



GOBIERNO DE ESPAÑA

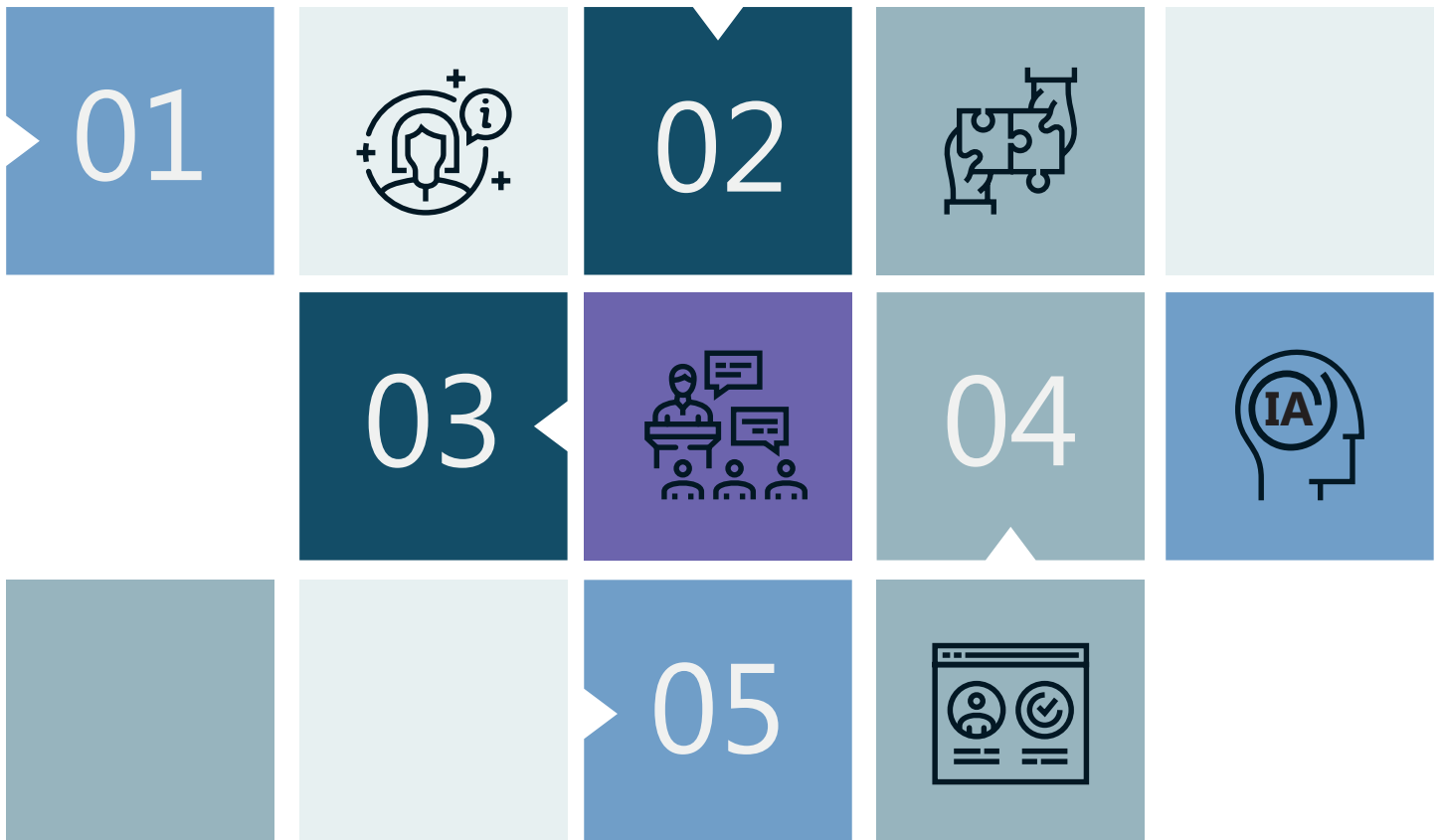
MINISTERIO DE CIENCIA, INNOVACIÓN Y UNIVERSIDADES



Aportación de la IA en la evaluación ex-ante del impacto social de proyectos de emprendimiento tecnológico

Una aproximación experimental

SHADOW EXPERIMENT NEOTEC



Créditos

Basado en el "Informe de resultados preliminares – Shadow Experiment Neotec", elaborado por Teo Firpo y Hugo Cuello, con la colaboración de Ascensión Barajas. Febrero de 2026.

Los resultados presentados en este informe proceden del proyecto de investigación en curso "*Human Experts, Artificial Intelligence, and the Assessment of the Social Impact of R&D Grants.*", de Teo Firpo.

Publicación incluida en el Programa Editorial 2026 del Ministerio de Ciencia, Innovación y Universidades.

Catálogo de publicaciones oficiales: <https://cpage.mpr.gob.es>

e-NIPO: 154260109

Mayo, 2026

Edita:

Centro para el Desarrollo Tecnológico y la Innovación (CDTI E.P.E.)

Departamento de Estrategia y Evaluación

C/Cid, 4, 28001, Madrid.

