

Cuando la IA generativa incrementa la desigualdad: evidencia experimental de una competición de debates universitarios

Research Brief Septiembre 2024

IDEAS CLAVE

- Este *research brief* resume los resultados de una investigación que evalúa el impacto de la inteligencia artificial (IA) generativa en la productividad y las desigualdades a través de un experimento (RCT) en un torneo de debate, en el cual asigno aleatoriamente apoyo de ChatGPT a estudiantes para preparar una serie de debates uno contra uno.
- En contra de los resultados hasta ahora en la literatura que apuntan a una reducción de las desigualdades en productividad por la mejora de los trabajadores de habilidades más bajas en una variedad de tareas escritas como consecuencia del uso de IA generativa, los resultados de este experimento muestran que los individuos con altas habilidades se benefician significativamente más de ella que sus compañeros de menor habilidad.
- Propongo una posible explicación para reconciliar estos resultados con los hallazgos anteriores: en interacciones verbales impredecibles que requieren pensamiento crítico y habilidades sociales, cuando las respuestas no pueden ser copiadas y pegadas, los trabajadores con alta capacidad se benefician más de estas herramientas.

AUTOR

Toni Roldán Monés

Director, EsadeEcPol

Research brief a partir del
Working Paper "[When GenAI
increases inequality: evidence
from a university debating
competition](#)" Toni Roldán
Monés

Resumen de la investigación

En el último año y medio las plataformas de inteligencia artificial (IA) generativa se han expandido por todo el mundo, convirtiéndose en herramientas utilizadas por millones de usuarios a diario. Tan solo dos meses después de su lanzamiento, ChatGPT alcanzó los 100 millones de usuarios activos mensuales, experimentando un crecimiento más rápido que cualquier app, servicio de internet o red social.

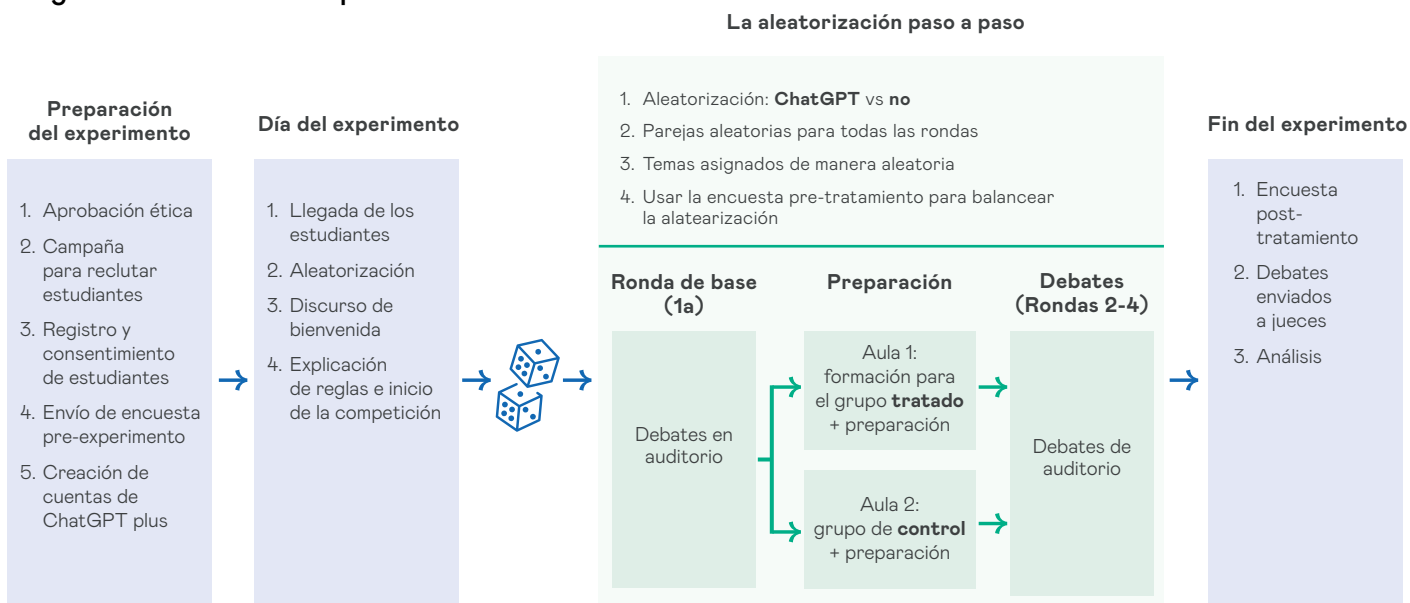
En este contexto, una de las preguntas más importantes que están abordando los investigadores es el **impacto de estas tecnologías basadas en modelos de lenguaje extenso** (LLM, por sus siglas en inglés) **en el mercado de trabajo**. Anteriores olas de automatización tuvieron un efecto negativo en los trabajos "rutinarios" y de baja cualificación, al tiempo que aumentaron la productividad de trabajadores altamente calificados, lo que condujo al incremento de las desigualdades (Autor et al, 2003; Acemoglu y Autor, 2011; Acemoglu y Restrepo, 2017; Felten et al, 2019). Sin embargo, sigue siendo una incógnita quién se beneficiará de la IA generativa y en qué tareas específicas (Wilmers, 2024). En algunos casos, podría reducir las diferencias de productividad entre trabajadores, y en otros, podría aumentar las desigualdades (Autor, 2024).

