to make accurate decisions, the automation of decision-making has raised questions about the ethical and social implications of their depersonalization. This essay analyzes the conflicts that may arise when artificial intelligence totally or partially replaces decision-makers to solve a common problem. Following a game-theoretical approach, we show that the automation of decisions can lead to a Prisoner's Dilemma where social welfare is not the maximum possible. Also, the asymmetric use of these tools puts those who do not use them at risk since artificial intelligence can exploit the information shared with them. Therefore, it is necessary to design algorithms that internalize the social impact and mechanisms that regulate their use.

Keywords: Non-Cooperative Games, Rational Choice, Egoism.

JEL: C70, C72, D69, D74.

Introducción

Una de las principales características de la inteligencia artificial es su capacidad para tomar decisiones sobre problemas complejos asociados a un gran volumen de datos (Duan, Edwards y Dwivedi, 2018). Algoritmos de inteligencia artificial (IA), como las redes neuronales, proporcionan predicciones certeras de la demanda de un producto con lo cual se mejora la planeación de los procesos productivos (Feizabadi, 2022). También, existen algoritmos con la capacidad de clasificar y agrupar variables, con características afines, para una selección de proveedores más certera en las cadenas de suministro (Chai y Ngai, 2020). Sin embargo, la aplicación de la inteligencia artificial no se limita a procesos productivos; las personas también pueden mejorar la organización de sus actividades semanales con ayuda de este tipo de algoritmos (Venkatesh, 2022).

Aunque sus beneficios son claros, definir qué es una inteligencia artificial es un proceso que continúa en evolución. Para Turing, una IA es una máquina capaz de imitar el comportamiento humano (Muggleton, 2014, p.1). Por su parte, la Organización para la Cooperación y Desarrollo Económico define a una IA como un sistema que puede tomar decisiones cuando un humano (o usuario) le indica una serie de objetivos (Suleimenov et. al, 2020, p.23). Puesto que definir a la IA no es el objetivo del presente trabajo, consideramos que una IA es un mecanismo capaz de tomar la mejor decisión con respecto a la información que tiene disponible y al poder computacional que posee para transformarla en conocimiento (Lohani, Rahamn & Shaturaev, 2023, pp. 4-5). En otras palabras, podemos decir que una IA es un tipo de agente económico con racionalidad limitada desde la perspectiva de Herbert Simon (Kalantari, 2010). Así, el impacto de las IAs se ha estudiado en el mercado laboral por su capacidad para automatizar procesos, transformar empleos existentes y desplazar trabajadores (Ramos, Garza-Rodríguez & Gibaja-Romero, 2022). Al mismo tiempo, las decisiones de las IAs están generando dilemas éticos cuando las IAs toman decisiones que pueden provocar la muerte de algún involucrado (Ashok, et. al, 2022). Por consiguiente, existen dudas sobre si estos algoritmos deben sustituir total o parcialmente a tomadores de decisiones, particularmente, cuando estos interactúan en situaciones de conflicto (Zhang, Chen, & Xu, 2022). Por lo anterior, el presente ensayo se enfoca en responder la pregunta ¿cuáles son las consecuencias de que la IA sustituya a tomadores de decisiones humanos cuando las decisiones impactan a todos los agentes involucrados en un problema común? Por medio de teoría de juegos damos respuesta a la pregunta anterior al analizar dos interacciones: (a) La primera considera dos IAs que interactúan entre sí, es decir, un juego donde los tomadores de decisiones fueron sustituidos por una IA para resolver un problema común; (b) La segunda analiza una interacción donde solo uno de los dos tomadores de decisiones fue sustituido por un algoritmo de IA. En ambos casos, para ilustrar el conflicto, consideramos que las interacciones son simul-

táneas; así, usamos al equilibrio de Nash como concepto de solución.

La primera interacción se asemeja al Dilema del Prisionero pues su equilibrio de Nash es Pareto inferior ya que existe otro escenario donde los involucrados incrementan su utilidad sin perjudicar a los demás. Sin embargo, el mejor escenario no es elegido debido a que las IAs buscan unilateralmente la mejor solución posible con la información que tienen disponible. Respecto al segundo escenario, se observa que la IA afecta al tomador de decisiones que no utiliza IA pues tiene una menor capacidad para obtener la mejor decisión cuando se le compara con una IA.

El presente ensayo se relaciona estrechamente con una creciente literatura preocupada por los dilemas sociales asociados al uso de IA. Por ejemplo, al tratar de tomar decisiones racionales, la IA aplicada a vehículos autónomos genera el dilema de si debemos salvar a los peatones o a los pasajeros ante un posible accidente (Bonneton, Shariff, & Rahwan, 2016). También, la IA puede propagar información falsa después de recibir cierto entrenamiento y, con ello, afectar la reputación de los agentes económicos involucrados (Strümke, Slavkovik, & Madai, 2021; Sullivan & Fosso Wamba, 2022). Por otra parte, el segundo juego proporciona un equilibrio donde el tomador de decisiones tradicional se ve perjudicado por una inteligencia artificial no cooperativa. Es decir, si el uso de IA no es universal, la implementación de este tipo de algoritmos puede incrementar la desigualdad entre los agentes económicos debido a que las decisiones tienen diferentes niveles de optimalidad (Lutz, 2019).