Licencia Creative Commons: Atribución/Reconocimiento-NoComercial-SinDerivados 4.0 Internacional
Usted es libre de:

Compartir — copiar y redistribuir el material en cualquier medio o formato. La licenciante no puede revocar estas libertades en tanto usted siga los términos de la licencia.

Bajo los siguientes términos:

- **Atribución** — Usted debe dar crédito de manera adecuada, brindar un enlace a la licencia, e indicar si se han realizado cambios. Puede hacerlo en cualquier forma razonable, pero no de forma tal que sugiera que usted o su uso tienen el apoyo de la licenciante.
- **NoComercial** — Usted no puede hacer uso del material con propósitos comerciales.
- **SinDerivadas** — Si remezcla, transforma o crea a partir del material, no podrá distribuir el material modificado.
- **No hay restricciones adicionales** — No puede aplicar términos legales ni medidas tecnológicas que restrinjan legalmente a otras a hacer cualquier uso permitido por la licencia.

Avisos:

- No tiene que cumplir con la licencia para elementos del material en el dominio público o cuando su uso esté permitido por una excepción o limitación aplicable.
- No se dan garantías. La licencia podría no darle todos los permisos que necesita para el uso que tenga previsto. Por ejemplo, otros derechos como publicidad, privacidad, o derechos morales pueden limitar la forma en que utilice el material.

Índice

1. Introducción.....	2
2. El experimento: la metodología	2
3. La capacidad predictiva de la evaluación ex ante del impacto social	3
4. La aportación de la IA en la evaluación ex ante del impacto social	7
5. Conclusiones	10
Capacidad de predicción	10
Aportación de la IA	11
Aprendizajes para la mejora del proceso de evaluación del impacto social	12

1. Introducción

Este informe presenta los resultados del experimento aleatorizado diseñado para analizar el uso de evaluaciones asistidas por inteligencia artificial (IA) en la valoración del impacto social de las solicitudes recibidas en el programa Neotec, de acuerdo con el protocolo del ensayo V1.2 (octubre de 2025). El experimento combina evaluaciones realizadas por personal del CDTI y por expertos externos en impacto social, con la incorporación controlada de información generada por un modelo de IA.

El informe describe, en primer lugar, la ejecución del experimento, incluyendo el reclutamiento de evaluadores y la implementación operativa, así como el tamaño y la composición de las muestras analizadas. A continuación, presenta resultados preliminares sobre la capacidad predictiva de las evaluaciones humanas y sobre los efectos asociados a la exposición a la información generada por la IA, distinguiendo entre tipos de evaluadores y métricas de análisis.

Finalmente, el informe sintetiza las principales conclusiones preliminares y delimita el alcance de los resultados presentados.

2. El experimento: la metodología

La muestra final del experimento ha estado compuesta por un total de 116 personas participantes. De ellas, 30 correspondían a expertos externos en impacto social, 52 a personal voluntario del CDTI que no participa habitualmente en el programa Neotec y 34 a personal evaluador del CDTI con experiencia en el programa Neotec. En conjunto, las personas participantes revisaron un total de 348 propuestas, dando lugar a 828 evaluaciones individuales.

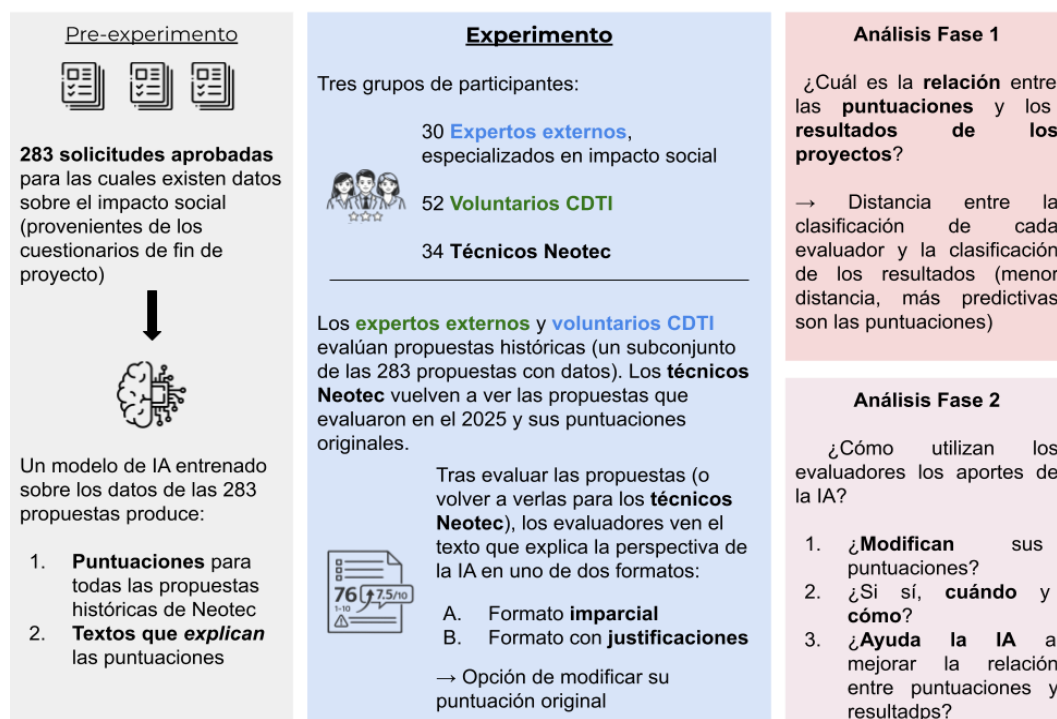
El experimento ha constado de dos fases. En la primera de ellas el objetivo ha sido comparar los resultados de la evaluación ex ante del impacto social de las solicitudes por parte de los tres grupos seleccionados en la muestra y una herramienta de inteligencia artificial (IA). Para ello se ha desarrollado un sistema de IA específico basado en el análisis automático de parte de los planes de empresa presentados por las empresas solicitantes, debidamente anonimizados. De esta manera, se genera una estimación estructurada y consistente del potencial impacto social a partir de la información disponible en las propuestas. En total se utilizaron 279 solicitudes aprobadas en los años 2021-2023 para las cuales existen datos de las encuestas internas realizadas por el CDTI al finalizar los proyectos. Estas encuestas contienen datos sobre distintos indicadores de impacto social de los proyectos, según las declaraciones de las propias empresas, y se utilizan como referencia en el sistema de IA para la cuantificación de dicho impacto.

