



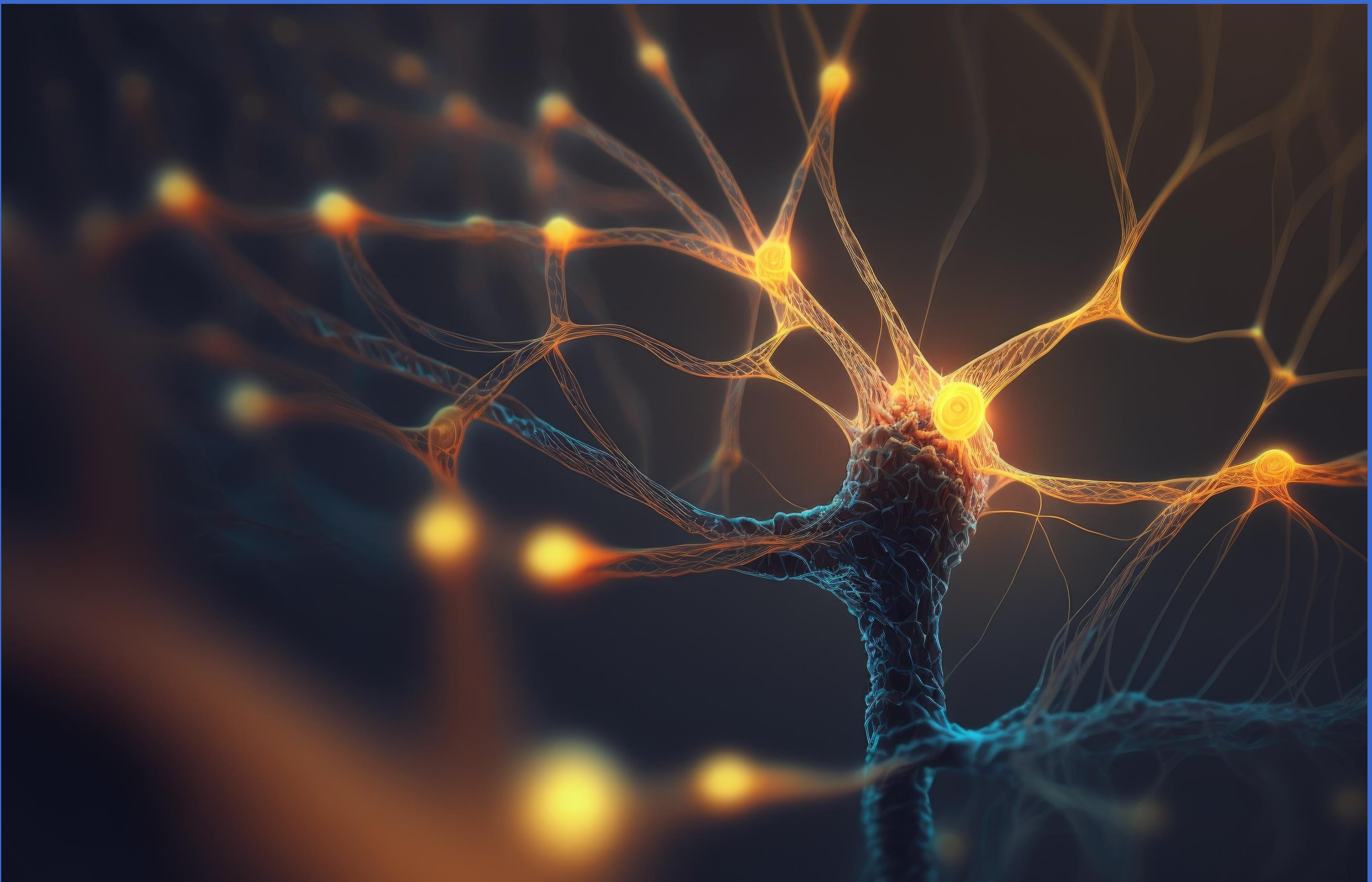
03



**Artículos
de Interés**

3.1.

Nuevos retos y oportunidades: La fusión de la Inteligencia Artificial y las Ciencias del Comportamiento



Este bloque explora cómo la disciplina se reinventa gracias a la tecnología y la ética adaptativa. Desde el uso de participantes sintéticos que aceleran la investigación, hasta la influencia de la IA generativa en decisiones financieras, pasando por los retos éticos del neuro-nudging, aquí se dibujan los horizontes que marcarán la evolución de las ciencias del comportamiento en los próximos años.



Una nueva herramienta para los científicos conductuales: explorar el potencial de los participantes sintéticos para acelerar el estudio de las actitudes y los comportamientos.

Antonio Hanna-Amodio¹ | Chiara Cappellini | Rasha Attar | Ailidh Finlayson

Behavioral Science Group

Este estudio explora si los participantes sintéticos —proxies generados por IA a partir de datos demográficos reales— pueden ofrecer una herramienta práctica para los científicos conductuales. Realizado por el Grupo de Ciencias del Comportamiento (BSG, por sus siglas en inglés Behavioral Science Group) de los Emiratos Árabes Unidos (EAU), el proyecto utilizó el modelo Falcon de código abierto para simular respuestas de una muestra representativa a nivel nacional ($n = 7,652$) y comparó los resultados experimentales con datos reales de dos experimentos en línea ($n = 4,577$). Los resultados sugieren que los participantes sintéticos pueden aproximar con precisión actitudes y opiniones públicas en poblaciones diversas y no WEIRD. Sin embargo, la predicción de resultados conductuales fue más desigual. Aunque los participantes sintéticos a menudo identificaron correctamente las intervenciones con mejor y peor desempeño, tendieron a sobreestimar los tamaños del efecto. En un caso, no obstante, sus predicciones superaron los pronósticos de expertos. Los participantes sintéticos no son una bala de plata, pero sí muestran potencial como una nueva herramienta en la caja de herramientas de la ciencia del comportamiento. A medida que evolucionen los métodos, podrían ayudar a los responsables de políticas a hacer una comprobación inicial de plausibilidad de ideas tempranas, priorizar qué probar y acelerar el diseño de intervenciones efectivas.

¿Qué papel pueden desempeñar los participantes sintéticos en la ciencia del comportamiento?

La inteligencia artificial (IA) está llamada a transformar la forma en que estudiamos el comportamiento humano (Flahavan, 2024). Una aplicación prometedora son los participantes sintéticos, que pueden entenderse simplemente como el uso de datos de encuestados humanos para generar proxies impulsados por IA que pueden simular opiniones y comportamientos.

¹Autor Correspondiente: antonio.hanna-amodio@bsg.ae

Por ejemplo, se puede utilizar un conjunto de datos demográficos para generar participantes sintéticos que simulen cómo responderían personas reales a preguntas sobre temas de interés. Cada participante sintético se crea a partir de características conocidas, como datos demográficos y datos sobre actitudes y comportamientos. En conjunto, estos participantes forman una muestra sintética que puede proporcionar a los responsables políticos información casi instantánea.

Los participantes sintéticos ofrecen una nueva forma de explorar actitudes, intervenciones previas a las pruebas y estudiar patrones de comportamiento a gran escala. Las primeras investigaciones sugieren que las respuestas generadas por la IA pueden aproximarse a las actitudes y comportamientos humanos reales (Argyle et al., 2023; Horton, 2023).