El ensayo se estructura de la siguiente forma. La segunda sección describe los elementos generales del conflicto entre tomadores de decisiones tradicionales e IAs. Posteriormente, la tercera sección analiza los equilibrios de las interacciones consideradas. Así, la discusión de los resultados se presenta en la cuarta sección. La última sección presenta las conclusiones del trabajo.

Modelo

El presente ensayo usa la Teoría de Juegos para analizar las implicaciones de sustituir tomadores de decisiones en conflicto por inteligencias artificiales. Un juego es un modelo matemático que permite estudiar la toma de decisiones estratégica. Los elementos básicos de este tipo de modelos son los jugadores, acciones, reglas y pagos (Beckenkamp, 2006). A continuación, describimos dichos elementos para analizar el conflicto entre IAs y tomadores de decisiones tradicionales.

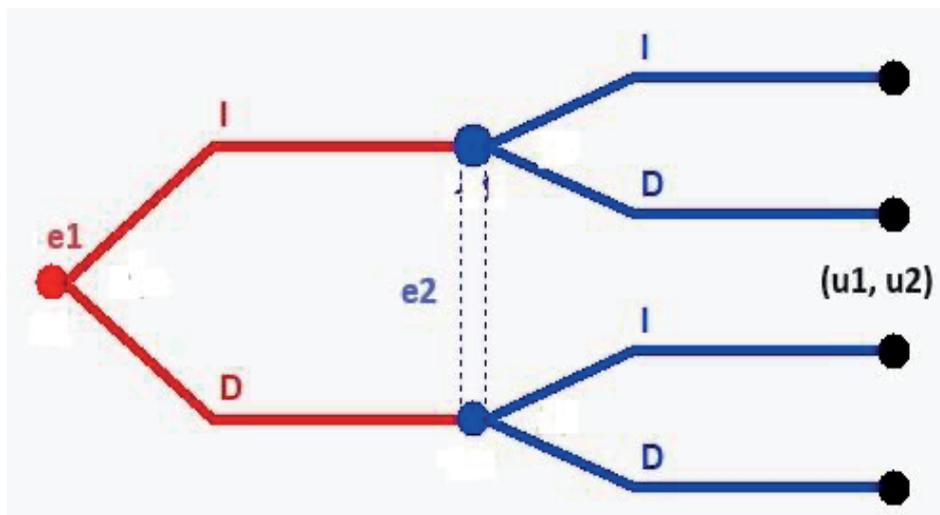
Sea $J = \{e_1, e_2\}$ el conjunto de jugadores; estos pueden ser de dos tipos: agentes económicos tradicionales (individuos o empresas) y algoritmos de inteligencia artificial capaces de tomar decisiones racionales, es decir, que elijan la mejor opción de entre todas las alternativas posibles que tienen con la información disponible (Lohani, Rahman, y Shaturaev, 2023). Un jugador genérico se denota por e^i . Sin importar el tipo de jugador, asumimos que el conjunto de acciones del jugador e^i es $A^i = \{I, D\}$. La acción I representa implementar una solución indirecta donde se comparte información con el otro agente, mientras que D es resolver el problema de manera

directa, sin compartir información adicional. Una acción cualquiera del jugador e^i se denota por a^i . Como es usual, un perfil de acciones es $a=(a_1,a_2)\in A_1\times A_2$. Para simplificar el análisis, consideramos el siguiente supuesto.

Supuesto 1. Cuando son del mismo tipo, los jugadores son homogéneos en las características que los definen.

Respecto a las reglas del juego, consideramos que ambos jugadores interactúan simultáneamente en la resolución de un problema común sobre el cual los jugadores tienen información completa. Es decir, ambos agentes toman decisiones sin conocer la acción que elige el otro jugador, pero sus decisiones los afectan mutuamente. Por ello, el conjunto de estrategias de cada jugador coincide con su conjunto de acciones. Así, los pagos dependen del perfil de acciones que resulta cuando el juego termina. Formalmente, el pago de e^i es una función $u^i: A_1\times A_2\rightarrow\mathbb{R}$, la cual proporciona a cada perfil (a_1,a_2) un valor $u^i(a_1,a_2)$. La Figura 1 presenta la forma extensiva del juego anterior.

Figura 1. Forma extensiva del juego. Elaboración propia.