En una segunda fase, el objetivo es comprobar hasta qué punto un modelo de IA generativa puede ser utilizado como herramienta de apoyo en el proceso de evaluación ex ante llevado a cabo por el personal técnico. Para ello, se genera un modelo basado en Google Gemini, utilizando el texto de la propuesta y la

puntuación ya calculada en la fase anterior, dándola por válida, para identificar evidencias relevantes y construir una justificación coherente de la nota asignada. Toda la información relevante de las propuestas y el *feedback* de la IA se integró en una plataforma específica desarrollada para este ejercicio.

En el cuadro 1 se resume el diseño y la muestra utilizada en el experimento en su conjunto, para facilitar la comprensión de su ejecución y los resultados que se describen a continuación.

Cuadro 1: Esquema del diseño del experimento



3. La capacidad predictiva de la evaluación ex ante del impacto social

En la Fase 1 del experimento se comparan las puntuaciones otorgadas por los tres grupos de evaluadores (expertos externos en impacto social, personal voluntario del CDTI y evaluadores técnicos del programa Neotec) con los resultados efectivos de los proyectos evaluados entre 2021 y 2023, declarados por las empresas en las encuestas de fin de proyecto. Este indicador de resultados se considera el mejor disponible, aunque, como se detalla en el Cuadro 2 tiene sus limitaciones.

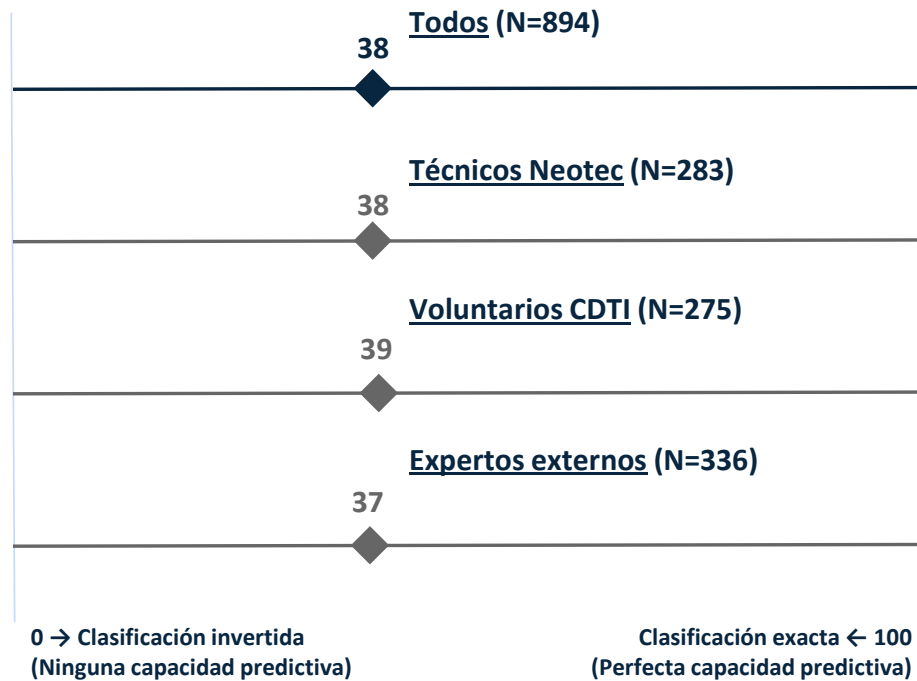
La pregunta que se quiere responder es hasta qué punto la clasificación implícita en las puntuaciones de los evaluadores coincide con la clasificación basada en los resultados observados en las encuestas de monitorización. La expectativa es que, de media, ambas clasificaciones estén alineadas: si un proyecto se sitúa entre los primeros en términos de resultados, también debería situarse entre los primeros en

términos de puntuación. A partir de estas dos clasificaciones se calcula el indicador que servirá para realizar el análisis: Distancia absoluta media entre la clasificación basada en resultados y la clasificación implícita en las puntuaciones, para todas las combinaciones de evaluador-solicitud (un total de 894).