El atractivo de los participantes sintéticos es intuitivo. Permiten a los investigadores generar conocimientos instantáneos de forma relativamente económica (Argyle et al., 2023). Los conjuntos de datos de código abierto pueden utilizarse para crear conjuntos de datos sintéticos, lo que a su vez permite responder a nuevas preguntas de investigación. El trabajo de campo en el mundo real, por el contrario, suele implicar un largo proceso de selección de participantes. Si se considera que son suficientemente precisos, los participantes sintéticos podrían convertirse en una nueva metodología valiosa, que complementaría, en lugar de automatizar, los métodos de trabajo típicos utilizados en las ciencias sociales (Handa et al., 2025).



A pesar de su potencial, su uso y adopción siguen sin explorarse lo suficiente, y se necesita más investigación para avanzar en este enfoque. Por ejemplo, la mayor parte del trabajo en este ámbito se ha centrado en las encuestas de opinión en lugar de en las respuestas conductuales, lo que significa que se sabe poco sobre si los participantes sintéticos pueden predecir los efectos de las políticas y las intervenciones destinadas a cambiar el comportamiento. Y, sin embargo, la ciencia del comportamiento se ocupa en última instancia de comprender y cambiar el comportamiento, no solo en medir las actitudes. Por lo tanto, se trata de una laguna importante que hay que explorar.

Comprender las actitudes y el comportamiento en los EAU con Falcon

Este documento describe un proyecto dirigido por el Grupo de Ciencias del Comportamiento (BSG), un equipo especializado de la Oficina de Asuntos de Desarrollo de los Emiratos Árabes Unidos dedicado a generar un impacto social positivo mediante la aplicación de las ciencias del comportamiento. El BSG se basa en datos sobre las actitudes y los comportamientos nacionales (recogidos a través de encuestas, ensayos e investigaciones colaborativas) para informar y poner a prueba nuevos enfoques metodológicos, como el uso de participantes sintéticos.

El proyecto se basó en fuentes de datos gubernamentales a gran escala de los EAU (n=7652), combinadas con datos de experimentos en línea realizados con muestras representativas a nivel nacional (n=4577). En conjunto, estos datos proporcionaron una base completa para comprender las tendencias y preferencias conductuales dentro de los EAU, un país muy heterogéneo en el que alrededor del 90 % de la población son expatriados de más de 200 naciones.

El estudio se llevó a cabo en colaboración con AI712, utilizando Falcon 180B3, el modelo de lenguaje grande líder en los EAU. Es fundamental destacar que Falcon es de código abierto, lo que es importante para los responsables políticos porque, a diferencia de los modelos cerrados, los sistemas de código abierto pueden implementarse en hardware local, lo que ofrece un mayor control y seguridad de los datos.

Esta es una ventaja nada desdeñable en un mundo en el que la seguridad de los datos es una consideración cada vez más importante. Cabe destacar que este estudio específico se llevó a cabo como un caso de uso pionero de la plataforma AI7I, que ofrece mayor facilidad y flexibilidad a los investigadores.

El uso de Falcon tiene ventajas adicionales. La mayoría de las investigaciones existentes se han basado en modelos basados en GPT de Open AI, que se entrenan principalmente con conjuntos de datos centrados en Occidente (Argyle et al., 2023; Shrestha et al., 2025). Falcon ha demostrado ser comparable a otros modelos de código abierto (Kemper, 2024), lo que lo convierte en una alternativa potencialmente más relevante a nivel regional para generar y evaluar participantes sintéticos.

Nuestras tres preguntas orientativas para la exploración

La creación, el ajuste y la prueba de audiencias sintéticas es un campo aún en evolución. Teniendo en cuenta la bibliografía incipiente pero creciente, este proyecto exploró tanto el potencial de los participantes sintéticos como sus limitaciones actuales en términos de dónde podrían ofrecer ya valor a los responsables políticos y dónde es necesario seguir investigando. Por lo tanto, nos centramos en preguntas de investigación clave para generar conocimientos prácticos para aplicaciones en el mundo real.

Pregunta 1 >>>

¿Pueden los participantes sintéticos simular las actitudes de audiencias no WEIRD?

Los participantes sintéticos han demostrado ser prometedores a la hora de simular con precisión las actitudes de poblaciones reales (Argyle et al., 2023).

Sin embargo, las investigaciones existentes se han centrado en gran medida en poblaciones WEIRD (4), especialmente en los Estados Unidos, lo que ha suscitado inquietudes sobre la capacidad de los participantes sintéticos para generalizar a otros contextos (von der Heyde et al., 2024). Esta tendencia es coherente con una dependencia excesiva de las muestras WEIRD en las ciencias sociales (Henrich, 2010; Muthukrishna et al., 2020).

Una excepción notable es Shrestha et al. (2025), quienes compararon respuestas sintéticas y humanas en encuestas realizadas en los Emiratos Árabes Unidos, Arabia Saudita y Estados Unidos. Utilizando GPT-4, descubrieron que la alineación entre las respuestas sintéticas y las reales era, en general, más débil en las regiones del Golfo que en los Estados Unidos, lo que

sugiere que las predicciones sintéticas pueden generalizarse peor en diferentes contextos culturales. Estos hallazgos ponen de relieve la necesidad de seguir investigando para comprender mejor si la precisión puede disminuir en poblaciones no occidentales y, además, si existen modelos alternativos o técnicas de condicionamiento que puedan mejorar la fiabilidad en entornos diversos.

El objetivo de este estudio es dar un paso adelante mediante la creación de una muestra sintética amplia y representativa de la población de los EAU y evaluar su precisión predictiva en relación con los datos de encuestas existentes.



Pregunta 2 >>>

¿Pueden los participantes sintéticos reflejar con precisión las opiniones públicas sobre políticas?

Como profesionales, nuestro principal interés radica en comprender cómo se puede utilizar a los participantes sintéticos para informar la formulación de políticas en el mundo real. Uno es evaluar la opinión pública sobre políticas futuras. Tradicionalmente, esto requiere encuestas o grupos focales, que a menudo son costosos y requieren mucho tiempo, especialmente cuando se necesita una respuesta rápida.

Si los participantes sintéticos pueden simular de forma fiable cómo podría responder una población a una política existente o propuesta, esto podría ofrecer una herramienta poderosa para comprobar la opinión, señalar áreas de posible resistencia o apoyo y orientar la toma de decisiones tempranas, antes de las pruebas en el mundo real.

Pregunta 3 >>>

¿Pueden los participantes sintéticos predecir resultados experimentales?

La mayoría de los estudios evalúan las respuestas generadas por la IA en entornos actitudinales, por lo que no está claro si los participantes sintéticos pueden predecir con precisión las respuestas conductuales del mundo real a las intervenciones (Bisbee et al., 2024; Durmus et al., 2024). Este estudio va más allá de las actitudes y comprueba si los participantes sintéticos pueden predecir resultados similares a partir de dos experimentos en línea realizados anteriormente por el BSG con muestras representativas a nivel nacional. El primer experimento se centró en la promoción de la salud y evaluó mensajes diseñados para fomentar la inscripción en un evento deportivo (n=4397). El segundo se centró en el comportamiento climático, utilizando marcos conductuales para fomentar la reducción del uso del aire acondicionado (n=4577).

Métodos para crear participantes sintéticos

Indicaciones en un solo paso frente a Indicaciones tipo retrato narrativo

Generar participantes sintéticos es actualmente tanto un arte como una ciencia. Aunque la investigación está evolucionando, seguimos las primeras buenas prácticas identificadas en la bibliografía. Nuestro estudio emplea dos métodos distintos para generar participantes sintéticos: Indicaciones en un solo paso frente a Indicaciones tipo retrato narrativo.