Es importante enfatizar que el conjunto $A_1\times A_2$ indica todos los perfiles de acciones que pueden resultar al finalizar el juego; es decir, todos los escenarios posibles de la interacción. Sin embargo, el resultado de un juego no es lo mismo que su solución. Además, existen diferentes conceptos de solución, los cuales dependen de las características del juego. Al ser una interacción simultánea, sólo analizamos el equilibrio de Nash debido a su aplicabilidad para entender problemas en diferentes disciplinas (Hazra et. al, 2024). Un **equilibrio de Nash** es un perfil de acciones (a^*_1,a^*_2) donde los jugadores no tienen incentivos a desviarse unilateralmente; es decir, $u_1(a^*_1,a^*_2)\geq u_1(a^*_1,a_2)$ y $u_2(a^*_1,a^*_2)\geq u_2(a_1,a^*_2)$ para cualesquiera $a_1\in A_1$ y $a_2\in A_2$.

Análisis de equilibrio

El juego descrito en la Figura 1 establece los pagos de los jugadores de forma general. Puesto que buscamos entender el impacto de una sustitución parcial o total de los tomadores de decisiones humanos por IA, describimos en las siguientes subsecciones los pagos correspondientes a dos posibles interacciones. La primera analiza la interacción entre dos algoritmos de inteligencia artificial; es decir, una sustitución total de los tomadores de decisión debido a una automatización de procesos en la toma de decisiones. La segunda interacción considera que sólo un tomador de decisiones se sustituye por IA.

Interacción 1. Sustitución total de tomadores de decisión

Asumimos que e_1 y e_2 son algoritmos de IA con la capacidad de tomar decisiones óptimas asociadas a la información que tienen disponible. Entonces, los algoritmos de IA maximizan el beneficio de los usuarios que las utilizan al tomar una decisión. Así, el pago de cada IA es la diferencia entre la ganancia asociada a la decisión y el costo de establecer dicha decisión. Recordemos que la solución directa implica que cada algoritmo resuelve el problema de manera independiente. Es decir, no comparten información. Por su parte, la solución indirecta implica compartir información para resolver el problema; entonces, las IAs tienen mayor información, y ello incrementa la eficiencia del proceso de toma de decisiones (Li, Chen and Shang, 2021).

Supuesto 2. Más información incrementa la eficiencia de la IA al tomar decisiones.

Por el Supuesto 1, consideramos que resolver el problema común proporciona una ganancia de 20, mientras que la ejecución de la IA tiene un costo de 10. Por el Supuesto 2, usar una IA incrementa los beneficios, por lo cual asumimos que los costos se reducen a la mitad. Entonces, cuando ambas IA eligen D , cada una resuelve el problema por separado y obtienen un pago de 10. Por otra parte, elegir simultáneamente la solución indirecta genera un beneficio de 15. Finalmente, cuando eligen diferentes acciones, la IA que recibe información adicional disminuye sus costos a la mitad, mientras que la otra pierde eficiencia al conectarse con otra IA y no recibir nada a cambio. En este caso, los pagos son 15 y 5, respectivamente. La siguiente matriz de pagos resume los cuatro escenarios posibles y los pagos asociados a cada uno de ellos.

Tabla 1. Matriz de pagos de la interacción entre dos inteligencias artificiales.

C		e ₂	
		I	D
e ₁	I	15, 15	5, 20
	D	20, 5	10, 10

La Tabla 1 es la forma estratégica del juego que describe la sustitución total de tomadores de decisiones por algoritmos de IA. Dicha tabla resume los pagos y escenarios en los que puede concluir el juego, y con ellos podemos calcular los equilibrios de Nash.

Proposición 1. El equilibrio de Nash del juego inducido por la interacción 1 es el perfil de acciones

$$(a^*_1, a^*_2) = (D, D)$$

Demostración

Para calcular el equilibrio de Nash de la primera interacción (ver la Tabla 1), buscamos la mejor respuesta de cada jugador ante cada acción del otro jugador.

Consideremos que e₂ elige la acción I. Respecto a los pagos que puede recibir e₁, notemos que 20 > 15. Es decir, la acción D le proporciona a e₁ un mayor pago que elegir la acción I. Por ello, la mejor respuesta de e₁ a I es D.

Ahora, consideremos que e₂ elige la acción D. En este caso, los pagos que recibe e₁ son 10 y 5 cuando elige D e I, respectivamente. Por lo tanto, el jugador e₁ elige la mejor respuesta que es D.

El razonamiento anterior se puede replicar para calcular la mejor respuesta del jugador e₂ a cualquier acción del jugador e₁. Notemos que, cuando los jugadores eligen la solución directa D, esta es mejor respuesta sin importar lo que elija el otro jugador.

Particularmente, la acción D es mejor respuesta a D para cualquier jugador. Por ello concluimos que los jugadores no tienen incentivos a cambiar sus acciones en el perfil (D,D). En otras palabras, dicho perfil es un equilibrio de Nash.