Para facilitar la interpretación de este indicador, se estandariza en una escala de 0 a 100, donde 100 representa una clasificación perfecta, es decir, la clasificación basada en las puntuaciones coincide exactamente con la clasificación basada en los resultados, y 0 representa una clasificación completamente invertida (es decir, la propuesta situada en primer lugar según las puntuaciones ocupa el último lugar según los resultados, y así sucesivamente).

Como puede observarse en el gráfico 1, los valores observados son muy similares entre los tres grupos de evaluadores.

Gráfico 1: Capacidad predictiva de los tres grupos de evaluadores

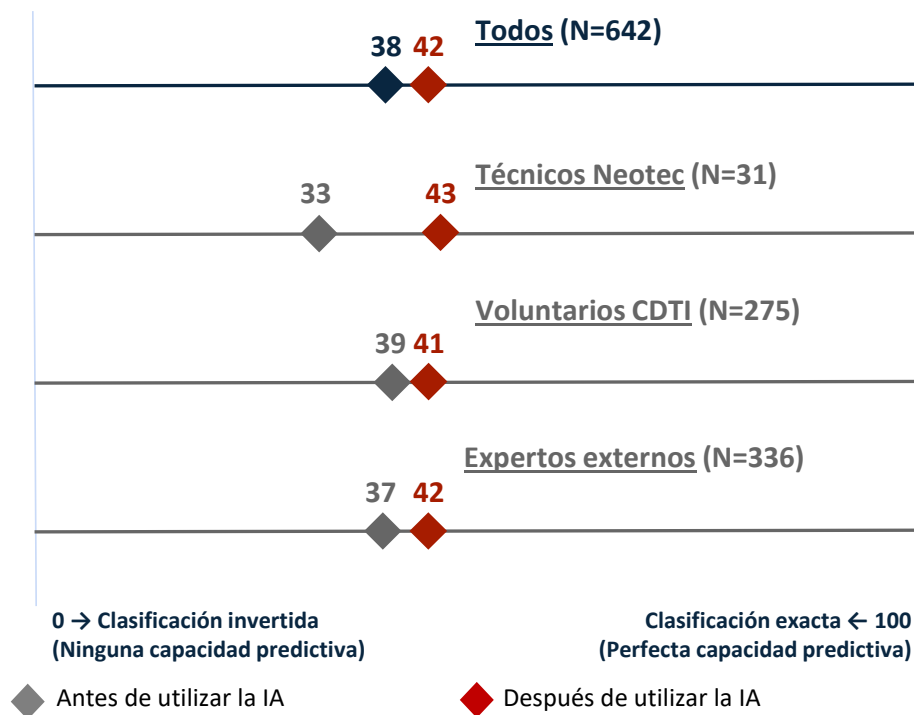


Con el objetivo de comprobar si el uso de la IA mejora la capacidad predictiva del personal evaluador, se incorporan en el análisis las puntuaciones revisadas por estas personas tras recibir la información generada por el modelo de IA utilizado en la Fase 2 del experimento. Los datos correspondientes al personal de CDTI con experiencia en la evaluación de Neotec deben interpretarse con cautela, ya que, en este caso, la muestra utilizada es tan solo de 31 propuestas aprobadas en el periodo 2021-2023. En el

experimento este grupo se centró en revisar principalmente propuestas de 2025, para las cuales aún no se dispone de resultados de las encuestas ex post.

Como se observa en la Tabla 2, los resultados sugieren que la información generada por la IA puede contribuir a mejorar la capacidad predictiva de las puntuaciones, especialmente en el caso del personal del CDTI sin experiencia en la evaluación de Neotec y en los expertos externos. Sin embargo, el efecto observado es modesto y no altera sustancialmente el patrón general identificado en la fase de evaluaciones paralelas.

Gráfico 2: Capacidad predictiva de los tres grupos de evaluadores antes y después de utilizar IA¹



En este contexto, los valores observados sugieren una baja capacidad predictiva de las evaluaciones ex ante, ya que las puntuaciones sólo reproducen parcialmente el ordenamiento de los proyectos según los resultados: aportan información, aunque de forma limitada. Este resultado debe interpretarse con cautela, teniendo en cuenta posibles factores explicativos, como se detalla en el siguiente cuadro.

¹ Los resultados del personal CDTI con experiencia en Neotec se basan en 31 propuestas del periodo 2021–2023, para las que existen datos de cuestionarios de fin de proyecto. Esta diferencia en la muestra da lugar a resultados no directamente comparables con los del conjunto completo.

Cuadro 2: ¿A qué se debe la baja capacidad predictiva de las evaluaciones ex ante del impacto social?