Indicaciones en un solo paso:

se generan participantes sintéticos utilizando una sola sugerencia que contiene variables de entrada estructuradas (n= 7,652). Estas variables se extrajeron de conjuntos de datos del Gobierno de los EAU representativos a nivel nacional e incluyen información demográfica y actitudinal que abarca el bienestar, los valores sociales y las actitudes climáticas. Cada sugerencia se formula en segunda persona (por ejemplo, "Vives en los EAU. Eres mujer. Tienes 29 años"). A continuación, se pide al modelo que genere respuestas a preguntas de la encuesta o intervenciones conductuales basadas en esa identidad. Este método es relativamente más eficiente desde el punto de vista computacional y requiere menos tiempo.



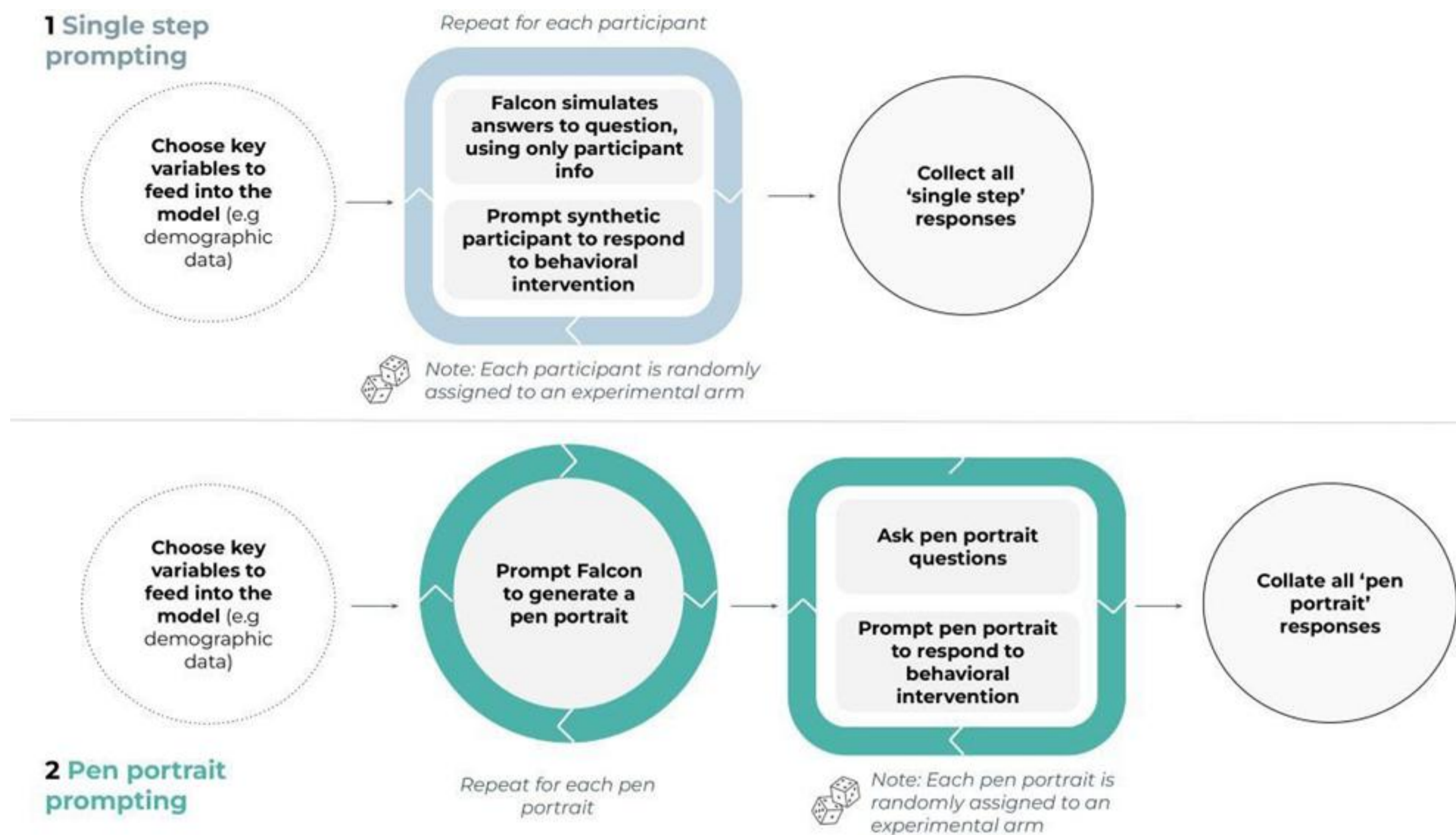


Figura 1: Indicaciones en un solo paso frente a Indicaciones tipo retrato narrativo.

Indicaciones tipo retrato narrativo:

en este enfoque, se proporcionan, al Falcon LLM, las mismas variables de entrada, pero se le indica que genere un retrato literario narrativo: una descripción breve pero detallada de una persona basada en estos atributos. A continuación, estos personajes sintéticos se introducen en el modelo, al que se le pide que prediga cómo respondería esa persona a las preguntas de la encuesta o a las intervenciones conductuales.

Un número cada vez mayor de investigaciones sugiere que este tipo de condicionamiento —proporcionar un contexto personal más amplio— puede mejorar el realismo y la fiabilidad de las respuestas generadas por LLM (Argyle et al. 2013; Flahavan, 2023). Este proceso de dos pasos requiere comparativamente más recursos; sin embargo, se planteó la hipótesis de que podría producir respuestas más precisas.

Evaluación de participantes sintéticos

Actualmente, no existe una métrica ampliamente aceptada para evaluar la precisión de los participantes sintéticos. Nuestro enfoque de evaluación comparó las distribuciones de respuestas sintéticas con las distribuciones de datos reales.

Se trata de una prueba más rigurosa que la simple comparación de puntuaciones medias. Por ejemplo, en una escala de respuestas, si el 40 % de los encuestados reales están muy de acuerdo y el 40 % muy en desacuerdo, un modelo que produzca respuestas mayoritariamente neutras podría parecer engañosamente que tiene una puntuación media alta, aunque se pasa por alto una importante polarización de opiniones. Nuestro protocolo para evaluar a los participantes sintéticos combinó tres enfoques:

Comparación sencilla de resultados:

superposición de gráficos de barras y diagramas de densidad de datos sintéticos frente a datos reales de la encuesta para comprobar la alineación y la divergencia. La comprobación de si las distribuciones son direccionalmente correctas proporcionó una primera evaluación intuitiva de la precisión.

Cálculo de la puntuación de similitud:

esta métrica se derivó directamente de la divergencia de Jensen-Shannon (JS), una medida estadística que compara las distribuciones de probabilidad entre datos reales y sintéticos (Apellániz et al., 2024).

En concreto, la puntuación de similitud es 1 - puntuaciones de divergencia JS, que van de 0 a 1. Una puntuación de 1 indica una alineación perfecta entre las distribuciones sintéticas y reales, mientras que 0 representa una divergencia completa. Esta puntuación de similitud limitada permite una comprensión y comparación más fáciles entre diferentes preguntas y conjuntos de datos.