Hasta el momento la investigación especializada ha producido dos aprendizajes provisionales. La primera es que debido a las mayores capacidades técnicas de estos sistemas se espera que tengan un **fuerte impacto en algunas profesiones de alta cualificación** como abogados, programadores o publicistas, que habían estado protegidas de la automatización en el pasado (Eloundou et al, 2023; Felten et al, 2023). La segunda conclusión es que, contrariamente a anteriores revoluciones tecnológicas, **estas tecnologías podrían ayudar a mejorar más la productividad de trabajadores de bajo rendimiento**, contribuyendo a reducir desigualdades de productividad.

Existen resultados claros de este efecto de compresión de la distribución de productividad en una gran variedad de tareas, incluyendo programación (Peng et al, 2023), escritura profesional, como la redacción de correos electrónicos o comunicados de prensa (Noy y Zhang, 2023), exámenes de derecho (Choi y Schwart, 2023), o escritura creativa (Doshi et al. 2023). Dell'Acqua y coautores (Dell'Acqua et al, 2023), por ejemplo, utilizan una muestra amplia de consultores en estrategia para estudiar el efecto de ChatGPT en 18 tareas escritas, como redactar un memorandum de 500 palabras para el CEO o proponer ideas para buenos eslóganes de marketing. Por su parte, (Brynjolfsson et al, 2023) analiza el efecto de un modelo de aprendizaje automático de atención al cliente. Todos estos estudios encuentran que **la IA generativa aumenta la productividad de todos los trabajadores**, al mismo tiempo que **reduce la desigualdad inicial en el rendimiento de las tareas debido a mayores mejoras en los trabajadores con peor desempeño**.

En este *research brief* difundimos los resultados de un **experimento aleatorizado controlado** (RCT, por sus siglas en inglés) que **estudia el impacto de ChatGPT** (la herramienta de IA generativa más usada en el mundo) en **un concurso de debates universitarios**¹. El entorno único de los debates permite analizar una serie de habilidades esenciales de los trabajadores que son habitualmente difíciles de testar. Los debates exigen un desarrollo sofisticado de la argumentación, tener pensamiento crítico, capacidad de persuasión y habilidades sociales y cognitivas muy exigentes para todo ello². En contraste con los resultados que muestra hasta ahora la literatura experimental de IA generativa - donde el análisis se ha centrado en su impacto sobre pruebas o habilidades escritas y acotadas - **este experimento muestra que ChatGPT favorece significativamente más a los individuos de mayores habilidades**, aumentando las desigualdades de productividad.

Figura 1. Diseño del experimento



El experimento tuvo lugar en tres sesiones distintas con un total de 142 estudiantes de Esade y CUNEF Universidad en Madrid y Barcelona. Los estudiantes se enfrentaron en tres o cuatro rondas de debates cortos, uno contra uno, siguiendo una versión simplificada de las competiciones tipo “British Parliament”. Después de la primera ronda de debates (la línea base), a la mitad de los estudiantes (grupo tratado) se les asignó aleatoriamente un entrenamiento intensivo de 20 minutos con ChatGPT y se les permitió usarlo como apoyo durante todo el concurso. El grupo de control solo pudo utilizar recursos convencionales en internet.