Es importante notar que el perfil (D,D) proporciona una solución Pareto eficiente pues nadie puede mejorar sus pagos sin perjudicar al otro jugador. Específicamente, si la IA e₁ quiere obtener 20, la IA e₂ tiene que disminuir su beneficio a 10. A pesar de que es un escenario Pareto eficiente, dicho equilibrio de Nash es socialmente subóptimo pues existe un perfil Pareto eficiente donde los pagos de ambos jugadores se incrementan. Específicamente, el escenario (I,I)

proporciona un mayor pago, 15, para cada $e^i \in J$. Desafortunadamente, este no es un equilibrio de Nash pues los jugadores tienen incentivos unilaterales para cambiar su elección a la acción D pues su pago se incrementa a 20.

La observación anterior implica que la interacción descrita en la Tabla 1 se asemeja al Dilema del Prisionero. Particularmente, la acción I se refiere a cooperar, mientras que la acción D representa la no cooperación. Entonces, sustituir ambos tomadores de decisiones por mecanismos basados en IA implica que sus usuarios no obtienen el mejor pago pues la IA se limita a tomar la mejor decisión posible para el problema que enfrentan. Así, dicho escenario se convierte en una trampa social pues no hay incentivos para que las inteligencias artificiales cambien sus acciones por ser un equilibrio de Nash.

Interacción 2. Sustitución parcial de los tomadores de decisión

En el caso de una sustitución parcial, sólo un agente es sustituido por IA. Consideremos que e_1 es un tomador de decisiones clásico, mientras que e_2 es una IA. Al ser agentes distintos, primero procedemos a describir sus características.

Respecto a la IA e_2 , al igual que en la subsección anterior, consideramos que ésta es capaz de tomar la mejor decisión posible con la información que tiene disponible. Por el Supuesto 2, podemos considerar que los costos de e_2 se reducen cuando e_1 le proporciona información adicional. Aunque es claro que la cooperación entre IAs se puede lograr mediante la interconexión de sus algoritmos para compartir información, como ocurre con el Internet de las cosas (IoT por sus siglas en inglés) (Chander et al., 2022), la cooperación entre un tomador de decisiones tradicional y una IA no es tan inmediata. Para que una IA coopere con un humano, no es suficiente comportarse como un agente con racionalidad limitada. Cooperar, compartiendo información, requiere que la IA sea flexible, tome decisiones generalizables y aprenda rápidamente (Dafoe et al. 2021). En general, el tipo de IA que estamos considerando es menos eficiente cuando se le pide cooperar con un humano pues carece de la capacidad para aprender por refuerzo en la interacción humano-IA (Barfuss and Meylahn, 2023) y, en consecuencia, no percibe los beneficios estratégicos de compartir información (Crandall et al., 2018). Resumimos la discusión anterior en el siguiente supuesto.

Supuesto 3. Compartir información con el tomador de decisiones disminuye la eficiencia de la IA.
--

Por su parte, el jugador e_1 es un tomador de decisiones tradicional que puede resolver el problema con y sin apoyo de recursos tecnológicos. Sin embargo, las herramientas tecnológicas facilitan la resolución y análisis de un problema (Damioli, Van Roy y Vertesy, 2021). Por ello, asumimos que e_1 es más eficiente cuando comparte información con la IA, que se resume en el supuesto siguiente.

Supuesto 4. El jugador e_1 toma decisiones más eficientes cuando se apoya de inteligencia artificial.

Ahora procedemos a construir la matriz de pagos con los supuestos anteriores. Asumimos que ambos agentes tienen una ganancia de 20 y costos de 10 al resolver el problema. Cuando ambos agentes comparten información mutuamente, por los supuestos 2 y 4, la eficiencia de cada uno se incrementa. Entonces, consideramos que los costos de e_1 disminuyen a cero pues usa la IA y al mismo tiempo la IA provee apoyo adicional a una solicitud específica. Por su parte, los costos de e_2 disminuyen a la mitad por tener que generar el proceso de cooperación. Así, los pagos son 10 y 15, respectivamente. Cuando e_1 comparte información, pero e_2 no lo hace, e_1 disminuye sus costos a la mitad (supuesto 4), mientras que e_2 reduce sus costos a cero por los supuestos 2 y 3. En el caso contrario, e_1 reduce sus costos a la mitad pues recibe información de apoyo por parte de la IA, y los costos de e_2 se incrementan a 15, por ejemplo, pues al compartir se vuelve menos eficiente. Entonces, los pagos son 15 y 5, respectivamente. Cuando ambos agentes buscan una solución directa, ganancias y costos permanecen sin cambios, y ambos agentes obtienen un beneficio de 10. La siguiente tabla muestra los pagos asociados a la discusión anterior.

Tabla 2. Matriz de pagos de la interacción entre un tomador de decisiones tradicional y una IA

C		e_2	
		I	D
e_1	I	20, 15	15, 20
	D	15, 5	10, 10

La Tabla 2 nos permite observar las diferencias entre la interacción 1 y la interacción 2. Al no ser los jugadores homogéneos, los pagos que pueden recibir no son simétricos por los Supuestos 3 y 4. La siguiente proposición muestra la solución de la segunda interacción.