Reproducción de los resultados experimentales:

La prueba final intentó simular los resultados de dos ECA en línea realizados con muestras representativas de la vida real de los Emiratos Árabes Unidos. Comprobamos si los participantes sintéticos pueden (1)

Meet Ahmed, a 30-year-old Emirati man living in Abu Dhabi with his wife and two children. He was born and raised in the UAE and is proud of his Emirati heritage. Ahmed is outgoing and friendly, always willing to help others.

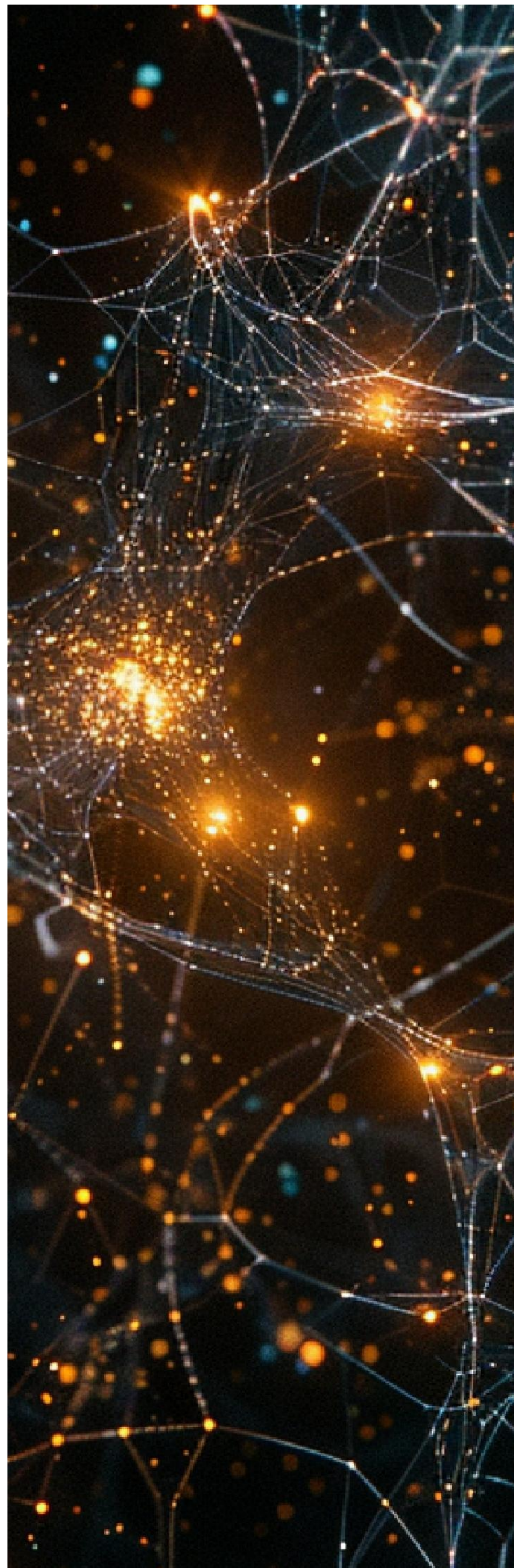
Ahmed comes from a large family and is very close to his parents and siblings. He values family above all else and loves spending time with his wife and children. Ahmed is also a devout Muslim, and his faith is an important part of his life.

Ahmed works as a manager at a local construction company, where he is highly respected by his colleagues for his leadership skills and work ethic. He completed his secondary education in Abu Dhabi and is always looking for ways to improve himself personally and professionally.

Ahmed enjoys playing football and going on family outings to the beach in his free time. He is also an avid reader and enjoys learning about new cultures and ideas.

Figura 2: Ejemplo estilo "Pen Portrait".

identificar correctamente las intervenciones más y menos eficaces y (2) aproximar los tamaños relativos del efecto observados en experimentos en línea con participantes reales. Los participantes sintéticos fueron asignados aleatoriamente a controles o tratamientos y luego se les pidió que predijeran resultados conductuales reales, como hacer clic en un enlace. En los ECA originales, la eficacia de las intervenciones se evaluó mediante un modelo de regresión adecuado, controlando así los datos demográficos. El mismo análisis se replicó en los conjuntos de datos sintéticos.





Resultados

Resultado 01

Los participantes sintéticos muestran potencial para simular respuestas actitudinales de audiencias no WEIRD

En general, los participantes sintéticos obtuvieron resultados muy similares a los de sus homólogos humanos. En una serie de preguntas, las respuestas generadas por la IA reflejaron en gran medida las distribuciones del mundo real, aunque se observaron algunas variaciones, como ilustran las figuras siguientes. El hecho de que estas respuestas procedieran de una muestra representativa a nivel nacional de los EAU, una población muy diversa y predominantemente no WEIRD, indica el potencial de los métodos sintéticos para replicar muestras diversas y heterogéneas.

Contrariamente a lo esperado, las indicaciones para la creación de retratos narrativo no dieron, en promedio, una mayor precisión que las indicaciones de un solo paso. El enfoque de indicaciones de un solo paso produjo una puntuación de similitud de 0,734, mientras que el enfoque de indicaciones para la creación de retratos narrativo produjo una puntuación de 0,730. Dado que las indicaciones para la creación de retratos requieren recursos computacionales y tiempo adicionales (por ejemplo, un mayor uso de API y costes de procesamiento), su valor práctico, en este caso, fue limitado.

Aplicación: Simular percepciones sobre la inteligencia artificial

Un interesante (y muy apropiado) estudio de caso exploró si los participantes generados por IA pueden predecir lo que piensan los seres humanos reales sobre la IA. Cuando se les preguntó si la IA ayudará o perjudicará en mayor medida a las personas en los Emiratos Árabes Unidos durante los próximos 20 años, los participantes sintéticos y las personas reales produjeron distribuciones sorprendentemente similares. En ambos casos, la gran mayoría de los encuestados indicó que la IA ayudará en mayor medida a las personas, y solo una pequeña proporción cree que causará daños.

Resultado 02

Los participantes sintéticos muestran potencial para simular la respuesta pública a políticas de cambio de comportamiento

Una ventaja clave de los participantes sintéticos es su capacidad para evaluar rápidamente la opinión pública sobre cuestiones políticas de actualidad, lo que a su vez crea oportunidades para comprobar rápidamente los niveles de apoyo u oposición. De cara al futuro, se podría imaginar la posibilidad de comprobar cómo cambia la opinión pública en respuesta a los cambios en la formulación de las políticas, lo que ayudaría a los responsables políticos a anticipar mejor cómo serán recibidas sus propuestas.

Aplicación: Simular la opinión pública sobre la prohibición de bolsas de plástico de un solo uso

La prohibición de las bolsas de plástico de un solo uso en los Emiratos Árabes Unidos, anunciada en 2023, constituyó un caso de prueba relevante. También en este caso, los participantes sintéticos mostraron una respuesta similar a la de los encuestados humanos, es decir, un apoyo abrumador a la política.

These days, there are machines that are able to perceive, synthesize and infer information on their own to make decisions - known as artificial intelligence (AI). Do you think that AI will mostly HELP or mostly HARM people in the UAE in the next 20 years?