1. **Brecha entre los criterios de puntuación de las convocatorias y las preguntas del cuestionario de fin de proyecto sobre impacto social.** Existe una discrepancia entre los aspectos de impacto social que los evaluadores tuvieron que considerar en la fase de valoración y los objetivos que posteriormente se midieron en las encuestas de fin de proyecto, a partir de las cuales se construyen nuestros datos de “resultado”.
2. **Cambios en los criterios de las propuestas.** El CDTI cambió el criterio de impacto social incluido en las convocatorias de Neotec en 2024. Para las propuestas consideradas en este trabajo, el criterio que usó el personal con experiencia en la evaluación de Neotec es el anterior a esta fecha, mientras que el personal CDTI sin experiencia en NEOTEC y las personas expertas utilizaron el criterio actual.
3. **Los resultados observados pueden contener mucho ruido o no reflejar plenamente el impacto social real.** Los resultados se miden a partir de encuestas ex post autodeclaradas por las empresas, lo que introduce varias fuentes potenciales de error: dificultades de autoevaluación, interpretaciones heterogéneas de las preguntas o posibles sobreestimaciones del impacto.
4. **Existe una elevada incertidumbre en la materialización del impacto social.** Las evaluaciones ex ante pueden capturar razonablemente el potencial de impacto, pero la realización efectiva de ese impacto depende de factores difíciles de anticipar. Proyectos con alto potencial pueden no materializarlo, mientras que otros con expectativas más modestas pueden obtener buenos resultados.
5. **Las propias puntuaciones pueden contener sesgos.** En el caso de expertos externos y personal voluntario del CDTI, es posible que el nivel de dedicación o atención al ejercicio haya sido desigual. No obstante, este factor no explicaría por sí solo la baja capacidad predictiva observada entre el personal del CDTI con experiencia en la evaluación del programa Neotec.
6. **Las puntuaciones pueden verse influidas por características no directamente relacionadas con el impacto social.** Elementos como el estilo de redacción, la calidad percibida de la idea tecnológica o la coherencia general de la propuesta pueden influir en la puntuación de impacto social, incluso cuando no guardan una relación clara con los resultados finales.
7. **Las propuestas no siempre describen adecuadamente la realidad del proyecto.** Las empresas pueden tener dificultades para describir de forma precisa su situación y su potencial de impacto, o bien proporcionar información insuficiente o excesiva para realizar una estimación informada de los resultados futuros.
8. **El impacto social es un concepto complejo y multidimensional.** No se trata de un resultado único y lineal, sino de un conjunto de dimensiones que pueden ser parcialmente independientes o incluso contradictorias. En consecuencia, distintas personas pueden asignar puntuaciones muy diferentes al mismo proyecto en función de qué dimensiones consideren más relevantes.

4. La aportación de la IA en la evaluación ex ante del impacto social

La Fase 2 del experimento consistió en analizar cómo la información generada por la IA influye en el comportamiento de los evaluadores y, en particular, en la actualización de las puntuaciones asignadas a las propuestas.

El diseño del experimento permite analizar qué efecto tiene la IA en la decisión de modificar la puntuación y en la magnitud de dicho cambio, teniendo en cuenta cuatro categorías de variables explicativas:

1. Recibir la aportación de la IA
2. Recibir la aportación de la IA con distintos formatos: imparcial vs. con juicios de valor, asignado aleatoriamente
3. Divergencia entre la aportación de la IA y la valoración inicial del personal evaluador: IA sugiere una puntuación más baja vs. IA sugiere una puntuación más alta. Magnitud de la diferencia.
4. Divergencia entre la aportación de la IA y los resultados reales del proyecto: IA más alineada con los resultados reales que la puntuación del personal evaluado (IA correcta) vs. IA menos alineada con los resultados reales que la puntuación del personal evaluador (IA incorrecta).

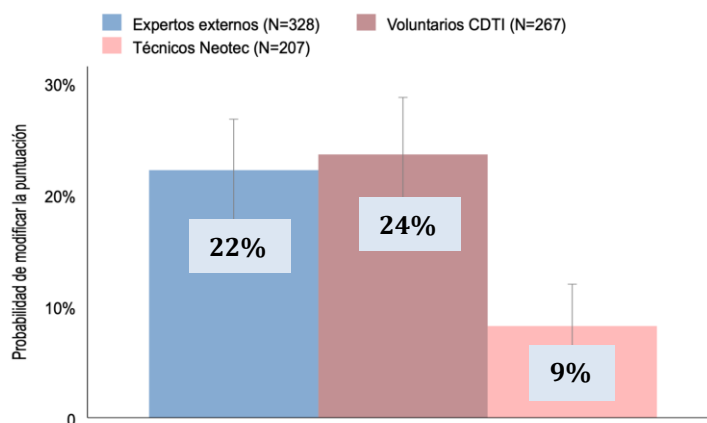
En la Tabla 1 se presentan los resultados detallados de cada una de estas indagaciones.

Los principales mensajes de estos resultados se pueden resumir en dos puntos:

1. Los evaluadores utilizan la IA selectivamente.

- ❖ Como se observa en el Gráfico 3, en la mayoría de los casos los evaluadores no modifican su puntuación (aproximadamente en el 80% de propuestas). Los Técnicos Neotec son los que menos cambian (75% no modifica ninguna propuesta en comparación al 30% en los demás grupos).