1 Este estudio fue pre-registrado en el registro de ensayos controlados aleatorizados de la American Economic Association, con el número de registro AEARCTR-0011113. Ver <https://www.socialscienceregistry.org/trials/11113>

2 Los resultados completos de este cuyos resultados están publicados en [Roldán-Monés \(2024\)](#).

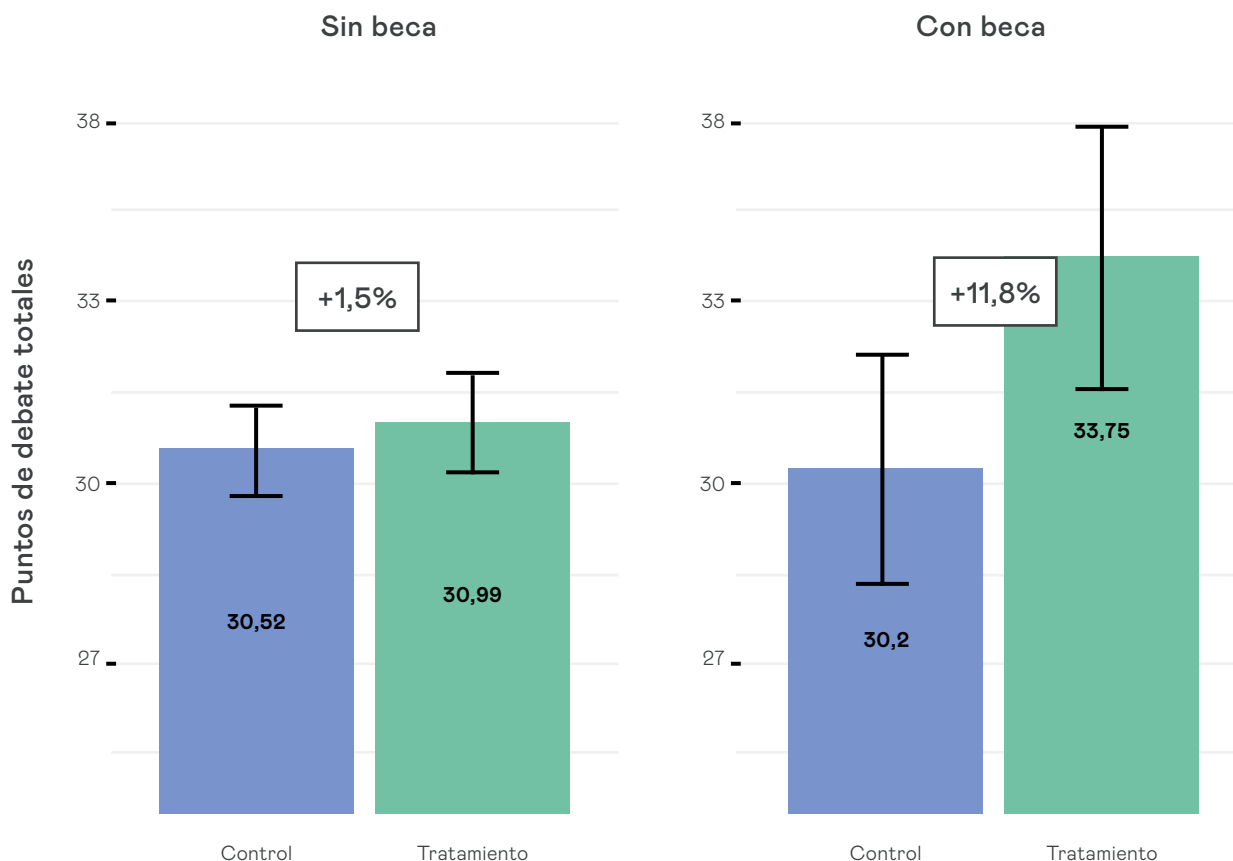
Los estudiantes también fueron asignados aleatoriamente a posiciones (a favor/en contra) y parejas de contrincantes. Cada debate fue grabado en audio y enviado a tres jueces expertos independientes que no tenían ninguna información sobre el experimento. La rúbrica para la evaluación siguió las métricas habituales en competiciones internacionales de debate. También se ofrecieron diez premios de 100€ en vales de Amazon a los ganadores para incentivar el máximo esfuerzo entre los estudiantes.