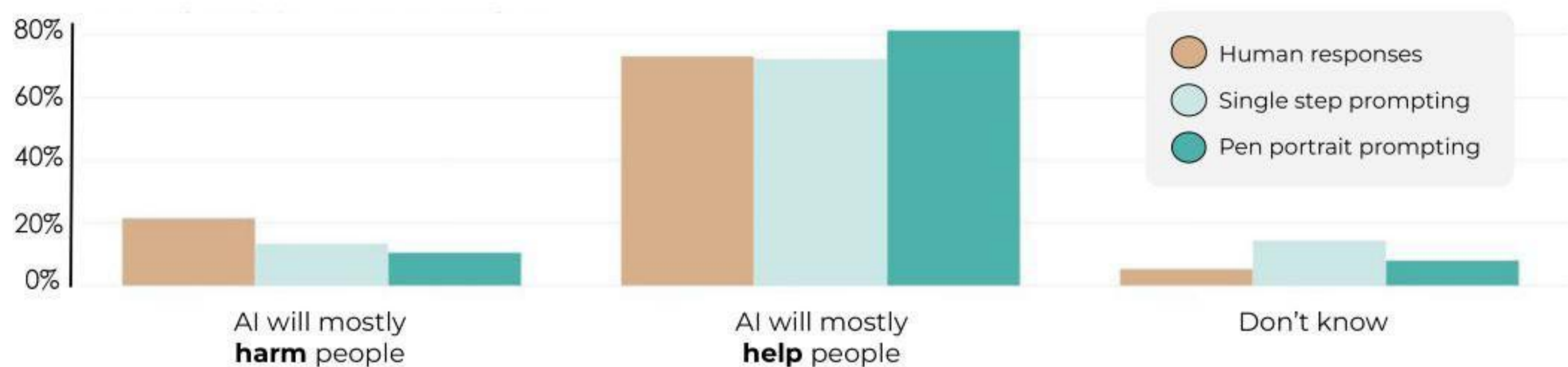


Figura 3: Percepción pública de la IA.

In 2023, the UAE announced it would ban single use plastic bags at the start of 2024, and other single-use plastics by 2026. Do you support or oppose this policy?

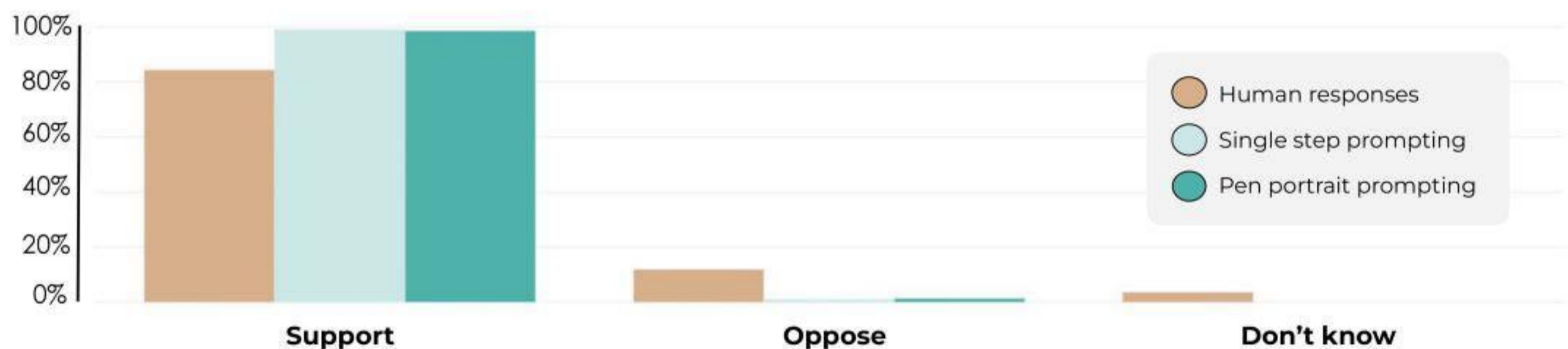


Figura 4: Apoyo a la prohibición del plástico de un solo uso.

Los participantes realizaron predicciones ligeramente más extremas, pronosticando un mayor apoyo y una menor oposición que la opinión pública real (véase la figura 4). Esto sugiere que las respuestas sintéticas se ajustan bien a las tendencias generales, pero pueden amplificar el sentimiento dominante, un patrón observado en otros estudios (Bisbee et al., 2024).

Curiosamente, en nuestro estudio, los participantes sintéticos obtuvieron resultados menos precisos en las preguntas relacionadas con el clima que en otros temas. A diferencia del sesgo general a favor de la sostenibilidad descrito por Shrestha et al. (2025), el patrón aquí fue menos consistente. En un caso, los participantes sintéticos expresaron una mayor convicción sobre el cambio climático que los encuestados reales; en otro, subestimaron el número de personas dispuestas a realizar cambios significativos en su estilo de vida para abordarlo.

Resultado 03

La predicción de los resultados conductuales produjo resultados más dispares

Los resultados de las medidas conductuales fueron más variados que los observados en las simulaciones basadas en actitudes y opiniones. En el lado positivo, nuestro panel sintético pudo distinguir en términos generales entre los tratamientos más eficaces y los menos eficaces.

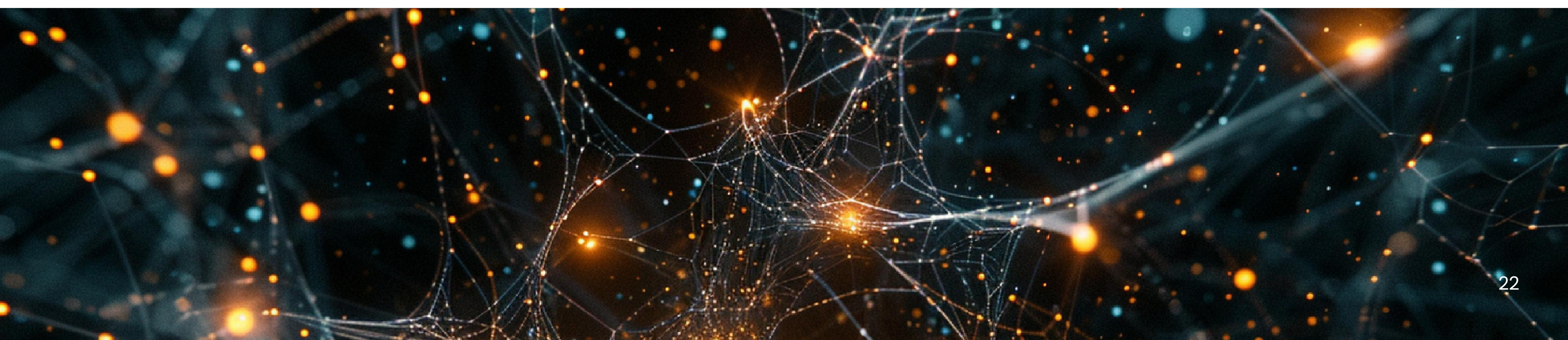
Por ejemplo, identificaron correctamente los grupos de tratamiento con mejor y peor rendimiento en tres de los cuatro resultados que analizamos. Sin embargo, en el otro caso, el tratamiento que se predijo como el más eficaz resultó ser el menos eficaz en el experimento real.

Aplicación: Replicar un experimento en línea que probaba mensajes para reducir el uso de AC

A continuación, analizamos con más detalle uno de los dos experimentos para ilustrar nuestros hallazgos. El estudio original, en el que participaron 4577 residentes de los Emiratos Árabes Unidos, tenía como objetivo reducir el uso del aire acondicionado. Los participantes fueron asignados aleatoriamente para recibir uno de varios mensajes, o ningún mensaje. Estos incluían dos mensajes centrados en el liderazgo (“Apoye la visión nacional”) y variaban la llamada a la acción entre “ajuste el aire acondicionado a 24 °C” o “suba la temperatura 1 °C”, un mensaje sencillo que ofrecía reglas básicas sobre la temperatura (para cuando se está en casa, fuera de casa o en habitaciones vacías) y un mensaje sobre salud que advertía de que el uso excesivo del aire acondicionado podía provocar sequedad en la piel o resfriados.