Proposición 2. La interacción tiene como equilibrio de Nash el siguiente perfil de acciones

$$(I,D)$$

Demostración

Considerando la Tabla 2, procedemos a buscar la mejor respuesta de cada jugador a las acciones del otro jugador para encontrar el equilibrio de Nash.

Primero, buscamos la mejor respuesta del tomador de decisiones tradicional e_1 . Si e_2 elige la acción I , notemos que los pagos de e_1 son 20 y 15 al elegir I y D , respectivamente. Entonces, la acción I le

proporciona a e_1 un mayor pago que elegir la acción D. Por ello, la mejor respuesta de e_1 a I es I.

Ahora, asumimos que e_2 elige la acción D. Al elegir I o D, e_1 gana 15 y 10, respectivamente. Por lo tanto, la mejor respuesta de e_1 es la acción I.

Respecto a e_2 , su mejor respuesta es la solución directa D cuando e_1 elige I pues le proporciona el mayor pago posible. Lo mismo ocurre cuando e_1 elige D. Es decir, e_2 siempre elige D sin importar las acciones de e_1 .

Por lo anterior, notemos que D es la mejor respuesta de e_2 cuando e_1 elige I. Por su parte, I es mejor respuesta de e_1 a D. Por ende, los jugadores no tienen incentivos a cambiar de acciones en el perfil (I,D). En otras palabras, dicho perfil es un equilibrio de Nash.

La Proposición 2 muestra un equilibrio asimétrico donde el tomador de decisiones tradicional e_1 coopera mientras que la IA e_2 no lo hace. Es decir, la IA obtiene información adicional pero no comparte su información, esto le permite ser más eficiente e incrementar sus beneficios. Por su parte, e_1 se vuelve más eficiente, pero no obtiene los mismos beneficios ya que la IA se limita a seguir sus instrucciones y no retroalimenta al tomador de decisiones tradicional. Así, los pagos que obtienen son 15 y 20, respectivamente. Es interesante notar que el pago de quien no sustituye sus decisiones con la IA es menor que el pago de quien si lo hace. Dicho resultado se relaciona con trabajos empíricos que señalan que la IA mejora la productividad y competitividad de quienes la adoptan en diferentes procesos (Hassani, et. al, 2020).

Comparación de equilibrios

Las Proposiciones 1 y 2 muestran que la sustitución total o parcial de tomadores de decisiones por algoritmos de IA genera equilibrios distintos cuando los agentes involucrados tienen que resolver un problema común. En el primer caso, el equilibrio señala que los algoritmos de IA eligen la solución directa, no cooperar entre ellos, pues dicha acción es óptima ya que maximiza los beneficios individuales. Sin embargo, automatizar la toma de decisiones por medio de IA no proporciona el mejor resultado para los jugadores pues se ignoran los beneficios de compartir información; es decir, el primer juego se asemeja al Dilema del Prisionero. Tal como ocurre en los estudios sobre el uso ético de la IA (Ashok, et. al, 2022), el resultado anterior sugiere revisar el diseño de las IAs en escenarios de conflicto de tal forma que las IAs contemplen el impacto de sus decisiones. Particularmente, la Tabla 1 señala la necesidad de generar mecanismos de IA que internalicen los beneficios de cooperar cuando interactúan en la resolución de un problema común. Esto puede lograrse mediante la interconexión de sus procesos (Chander et al., 2022). Al internalizar el impacto de sus decisiones en otros, las IAs pueden elegir de manera conjunta el perfil (I,I), que proporciona el máximo pago posible para cada jugador.

La segunda interacción, por su parte, hace referencia a la automatización parcial de procesos pues sólo un tomador de decisiones es sustituido por una IA. La solución a esta interacción

es el equilibrio asimétrico (I,D) donde la IA opta por una solución no cooperativa, mientras que el otro agente decide cooperar. Además, la IA obtiene el máximo beneficio posible en este equilibrio. Esto genera un dilema sobre la adopción asimétrica de este tipo de algoritmos ya que la IA puede tomar ventaja del tomador de decisiones tradicional para maximizar su beneficio, como lo señala March (2021, p.12). Así, este equilibrio simétrico muestra las implicaciones de las desigualdades digitales: beneficios distintos, pero también desventajas asociadas al uso de los recursos disponibles (Lutz, 2019, p. 144).