Una de las aplicaciones claramente relevantes de ChatGPT es que proporciona resúmenes rápidos de ideas y conceptos complejos, argumentos a favor o en contra de cualquier tema y un sinfín de ejemplos y metáforas. Por estas razones, ChatGPT puede ser un aliado poderoso para mejorar las habilidades de debate. El experimento testó **dos hipótesis** principales: (1) **si ChatGPT mejora el rendimiento general en los debates** y (2) **si contribuye a reducir la desigualdad** entre los estudiantes con mayores y menores habilidades iniciales. Los principales resultados que se miden son (a) puntos de debate (por estudiante) y (b) probabilidad de ganar el debate.

Los resultados muestran que ChatGPT tiene un efecto positivo, pero no significativo en el rendimiento general en los debates. **Los debatientes del grupo de tratamiento tienen un 9,2% más de probabilidades de ganar que los del grupo de control y obtienen, en promedio, un 2,2% más de puntos** (equivalente a un aumento de 0,15 desviaciones estándar) que los individuos del grupo de control, pero las diferencias no son estadísticamente significativas.

Sin embargo, este resultado oculta una heterogeneidad interesante. ChatGPT ayuda significativamente más a aquellos estudiantes que están en la parte superior de la distribución de habilidades. Utilizando una medida de la capacidad innata de los estudiantes, si tienen una beca de excelencia o no, muestro que los estudiantes con alta capacidad (es decir, con beca) experimentan mejoras significativamente mayores al usar ChatGPT que el resto de los estudiantes. La estimación del coeficiente muestra una **mejora del 12% entre los estudiantes del grupo de tratamiento que tienen una beca**, y un efecto del 1,7% (no significativo) de ChatGPT entre aquellos sin beca por mérito.

Figura 2. Heterogeneidad en los efectos de tratamiento de ChatGPT por habilidad de estudiantes



Nota: esta figura muestra la puntuación media de individuos en grupos de control o tratamiento para aquellos con o sin beca de excelencia, condicional en variables de control. Las líneas representan intervalos de confianza del 90% (márgenes predictivos del indicador de tratamiento). Los porcentajes indican el cambio en resultado entre grupos de tratamiento y de control.

Usando una medida adicional de capacidad, encuentro que entre aquellos que se encuentran en el 50% superior en puntos de debate en la ronda inicial (antes de que se implementara el tratamiento), los estudiantes del grupo de tratamiento tienen en promedio un 5,2% más de puntos que los estudiantes del grupo de control. Entre aquellos con puntos de debate más bajos en la ronda inicial (50% inferior), los individuos del grupo de tratamiento no se benefician en absoluto de ChatGPT. Estos hallazgos sugieren que, para tareas como el debate que requieren habilidades de "orden superior", la IA generativa es complementaria a la habilidad y puede aumentar las desigualdades en productividad.

También estudio una serie de mecanismos que pueden estar explicando los resultados. El análisis de la rúbrica de evaluación del debate por parte de los jueces muestra que ChatGPT **mejora significativamente las puntuaciones** en cuatro de los cinco indicadores de debate, como la **credibilidad** y **superioridad** de sus argumentos, como su capacidad de **refutación** y **retórica**. Mientras que *para los estudiantes de bajas habilidades la IA generativa solo tiene efectos en un indicador, asociado con la claridad*: una medida asociada con la estructura de la información presentada, más que con la calidad. En el resto de tareas asociadas con la capacidad de discriminar entre contenido de buena y mala calidad estos estudiantes no son capaces de beneficiarse del acceso a ChatGPT. Las respuestas de la encuesta final indican que los individuos con alta capacidad experimentan efectos positivos, significativos y grandes de ChatGPT en las percepciones auto-reportadas sobre “haber tenido suficiente tiempo para preparar los debates”.