Como se ilustra en la figura 5, los resultados demuestran el patrón general que observamos: los participantes sintéticos clasificaron sistemáticamente al grupo de control como el menos eficaz y mostraron una fuerte alineación con los participantes humanos reales en cuanto a las intervenciones más eficaces. Sin embargo, los participantes sintéticos sobreestimaron sistemáticamente el tamaño del efecto en los cuatro resultados, a menudo por un amplio margen.

Cabe destacar que estos resultados experimentales deben interpretarse como preliminares. Se basan en dos experimentos en línea, y la mayoría de los resultados miden intenciones declaradas en lugar de comportamientos reales. Estas simulaciones carecen obviamente de un “entorno conductual”, es decir las señales y limitaciones situacionales del mundo real que determinan cómo actúan las personas en cada momento (Lewin, 1951). Por lo tanto, se necesita más investigación para comprender cuándo y cómo los participantes sintéticos pueden predecir de manera fiable los efectos en el mundo real de las intervenciones conductuales.



Implicaciones Políticas

¿Pueden las audiencias sintéticas ser una nueva incorporación a la caja de herramientas del científico del comportamiento?

La evidencia emergente, incluido nuestro estudio, sugiere que una de las aplicaciones más prácticas de los participantes sintéticos es evaluar rápidamente la opinión pública y actitudes, especialmente en contextos limitados por el tiempo o recursos.

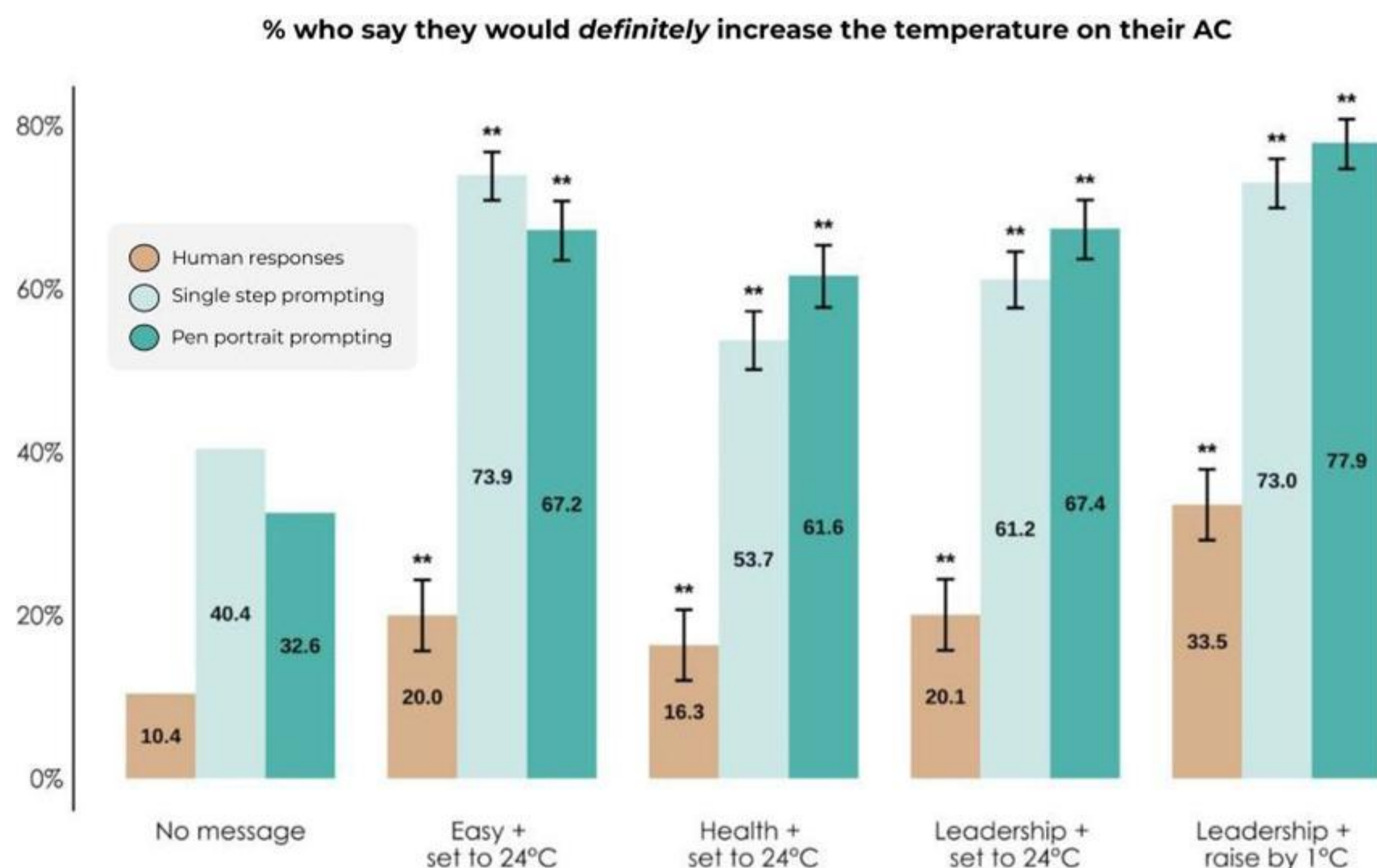


Figura 5: Resultados reales vs. predichos por IA para el experimento de AC. Las etiquetas del eje X se refieren a los distintos tratamientos de encuadre.

Aunque nunca sustituyen a la recopilación de datos del mundo real, las respuestas sintéticas pueden proporcionar una capa adicional de información para tomar decisiones tempranas. Lo hacen ofreciendo una forma de verificar las hipótesis, reducir las opciones y orientar el diseño de investigaciones más específicas. De este modo, los participantes sintéticos pueden desempeñar un papel valioso durante la fase de exploración y preensayo, ayudando a garantizar que los recursos limitados se dirijan hacia donde es más probable que tengan un mayor impacto.

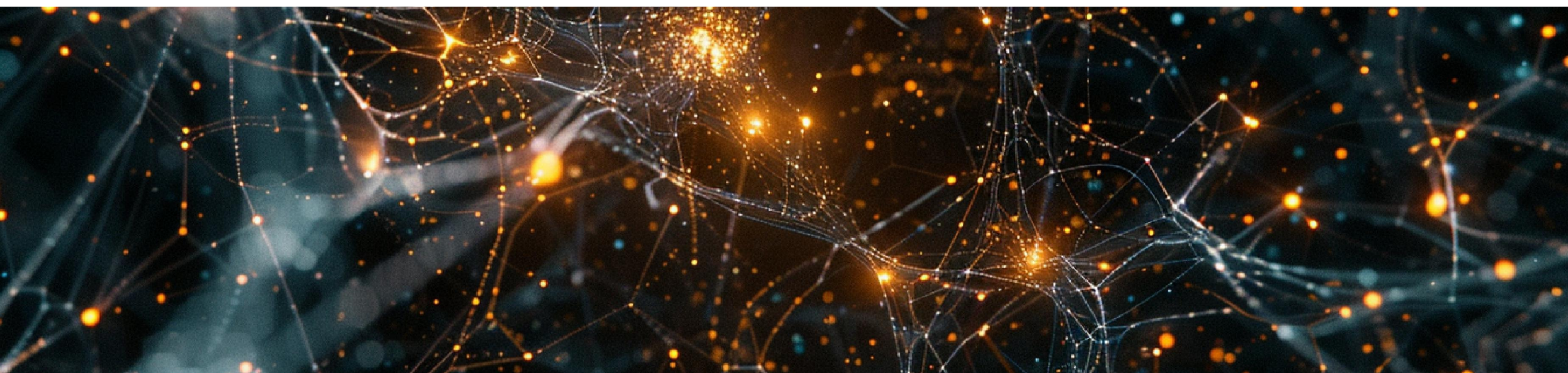
¿Cuándo es bueno “lo suficientemente bueno”?