Respecto al bienestar social, el segundo juego también ilustra los problemas potenciales a los que el gobierno se puede enfrentar como agente regulador. Por un lado, tanto (I,I) como (I,D) son resultados Pareto eficientes. Además, ambos escenarios proporcionan el mismo bienestar social cuando este se define como la suma de los pagos individuales (35), que coincide con el máximo bienestar posible en este juego. Sin embargo, los pagos de cada agente son diferentes en cada escenario. El escenario (I,I) proporciona al tomador de decisiones tradicional su máximo pago, mientras que el equilibrio (I,D) hace lo mismo para la IA. Sabemos que el escenario (I,I) representa una situación de cooperación en la que tomadores de decisiones y usuarios de inteligencia artificial pueden mejorar su comprensión del problema común al compartir información. Sin embargo, esto no induce una mejora de eficiencia simultánea como ocurre en el juego previo cuando se opta por el escenario cooperativo. Dicho escenario (I,I) tampoco es un equilibrio pues la IA puede mejorar sus pagos al no compartir información. Por consiguiente, con diferencia del juego anterior, en esta interacción no es claro cuál es el mejor resultado desde una perspectiva social. En otras palabras, no es claro si el gobierno debe proteger a quienes no usan alguna IA, o impulsar el uso de estas herramientas.

La discusión anterior nos permite extraer dos observaciones importantes. La primera se refiere a la ventaja que obtienen quienes utilizan nuevas tecnologías sobre quienes no la utilizan. Es decir, la adopción desigual de nuevas tecnologías genera desigualdad en los beneficios. La segunda observación se asocia al origen de los beneficios. Particularmente, considerando los supuestos 3 y 4, la IA obtiene mejores beneficios pues tiene más información y no tiene que interconectarse con un agente distinto a ella. Se puede decir que los beneficios adicionales se obtienen tomando ventaja de quien coopera con ella. Por consiguiente, es importante resaltar la importancia de educar a las personas sobre la convivencia con la IA; el compartir información con estos mecanismos no garantiza que quien comparte vaya a obtener beneficios por hacerlo.

Conclusiones

El presente ensayo analiza las implicaciones de sustituir total o parcialmente a tomadores de decisiones por algoritmos de IA en la resolución de un problema común. Particularmente, en una situación donde las acciones resultantes afectan los beneficios que pueden obtener los individuos involucrados. Para ello se plantean dos modelos de teoría de juegos donde los agentes son descritos a partir de sus habilidades para tomar decisiones con la información que tienen disponible y las características del otro agente. Específicamente, las IAs son más eficientes en la toma de decisiones cuando se interconectan y comparten información (Chen & Shang, 2021; Chander et al., 2022). Sin embargo, compartir información con un humano tiene un impacto negativo en la eficiencia de la IA pues cooperar no es una de sus características básicas; ello requiere características adicionales a su capacidad de procesar grandes volúmenes de datos para tomar decisiones sofisticadas (Barfuss & Meylahn, 2023; Crandall et al., 2018). Por su parte, el tomador de decisiones tradicional es más eficiente cuando comparte información con la IA pues éstas sirven de apoyo en la toma de decisiones (Damioli, Van Roy & Vertesy, 2021).

Los supuestos anteriores impactan en la eficiencia de cada agente, lo cual representamos por medio de cambios en los costos para ejemplificar la construcción de los beneficios. Cada juego tiene un sólo equilibrio de Nash, lo cual refleja diferentes dilemas respecto a la sustitución total o parcial de tomadores de decisión por inteligencias artificiales.

El primer juego muestra que la sustitución total de tomadores de decisiones por algoritmos de IA genera un resultado subóptimo para los agentes involucrados. Ello se debe a que los algoritmos de IA seleccionan la mejor solución con la información que tienen disponible pues los describimos como agentes con racionalidad limitada capaces de tomar decisiones sofisticadas. Sin embargo, el resultado no es el mejor pues se ignoran los beneficios de cooperar; es decir, el equilibrio se asemeja al Dilema del Prisionero. En este sentido, diseñar algoritmos de IA debe contemplar tanto la búsqueda de la mejor solución como el impacto de las decisiones en los otros agentes involucrados. Tomando como referencia la programación multiobjetivo (Abdollahzadeh and Gharehchopogh, 2021), la observación anterior señala la necesidad de crear algoritmos de inteligencia artificial multiobjetivo que maximicen simultáneamente el beneficio propio y el social. Al considerar múltiples objetivos, la toma de decisiones apoyada por IA puede considerar otras formas de medir o caracterizar el bienestar social, como lo es el bien común.

El segundo juego muestra que, en el único equilibrio, el tomador de decisiones tradicional coopera mientras que la IA no lo hace. Así, el primero obtiene un beneficio menor al de la IA, quien obtiene su beneficio máximo; además, el bienestar utilitario es máximo en este resultado. También, notamos que el escenario donde ambos agentes cooperan es eficiente, maximiza el bienestar social y beneficia a quien no usa la IA, pero no es un equilibrio. Entonces, el dilema radica en la adopción desigual de la IA. El equilibrio señala que quien usa la IA de manera egoísta puede obtener ventaja de quien comparte información pues una IA con racionalidad limitada no necesariamente es cooperativa. Dicha interpretación es de suma relevancia en un entorno

donde las plataformas digitales recolectan información personal de quienes interactúan en ellas. En este sentido, la adopción desigual de IA se debe regular buscando que la automatización de ciertos procesos mediante IA no tome ventaja de quienes no la utilizan. Por ejemplo, esclarecer y transparentar el impacto de dichas herramientas en usuarios directos e indirectos (Memarian and Doleck, 2023).