Gráfico 3: Probabilidad de modificar la puntuación

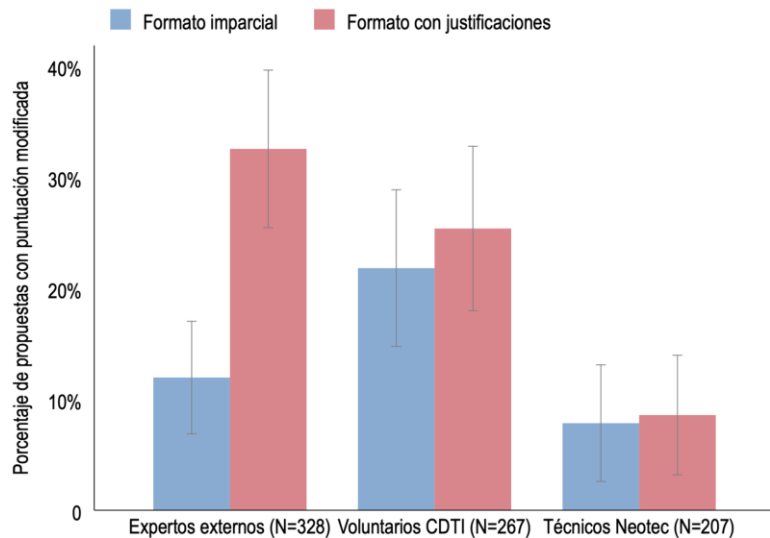


- ❖ La probabilidad de que los evaluadores modifiquen sus puntuaciones es más alta cuando la IA:
 - ➔ Sugiere bajar la puntuación (para evitar el riesgo de “falsos positivos”)
 - ➔ Sugiere una puntuación más alejada de la suya (se aporta información novedosa y útil)
 - ➔ Proporciona una puntuación más correcta (más cercana a los resultados reales)

2. El diseño importa: el formato de la IA cambia el comportamiento.

- ❖ Presentamos el texto de la IA en uno de estos dos formatos:
 - ➔ Formato imparcial: resume las pruebas identificadas por la IA sin ninguna valoración (solo hechos).
 - ➔ Formato con justificaciones: resume las mismas evidencias pero utilizando un lenguaje persuasivo (juicios de valor).
- ❖ Como se observa en el Gráfico 4, el formato narrativo de la IA altera la tasa de aceptación del evaluador

Gráfico 4: Porcentaje de propuestas con puntuación modificada



Decisión de modificar la puntuación inicial tras recibir la aportación de la IA

Variables explicativas	Variables dependientes			
	Decisión de modificar la puntuación	Magnitud y dirección del cambio		
Recibir la aportación de la IA	<p><i>Frecuencia con la que se modifica la puntuación (% evaluaciones):</i> Externos: 23%* CDTI no Neotec: 24%* CDTI Neotec: 8%*</p> <p><i>Personas que cambiaron al menos una puntuación (% personas):</i> Externos: 80% CDTI no Neotec: 65% CDTI Neotec: 25%</p>			
Recibir la aportación de la IA con distintos formatos (imparcial y con juicios de valor)	<p><i>Frecuencia con la que se modifica la puntuación con IA imparcial (% evaluaciones): 14%*</i> Externos: 12%* CDTI no Neotec: 22% CDTI Neotec: 8%</p> <p><i>Frecuencia con la que se modifica la puntuación con IA valorativa (% evaluaciones): 24%*</i> Externos: 33%* CDTI no Neotec: 26% CDTI Neotec: 9%</p>	<p><i>Magnitud de la modificación (puntos en escala 0-10) cuando:</i></p> <p><i>con IA imparcial</i></p> <p><i>con IA valorativa</i></p>		
		<p><i>la IA sugiere bajar la puntuación</i></p>	-0.3	-0.4
		<p><i>la IA sugiere subir la puntuación</i></p>	+0.05*	+0.4*
		<p><i>la diferencia de puntuaciones está por encima de la media</i></p>	+0.30	+0.59
		<p><i>la diferencia de puntuaciones está por debajo de la media</i></p>	+0.14	+0.26
Divergencia entre la aportación de la IA y la valoración del personal evaluador (IA sugiere bajar la puntuación vs. IA sugiere subir la puntuación y magnitud de la diferencia)	<p><i>Frecuencia con la que se modifica la puntuación (% evaluaciones):</i> -cuando IA sugiere bajar: 25% -cuando IA sugiere subir: 14%</p> <p>-cuando la diferencia de puntuaciones está por encima de la media: 25%* -cuando la diferencia de puntuaciones está por debajo de la media: 14%*</p>	<p><i>Magnitud de la modificación (puntos en escala 0-10):</i> - cuando IA sugiere bajar: -0.3 -cuando IA sugiere subir: +0.05</p> <p><i>Frecuencia con la que se modifica la puntuación en la misma dirección que la IA (% evaluaciones):</i> Externos: 89% CDTI no Neotec: 91% CDTI Neotec: 77%</p>		
Divergencia entre la aportación de la IA y los resultados reales del proyecto (IA correcta vs. IA incorrecta)	<p><i>Frecuencia con la que se modifica la puntuación (% evaluaciones):</i> -cuando IA es "correcta": 25%* Externos: 26%* CDTI no Neotec: 26% CDTI Neotec: 11%</p> <p>-cuando IA es "incorrecta": 16%* Externos: 14%* CDTI no Neotec: 20% CDTI Neotec: n.d.</p>	<p><i>Magnitud de la modificación (puntos en escala 0-10):</i> -cuando IA es "correcta": - con IA imparcial: +0.33 - con IA valorativa: +0.54 -cuando IA es "incorrecta": - con IA imparcial: +0.12 - con IA valorativa: +0.28</p>		

* Resultados estadísticamente significativos. Los coeficientes indican el efecto en las variables dependientes en relación a la unidad de medida de cada una de ellas.