Nuestros resultados sugieren que el valor de este enfoque radica en proporcionar información orientada en la dirección correcta, lo que ayuda a determinar la opinión general sobre un tema determinado. De hecho, la precisión pareció mejorar cuando se redujeron las escalas de respuesta. En otras palabras, agregar categorías para centrarse en la opinión general, en lugar de en distinciones sutiles, puede ofrecer una visión rápida del panorama general.

Por ejemplo, al evaluar la satisfacción con la vida (véase la figura 6), los participantes sintéticos se alinearon más con los datos del mundo real cuando las respuestas se agruparon en “Satisfecho” frente a “No satisfecho”, en lugar de utilizar una escala más granular. Esto quizá no sea sorprendente y concuerda con conclusiones más amplias de la ciencia del comportamiento. Como señalan Kahneman et al. (2021, p. 199), “la ambigüedad en la redacción de las escalas es un problema general”, lo que hace que la medición precisa de sentimientos similares sea intrínsecamente ruidosa. Por supuesto, en aplicaciones del mundo real, el agrupamiento óptimo puede no conocerse de antemano. No obstante, esto apunta a una implicación práctica: escalas más simples que capten el sentimiento general —por ejemplo, positivo, neutral o negativo— pueden ofrecer conocimientos más fiables en etapas tempranas.

¿Mejor que la intuición del profesional por sí sola?

Al diseñar intervenciones conductuales, las decisiones suelen guiarse por la intuición humana. A veces, esa intuición se basa en una profunda experiencia, pero las suposiciones pueden ser erróneas y los resultados suelen ser sorprendentes. En el espectro de la solidez, la intuición por sí sola se sitúa en el extremo inferior, mientras que la evaluación empírica, como los ensayos aleatorios, sigue siendo el estándar de referencia. Los participantes sintéticos pueden situarse en algún punto intermedio. En este estudio, se evaluó a los participantes sintéticos comparando sus respuestas con datos del mundo real, principalmente sobre opiniones y actitudes. Sin embargo, si su valor principal se encuentra en las primeras etapas del diseño de la intervención, la comparación más significativa podría ser con el criterio de los profesionales: concretamente, las decisiones sobre qué ideas seguir y cuáles se espera que sean más eficaces. Si los datos sintéticos pueden respaldar o mejorar ese criterio, podrían convertirse en una herramienta valiosa para perfeccionar y priorizar las ideas antes de los ensayos sobre el terreno.



On the whole, are you very satisfied, fairly satisfied, not very satisfied, or not satisfied at all with the life you lead?

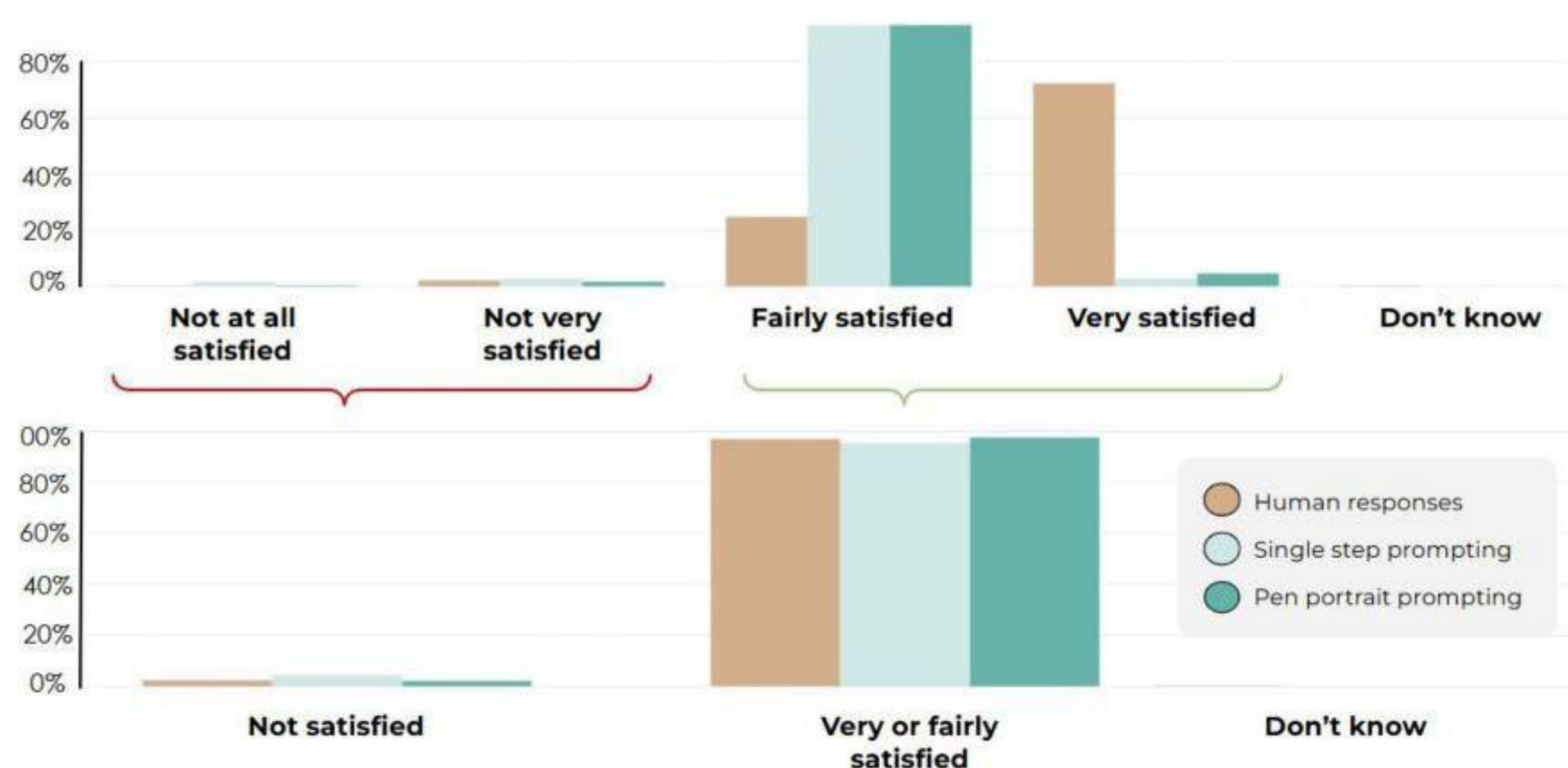


Figura 6: Ilustración del colapso de las escalas de respuesta.