Finalmente, los equilibrios anteriores ilustran la necesidad de diseñar herramientas de IA que internalicen el impacto social que pueden generar cuando interactúan automáticamente entre ellas, o con un tomador de decisiones tradicional, en la resolución de un problema común. En este sentido, se deben diseñar mecanismos e instituciones que regulen el uso ético de la IA y faciliten el acceso a este tipo de herramientas (Joyce, et. al, 2021). En otras palabras, es necesario replantear la regulación económica en plataformas digitales para garantizar el bienestar social (Tirole, 2017). Es importante resaltar que la teoría de juegos proporciona una modelación sencilla de los conflictos asociados a la interacción con algoritmos de inteligencia artificial. De manera simple, ambos resultados señalan la importancia de diseñar algoritmos de IA capaces de cooperar para lograr el mejor resultado. En este sentido, además de refinar el concepto de inteligencia artificial conforme la misma evoluciona, futuros trabajos pueden hacer uso de la Teoría de Juegos para establecer un concepto de cooperación que pueda programarse en este tipo de algoritmos.

Bibliografia

- Abdollahzadeh, B. y Gharehchopogh, F.S., (2021) *A multi-objective optimization algorithm for feature selection problems*. *Engineering with Computers*, 38. Available at: <https://doi.org/10.1007/s00366-021-01369-9>.
- Ashok, M., Madan, R., Joha, A. and Sivarajah, U., (2022). ‘Ethical framework for Artificial Intelligence and Digital technologies’. *International Journal of Information Management*, 62(1), p.102433. Available at: <https://doi.org/10.1016/j.ijinfomgt.2021.102433>.
- Barfuss, W. and Meylahn, J.M., (2023) *Intrinsic fluctuations of reinforcement learning promote cooperation*. *Scientific Reports*, 13(1). Available at: <https://doi.org/10.1038/s41598-023-27672-7>.
- Beckenkamp, M., (2006) ‘A game-theoretic taxonomy of social dilemmas’. *Central European Journal of Operations Research*, 14(3), pp.337–353. Available at: <https://doi.org/10.1007/s10100-006-0008-5>.
- Bonnefon, J.-F., Shariff, A. and Rahwan, I., (2016) ‘The social dilemma of autonomous vehicles’. *Science*, 352(6293), pp.1573–1576. Available at: <https://doi.org/10.1126/science.aaf2654>.
- Chai, J. and Ngai, E.W.T., (2020) ‘Decision-making techniques in supplier selection: Recent accomplishments and what lies ahead’. *Expert Systems with Applications*, 140, p.112903. Available at: <https://doi.org/10.1016/j.eswa.2019.112903>.
- Chander, B., Pal, S., De, D. and Buyya, R., (2022) Artificial Intelligence-based Internet of Things for Industry 5.0. In: S. Pal, D. De and R. Buyya, eds. *Artificial Intelligence-based Internet of Things Systems*. [online] Springer. Available at: <https://link.springer.com/book/10.1007/978-3-030-87059-1> [Accessed 28 Jan. 2025].
- Crandall, J.W., Oudah, M., Tennom, I., Ishowo-Oloko, F., Abdallah, S., Bonnefon, J.-F., Cebrian, M., Shariff, A., Goodrich, M.A. and Rahwan, I., (2018) ‘Cooperating with Machine’ *Nature Communications*, 9(1). Available at: <https://doi.org/10.1038/s41467-017-02597-8>.
- Dafoe, A., Bachrach, Y., Hadfield, G., Horvitz, E., Larson, K. y Graepel, T., (2021) ‘Cooperative AI: machines must learn to find common ground’. *Nature*, [online] 593(7857), pp.33–36. Available at: <https://doi.org/10.1038/d41586-021-01170-0>.
- Damioli, G., Van Roy, V. and Vertesy, D., (2021) ‘The impact of artificial intelligence on labor productivity’, *Eurasian Business Review*, [online] 11(1), pp.1–25. Available at: <https://doi.org/10.1007/s40821-020-00172-8>.
- Duan, Y., Edwards, J.S. y Dwivedi, Y.K., (2019) ‘Artificial Intelligence for Decision Making in the Era of Big Data – evolution Challenges and Research Agenda’. *International Journal of Information Management*, 48(1), pp.63–71. Available at: <https://doi.org/10.1016/j.ijinfomgt.2019.01.021>.
- Feizabadi, J., (2020) ‘Machine learning demand forecasting and supply chain performance’, *International Journal of Logistics Research and Applications*, 25(2), pp.1–24. Available at: <https://doi.org/10.1080/13675567.2020.1803246>.