5. Conclusiones

Capacidad de predicción

Los resultados de la Fase 1 permiten extraer varias conclusiones relevantes sobre la capacidad predictiva de las evaluaciones ex ante y sobre el papel que pueden desempeñar tanto los evaluadores humanos como la IA en la valoración del impacto social de los proyectos Neotec.

En primer lugar, las puntuaciones ex ante asignadas por **los evaluadores humanos muestran una capacidad predictiva limitada** con respecto a los resultados efectivos de los proyectos. Esta limitada capacidad predictiva no es un problema atribuible a un grupo específico de evaluadores, sino una característica más general del ejercicio de evaluación ex ante del impacto social. Los análisis descriptivos, las métricas de distancia y las regresiones confirman que ninguno de los perfiles evaluados destaca de forma clara en términos de alineación con los resultados ex post.

En segundo lugar, la evidencia apunta a que **la evaluación ex ante del impacto social está sujeta a una elevada incertidumbre estructural**. Las puntuaciones parecen capturar, en el mejor de los casos, el potencial de impacto de los proyectos, pero no su realización efectiva. Factores como la brecha entre el criterio de evaluación y las preguntas del cuestionario de fin de proyecto, el carácter autodeclarado de estas encuestas, la complejidad y multidimensionalidad del concepto de impacto social y la dificultad de anticipar la ejecución real de los proyectos limitan la capacidad de cualquier evaluación previa para predecir resultados con precisión.

En tercer lugar, las puntuaciones generadas por el modelo predictivo de la **IA presentan un rendimiento ligeramente superior** a los evaluadores humanos en la predicción del índice de resultados de impacto social, aunque cabe subrayar que la muestra es limitada. Esto indica que un apoyo de IA puede aportar valor al análisis de los evaluadores.

En cuarto lugar, **la exposición de los evaluadores a la información generada por la IA mejora ligeramente la capacidad predictiva** de las puntuaciones, especialmente en el caso del personal voluntario del CDTI, los evaluadores externos y en el análisis agregado. Estas mejoras son consistentes en dirección (reducción de la distancia entre rankings, aumento del R^2 y reducción del error), pero modestas en magnitud, y no alteran de forma sustancial el patrón general observado en las evaluaciones paralelas.

En conjunto, la Fase 1 pone de manifiesto tanto las limitaciones inherentes a la evaluación ex ante del impacto social como el potencial de herramientas como la IA para apoyar y complementar este proceso, especialmente cuando se combinan con estrategias de agregación de información. Para garantizar la robustez de estos resultados, el equipo de IGL realizó los análisis correspondientes (regresión paramétrica, análisis del R^2 y correlación de Spearman).

Aportación de la IA

Los resultados de la Fase 2 permiten extraer varias conclusiones robustas sobre cómo interactúan los evaluadores con la información generada por la IA y bajo qué condiciones esta influye efectivamente en sus decisiones.

En primer lugar, **la IA no sustituye el juicio humano ni induce revisiones automáticas**. En la mayoría de los casos, las puntuaciones iniciales se mantienen sin cambios tras la exposición a la IA. No obstante, una proporción relevante de evaluadores, especialmente entre expertos externos y personal voluntario del CDTI, revisa al menos una de sus puntuaciones, lo que indica que la IA es considerada y utilizada de forma selectiva, no mecánica.

En segundo lugar, **la receptividad a la IA varía de forma sustancial entre perfiles de evaluadores**. Los técnicos del programa Neotec muestran una propensión significativamente menor a modificar sus puntuaciones que los otros grupos, tanto a nivel de propuesta como a nivel de evaluador. Y cuando cambian su puntuación, también son menos propensos a hacerlo en la dirección que marca la IA en comparación con el resto. Esto sugiere diferencias estructurales en la forma en que los distintos perfiles integran señales externas en su proceso de evaluación.

En tercer lugar, cuando los evaluadores deciden modificar sus puntuaciones, **los cambios tienden mayoritariamente a alinearse con la orientación de la IA**, aunque con ajustes de magnitud moderada. Aproximadamente un 90% de las modificaciones se producen en la misma dirección que la valoración de la IA, lo que indica que esta actúa como una fuerte señal orientadora, aunque no lleva a modificaciones mecánicamente.

En cuarto lugar, **el formato en el que se presenta la información de la IA importa**. En promedio, el uso de un lenguaje más persuasivo incrementa de forma significativa la probabilidad global de revisión. Este efecto se percibe más en los expertos externos, y en menor medida para los voluntarios del CDTI, mientras que para los técnicos Neotec el formato no parece importar.

En quinto lugar, **la magnitud de los cambios depende tanto del formato como del contexto del desacuerdo con la IA**. El formato imparcial induce ajustes más graduales y simétricos, con magnitudes indistinguibles de 0 cuando la brecha entre la puntuación del evaluador y la puntuación subyacente de la IA es pequeña, o cuando el aporte de la IA sugiere aumentar la puntuación. En cambio, el formato con justificaciones produce cambios significativos en la puntuación en la dirección de la IA, incluso cuando la brecha con la IA es pequeña o cuando el aporte de la IA sugiere incrementar la puntuación.