Tabla 1. Efectos de ChatGPT en puntos totales de debate por subcategoría

	Claridad		Credibilidad		Retórica		Refutación		Superioridad de los argumentos	
	(1) Bajo	(2) Alto	(3) Bajo	(4) Alto	(5) Bajo	(6) Alto	(7) Bajo	(8) Alto	(9) Bajo	(10) Alto
Tratamiento	0,238*** (0,083)	-0,197 (0,306)	-0,134 (0,204)	1,756*** (0,491)	0,081 (0,093)	1,736 (0,259)	0,126 (0,106)	1,858*** (0,199)	0,090 (0,094)	1,430*** (0,230)
Constante	4,781*** (1,019)	8,684*** (0,379)	0,932 (2,486)	4,333*** (0,765)	3,777*** (1,107)	9,521*** (0,444)	1,225 (1,370)	2,881*** (0,354)	3,803*** (1,054)	6,657*** (0,320)
Media var. dep.	6,12	6,02	5,12	5,28	6,41	6,22	5,82	5,50	5,94	5,67
SD var. dep	0,73	1,07	1,73	1,62	0,81	0,84	0,93	1,14	0,81	1,04
R ²	0,30	0,61	0,20	0,62	0,31	0,60	0,23	0,64	0,22	0,61
Obs.	304	60	304	60	304	60	304	60	304	60

Los niveles de significancia están indicados por * < .1, ** < .05, *** < .01. La tabla muestra los coeficientes de regresiones de la ecuación 1, donde la variable dependiente es el total de puntos de debate del individuo i en la ronda de debate r para la subcategoría de la rúbrica (calculada como el promedio entre diferentes jueces que evalúan el mismo debate del individuo), por separado para estudiantes de baja habilidad (no beneficiarios de becas por mérito) y alta habilidad (beneficiarios de becas por mérito). Todas las especificaciones incluyen los mismos controles que los reportados en las notas de la Tabla 2. Los errores estándar están agrupados a nivel individual porque cada individuo es observado entre 2 y 3 veces, dependiendo del número de debates que completaron. El número total de individuos incluidos en cada regresión es 141 de los 142 individuos aleatorizados, ya que un individuo no completó la encuesta final.

Propongo una posible explicación teórica para reconciliar estos resultados con los hallazgos anteriores que apoyan la hipótesis de la “compresión de la productividad” de la IA generativa. En tareas escritas predecibles, los modelos de IA y de aprendizaje automático (ML) aprenden patrones de comportamiento de los mejores y peores desempeños (Brynjolfsson et al, 2023). Esto permite que los sistemas de IA reproduzcan el contenido de los mejores, ayudando a los de bajo desempeño a mejorar mediante interacciones simples, poca reflexión y copia y pega de respuestas. En ese tipo de entornos, se puede esperar que los individuos sean al menos tan competentes como el sistema de IA al que pueden acceder. Sin embargo, **en entornos sociales** con interacciones repetidas, **intensivas en conocimiento**, más cercanas a las que se encuentran en los trabajos de gestión de todo tipo, **los individuos de altas habilidades estarán mejor capacitados para extraer y utilizar la información proporcionada por la IA generativa** a su favor que los individuos de bajas habilidades.

Conclusiones e implicaciones de políticas

Una novedad de los sistemas de IA generativa es que adquieren conocimiento a través de la observación en lugar de reglas, lo que les permite realizar tareas sofisticadas, tradicionalmente reservadas para profesionales altamente capacitados (Autor, 2024). Los primeros experimentos que prueban el impacto de ChatGPT en la productividad laboral en una variedad de tareas escritas muestran un patrón común: los sistemas de GenAI ayudan más a los trabajadores de bajo rendimiento que a los de alto rendimiento, comprimiendo así la distribución de la productividad. Este estudio complementa los hallazgos anteriores al explorar los efectos de la IA en una tarea novedosa, un concurso de debates, que ofrece un entorno ideal para probar si la GenAI puede aumentar la productividad cuando diferentes habilidades sociales e intensivas en conocimiento son esenciales para la productividad.

Contrario a los resultados iniciales en la literatura sobre GenAI, encuentro que los individuos con altas habilidades se benefician más de la interacción con ChatGPT que los individuos con bajas habilidades. También propongo una posible explicación: cuando se requieren habilidades de orden superior en contextos sociales realistas e impredecibles, es probable que los trabajadores con altas habilidades disfruten de una complementariedad más fuerte con la IA. Si estos hallazgos se replican en otros contextos que involucren habilidades de orden superior, como la negociación o la venta, por ejemplo, tendrían implicaciones relevantes para comprender el impacto de la GenAI en las desigualdades laborales.