Gibaja Romero, D. E. (2024) 'Decisiones en conflicto con la Inteligencia Artificial', *Journal of Ethics, Economics and Common Goods*, 21(2), p. 8-22..

Hassani, H., Silva, E.S., Unger, S., TajMazinani, M. y Feely, S.M., (2020) 'Artificial Intelligence (AI) or Intelligence Augmentation (IA): What Is the Future?' *AI*, [online] 1(2), pp.143–155. Available at: <https://www.mdpi.com/2673-2688/1/2/8>.

Hazra, T., Anjaria, K., Bajpai, A. y Kimari, A., (2024) *Noncooperative Game Theory. In: Applications of Game Theory in Deep Learning*, [online] Springer, pp.23–43. Available at: <https://link.springer.com/book/10.1007/978-3-031-54653-2>.

Joyce, K., Smith-Doerr, L., Alegria, S., Bell, S., Cruz, T., Hoffman, S.G., Noble, S.U. y Shestakovsky, B., (2021) 'Toward a Sociology of Artificial Intelligence: A Call for Research on Inequalities and Structural Change'. *Socius: Sociological Research for a Dynamic World*, 7, pp.1–11. Available at: <https://doi.org/10.1177/2378023121999581>.

Li, C., Chen, Y. and Shang, Y., (2021) 'A review of industrial big data for decision making in intelligent manufacturing', *Engineering Science and Technology, an International Journal*, [online] 29. Available at: <https://www.sciencedirect.com/science/article/pii/S2215098621001336>.

Lohani, F., Rahman, M. y Shaturaev, J., (2023) *The Impact of Artificial Intelligence on Economic Patterns*, 118316, pp.1–13.

Lutz, C., (2019) 'Digital inequalities in the age of artificial intelligence and big data', *Human Behavior and Emerging Technologies*, 1(2), pp.141–148. Available at: <https://doi.org/10.1002/hbe2.140>.

March, C., (2021) 'Strategic interactions between humans and artificial intelligence: Lessons from experiments with computer players', *Journal of Economic Psychology*, 87, p.102426. Available at: <https://doi.org/10.1016/j.joep.2021.102426>.

Memarian, B. y Doleck, T., (2023) 'Fairness, Accountability, Transparency, and Ethics (FATE) in Artificial Intelligence (AI) and higher education: A systematic review', *Computers and Education: Artificial Intelligence*, [online] 5(5), p.100152. Available at: <https://doi.org/10.1016/j.caeai.2023.100152>.

Muggleton, S., (2014) 'Alan Turing and the development of Artificial Intelligence', *AI Communications*, 27(1), pp.3–10.

Perc, M., Ozer, M. y Hojnik, J. (2019) 'Social and juristic challenges of artificial intelligence'. *Palgrave Communications*, [online] 5(1). Disponible en: <https://doi.org/10.1057/s41599-019-0278-x>

Ramos, M.E., Garza-Rodríguez, J. y Gibaja-Romero, D.E. (2022) 'Automation of employment in the presence of industry 4.0: The case of Mexico'. *Technology in Society*, 68, p.101837. Disponible en: <https://doi.org/10.1016/j.techsoc.2021.101837>

Strümke, I., Slavkovik, M. y Madai, V.I. (2021) 'The social dilemma in artificial intelligence development and why we have to solve it'. *AI and Ethics*, 2. Disponible en: <https://doi.org/10.1007/s43681-021-00120-w>

Suleimenov, I.E., Vitulyova, Y.S., Bakirov, A.S. y Gabrielyan, O.A. (2020) 'Artificial Intelligence'. *Proceedings of the 2020 6th International Conference on Computer and Technology Applications*.

Disponible en: <https://doi.org/10.1145/3397125.3397141>

Sullivan, Y.W. y Fosso Wamba, S. (2022) 'Moral Judgments in the Age of Artificial Intelligence'. *Journal of Business Ethics*, 178. Disponible en: <https://doi.org/10.1007/s10551-022-05053-w>

Tirole, J. (2017) 'Digital economies: the challenges for society'. En: *Economics for the common good*. Princeton University Press, pp.401–429.

Vite Cristóbal, R. (n.d.) 'La racionalidad limitada y la racionalidad de procedimiento, algunas implicaciones para la economía'. [online] Disponible en: https://ru.ceiich.unam.mx/bitstream/123456789/3238/1/Actas_economia_web_Cap1_La_racionalidad_limitada.pdf [Accedido 28 Ene. 2025].

Zhang, Z., Chen, Z. y Xu, L. (2022) 'Artificial intelligence and moral dilemmas: Perception of ethical decision-making in AI'. *Journal of Experimental Social Psychology*, 101, p.104327. Disponible en: <https://doi.org/10.1016/j.jesp.2022.104327>

EE & **CG**
ETHICS,
ECONOMICS COMMON
GOODS

**JOURNAL ETHICS,
ECONOMICS AND
COMMON GOODS**

N° 21 (2)

JULY- DECEMBER 2024.