Por último, y de forma especialmente relevante, **los evaluadores no siguen ciegamente a la IA**. Cuando la orientación de la valoración de la IA es inconsistente con los resultados efectivos de los proyectos, los evaluadores tienen menos probabilidad de cambiar su puntuación y, bajo el formato imparcial, la magnitud del cambio es indistinguible de 0. En cambio, cuando la IA señala una dirección alineada con los resultados reales, los evaluadores ajustan sus puntuaciones en esa dirección, independientemente del formato. Esto sugiere que los evaluadores son sensibles a la calidad de la señal de la IA y ejercen un control crítico sobre su uso.

En conjunto, estos resultados muestran que la IA funciona principalmente como un dispositivo de apoyo cognitivo: no reemplaza el juicio experto, pero puede influir en cuándo, cómo y en qué dirección se revisan las puntuaciones, especialmente en contextos de mayor incertidumbre o desacuerdo.

Aprendizajes para la mejora del proceso de evaluación del impacto social

Los resultados del experimento permiten extraer varias lecciones transversales sobre cómo fortalecer la evaluación del impacto social en el CDTI y sobre el papel que pueden desempeñar, de forma realista, las herramientas de apoyo basadas en inteligencia artificial.

En primer lugar, los datos muestran que, en el estado actual, **la información generada por la IA no se tiene en cuenta en la mayoría de los casos**, incluso cuando aporta una señal potencialmente útil. Cerca del 80% de las puntuaciones no se modifican tras la exposición a la IA, lo que apunta a una fuerte inercia evaluativa y a una actitud de cautela, o desconfianza, frente a herramientas automatizadas. Esto es especialmente acusado entre los técnicos de Neotec.

Este comportamiento no debe interpretarse necesariamente como un fallo del evaluador individual, sino como el reflejo de un entorno institucional poco tolerante a la incertidumbre y a la revisión explícita del juicio inicial. En este contexto, **rectificar una puntuación puede percibirse como costoso**, mientras que mantenerla constituye la opción por defecto más segura. Desde esta perspectiva, el valor futuro de la IA no depende únicamente de su precisión técnica, sino también de crear condiciones organizativas que legitimen la revisión, el aprendizaje y el feedback.

En segundo lugar, los resultados sugieren que la **dificultad para anticipar el impacto social no depende principalmente del perfil del evaluador**. Los expertos externos en impacto social no muestran una capacidad predictiva claramente superior a la del personal del CDTI. Esto apunta a que las limitaciones observadas no se deben tanto a una falta de conocimiento especializado, sino a factores más estructurales: la incertidumbre inherente a la materialización del impacto, el marco conceptual del impacto social, la calidad y naturaleza de la información disponible en las propuestas y la complejidad multidimensional del propio concepto de impacto social. En este contexto, incorporar más especialización temática, por sí sola, no necesariamente garantiza

mejores predicciones. Sería necesario mejorar la explicación conceptual del impacto social tanto de cara a las empresas como a los técnicos, de forma que se alinee con la demanda real expresada en las encuestas de fin de proyecto. También sería necesario mejorar estas encuestas para que no dependan tanto de información autodeclarada.

En tercer lugar, **la IA muestra potencial como herramienta de apoyo cognitivo**. Su principal valor no reside en “acertar” sistemáticamente más que los evaluadores, sino en introducir una señal adicional que ayude a detectar discrepancias, cuestionar evaluaciones extremas o aportar coherencia intertemporal y departamental. El hecho de que los evaluadores no sigan mecánicamente a la IA, sino que la incorporen de forma selectiva, es una fortaleza desde el punto de vista institucional.

Conviene subrayar, además, que el modelo de IA utilizado en este experimento constituye una **primera versión operativa**, desarrollada por IGL con fines de validación y aprendizaje. Como cualquier sistema basado en datos, su rendimiento puede mejorar mediante la ampliación y depuración de los datos de entrenamiento, la incorporación de nuevas variables y el ajuste iterativo del modelo. En este sentido, los resultados obtenidos deben interpretarse no como una evaluación definitiva del potencial de la IA, sino como una línea de base a partir de la cual es posible seguir afinando y optimizando la herramienta

En cuarto lugar, los resultados subrayan que **el diseño de la interacción entre personas y sistemas de IA es clave**. No todas las señales funcionan igual: los formatos más explicativos parecen útiles cuando se requieren revisiones sustantivas, mientras que formatos más neutrales inducen ajustes más graduales. Esto sugiere que el valor de la IA para el CDTI dependerá tanto de cómo se integre en los flujos de trabajo como de la calidad técnica del modelo subyacente.

Por último, los resultados apuntan a que **mejorar la evaluación del impacto social requiere ir más allá del momento de la evaluación ex ante**. Dado el ruido y la incertidumbre observados, el mayor margen de mejora probablemente reside en reforzar los sistemas de seguimiento ex post, clarificar qué dimensiones de impacto se consideran prioritarias y utilizar esa información de forma sistemática para retroalimentar tanto a evaluadores humanos como a herramientas automatizadas. En este sentido, la IA puede desempeñar un papel complementario, pero su valor dependerá de un ecosistema institucional que combine datos de calidad, procesos de aprendizaje y criterios claros de decisión.