En todos los casos dos conclusiones son claras: (1) es **todavía pronto para saber los efectos en desigualdad** que tendrán estas tecnologías; (2) habrá que **analizar en cada caso específico**, en qué tareas concretas y qué perfiles de trabajadores, para **entender quien ganará o perderá**, y en qué medida estas tecnologías serán complementarias o sustitutivas al trabajo humano.

REFERENCIAS

- Acemoglu, Daron and David Autor, "Chapter 12 - Skills, Tasks and Technologies: Implications for Employment and Earnings," in David Card and Orley Ashenfelter, eds., *Handbook of Labor Economics*, Vol. 4, Elsevier, 2011, pp. 1043–1171.
- and Pascual Restrepo, "Robots and Jobs: Evidence from US Labor Markets," Working Paper 23285, National Bureau of Economic Research March 2017.
- Aristotle, *The Rhetoric of Aristotle*, New York: Appleton-Century-Crofts, Inc, 1960.
- Autor, David, "Applying AI to Rebuild Middle Class Jobs," Working Paper 32140, National Bureau of Economic Research February 2024.
- Autor, David H., Frank Levy, and Richard J. Murnane, "The Skill Content of Recent Technological Change: An Empirical Exploration," *The Quarterly Journal of Economics*, 11 2003, 118 (4), 1279–1333.
- Brynjolfsson, Erik, Danielle Li, and Lindsey R Raymond, "Generative AI at Work," Working Paper 31161, National Bureau of Economic Research April 2023.
- Choi, Jonathan H. and Daniel Schwarcz, "AI Assistance in Legal Analysis: An Empirical Study," *Minnesota Legal Studies Research Paper No. 23-22*, 2023.
- Dell'Acqua, Fabrizio, Edward McFowland, Ethan R. Mollick, Hila Lifshitz-Assaf, Katherine Kellogg, Saran Rajendran, Lisa Krayer, Francois Candelon, and Karim R. Lakhani, "Navigating the Jagged Technological Frontier: Field Experimental Evidence of the Effects of AI on Knowledge Worker Productivity and Quality," *Harvard Business School Technology & Operations Mgt*, 2023, Unit Working Paper No. 24-013.
- Deming, David J., "The Growing Importance of Social Skills in the Labor Market," *The Quarterly Journal of Economics*, 06 2017, 132 (4), 1593–1640. "Four Facts about Human Capital," *Journal of Economic Perspectives*, August 2022, 36 (3), 75–102.
- Doshi, Anil Rajnikant and Oliver Hauser, "Generative Artificial Intelligence Enhances Creativity but Reduces the Diversity of Novel Content," SSRN, 2023.
- Eloundou, Tyna, Sam Manning, Pamela Mishkin, and Daniel Rock, "GPTs are GPTs: An Early Look at the Labor Market Impact Potential of Large Language Models," March 2023, (2303.10130).
- Felten, Edward, Manav Raj, and Robert Channing Seamans, "The Effect of Artificial Intelligence on Human Labor: An Ability-Based Approach," *Academy of Management Proceedings*, 2019, 2019 (1), 15784.
- Felten, Edward W, Manav Raj, and Robert Seamans, "Occupational heterogeneity in exposure to generative ai," *Available at SSRN 4414065*, 2023.
- Noy, Shakked and Whitney Zhang, "Experimental Evidence on the Productivity Effects of Generative Artificial Intelligence," 2023.
- Peng, Sida, Eirini Kalliamvakou, Peter Cihon, and Mert Demirer, "The Impact of AI on Developer Productivity: Evidence from GitHub Copilot," 2023.
- Roldan-Monés, "When GenAI increases inequality: evidence from a University Debating Competition", POID Working Paper, London School of Economics and Political Science, 2024.
- Wilmers, Nathan, "Generative AI and the Future of Inequality," *An MIT Exploration of Generative AI*, mar 27 2024. <https://mit-genai.pubpub.org/pub/24gsgdix>.