Predicciones humanas vs. de máquinas de resultados experimentales

Para el estudio sobre el aire acondicionado, tuvimos acceso a predicciones previas al ensayo realizadas por cinco científicos del comportamiento y responsables políticos sobre la eficacia de las intervenciones. Como se muestra en la figura 5, ambos conjuntos de datos sintéticos identificaron con precisión las tres intervenciones más eficaces, y el retrato breve clasificó con precisión todas las intervenciones.

En comparación, ninguno de los cinco profesionales predijo correctamente qué grupo obtendría los mejores resultados. Además, tres de los cinco pronosticaron erróneamente que informar a las personas sobre las posibles implicaciones para la salud sería el motivador más eficaz para que redujeran el uso del aire acondicionado. En realidad, este fue el mensaje conductual que obtuvo peores resultados.

En resumen: los participantes sintéticos fueron más precisos que los profesionales en este escenario a la hora de identificar los grupos que obtendrían mejores resultados.

Evaluación preliminar sintética: una herramienta para refinar y preseleccionar ideas

De cara al futuro, los participantes sintéticos podrían desempeñar un papel clave en el diseño de políticas conductuales. Si bien nuestro estudio arrojó resultados dispares, otras investigaciones han sugerido el valor potencial del uso de participantes generados por IA para predecir la eficacia de las intervenciones (Hewitt et al., 2024). Este enfoque podría integrarse como un paso adicional en el flujo de trabajo típico de los proyectos de ciencias conductuales. Consideremos un ejemplo ilustrativo de una campaña destinada a impulsar las tasas de vacunación infantil, en un momento en el que las tasas de inmunización han descendido a nivel mundial (UNICEF, 2023). El proceso podría comenzar con una revisión bibliográfica para identificar temas generales de mensajes eficaces sobre la vacunación (paso 1), seguida de una prueba previa de 20 variantes de mensajes con participantes sintéticos generados por IA (paso 2). A continuación, los 10 mensajes más populares podrían probarse en un experimento en línea con participantes humanos (paso 3) antes de seleccionar los cinco mejores para realizar pruebas de campo en el mundo real (paso 4). Los participantes sintéticos pueden ayudar en la exploración inicial, sin sustituir la observación directa o las pruebas de campo, que a menudo revelan diferencias entre la intención y la acción. Dicho esto, la incorporación de pruebas previas sintéticas podría ayudar a los investigadores y a los responsables políticos a descartar rápidamente las intervenciones más débiles, ahorrando tiempo y recursos antes de llevar a cabo ensayos a gran escala.



CONCLUSIÓN

Los participantes sintéticos no son una solución milagrosa, pero sí ofrecen una promesa genuina. En este estudio, descubrimos que las respuestas generadas por IA pueden aproximarse de manera significativa a los datos de encuestas humanas sobre actitudes y opiniones de poblaciones no WEIRD. La predicción de los resultados conductuales fue más variada, aunque alentadora en algunos aspectos, lo que reveló tanto el potencial de este enfoque emergente como sus limitaciones actuales.

Si bien la recopilación de datos del mundo real es y seguirá siendo esencial, los participantes sintéticos pueden acelerar y ampliar este proceso al proporcionar una forma rápida y económica de comprobar la opinión pública o realizar pruebas previas de las intervenciones en las primeras fases del proceso de diseño de políticas. A medida que mejoren los métodos, también lo harán las oportunidades de incorporar a los participantes sintéticos en el conjunto de herramientas de análisis del comportamiento de forma práctica y complementaria.



“Si bien la recopilación de datos del mundo real es y seguirá siendo esencial, los participantes sintéticos pueden acelerar y ampliar este proceso al proporcionar una forma rápida y económica de comprobar la opinión pública o realizar pruebas previas de las intervenciones en las primeras fases del proceso de diseño de políticas”



LOS AUTORES

Antonio Hanna-Amodio | Chiara Cappellini | Rasha Attar | Ailidh Finlayson

Antonio Hanna-Amodio es asesor de políticas en el Grupo de Ciencias del Comportamiento (BSG), donde lidera trabajos en la intersección entre la inteligencia artificial y el comportamiento humano. Además, diseña políticas informadas por la economía del comportamiento para abordar desafíos sociales en contextos culturales y situacionales diversos. Es licenciado (BA) en Economics and World Philosophies por SOAS University of London y posee un MSc in Behavioural Science de la London School of Economics and Political Science.

Chiara Cappellini es Senior Behavioral Scientist en el BSG, donde supervisa la dirección estratégica de políticas. Asumió este emocionante rol tras años trabajando con BIT en el diseño y la ejecución de proyectos de cambio de comportamiento a escala global. Antes del BIT, Chiara trabajó como evaluadora en los sectores público, privado y del tercer sector. Es licenciada (BSc) en Social Policy por la London School of Economics y tiene un MSc in International Development and Research Methods por la University of Edinburgh.

Rasha Attar es directora del BSG y de la cartera de “insights” más amplia en la Oficina de Asuntos de Desarrollo. Cuenta con más de 15 años de experiencia generando investigación e impacto para el gobierno de los EAU y lideró la creación del BSG como la unidad central de ciencias del comportamiento del gobierno. Es licenciada en Economics por la American University of Beirut, ha completado el programa Behavioral Insights and Public Policy de la Harvard Kennedy School y posee un Executive MSc in Behavioral Science en la London School of Economics.

Ailidh Finlayson es investigadora en el Grupo de Ciencias del Comportamiento (BSG), donde utiliza métodos cuantitativos avanzados para diseñar experimentos y generar “insights” en diversas áreas de política pública. Su trabajo incluye el desarrollo y análisis de ensayos controlados aleatorizados, así como otros diseños experimentales. Posee un MRes in Advanced Quantitative Methods por la University of Bath y un BA in Experimental Psychology por la University of Oxford.

REFERENCIAS:

- Apellániz, P. A., Jiménez, A., Galende, B. A., Parras, J., & Zazo, S. (2024). Synthetic tabular data validation: A divergence-based approach. *arXiv Preprint*. <https://arxiv.org/abs/2405.07822>.
- Argyle, L. P., Busby, E. C., Fulda, N., Gubler, J. R., Rytting, C., & Wingate, D. (2023). Out of one, many: Using language models to simulate human samples. *Political Analysis*, 31(3), 337–351.
- Arora, N., Chakraborty, I., & Nishimura, Y. (2024). Revolutionizing marketing research with a large language model: A hybrid AI-human approach. *SSRN*. <https://doi.org/10.2139/ssrn.4683054>.
- Bisbee, J., Clinton, J. D., Dorff, C., Kenkel, B., & Larson, J. M. (2024). Synthetic replacements for human survey data? The perils of large language models. *Political Analysis*, 32(4), 401–416. <https://doi.org/10.1017/pan.2024.5>.
- Durmus, E., Nyugen, K., Liao, T., Schiefer, N., Askill, A., Bakhtin, A., Chen, C., Hatfield-Dodds, Z., Hernandez, D., Joseph, N., Lovitt, L., McCandlish, S., Sikder, O., Tamkin, A., Thamkul, J., Kaplan, J., Clark, J., & Ganguli, D. (2023). Towards measuring the representation of subjective global opinions in language models. *arXiv Preprint*. <https://arxiv.org/abs/2306.16388>.
- Flahavan, E. (2024). BIT's roadmap for AI & BI: How AI is changing and improving the methodologies we use. *Behavioural Insights Team Blog*. <https://www.bi.team/blogs/bits-roadmap-for-ai-bi/>.
- Flahavan, E. (2023). Can AI accurately simulate cross-national cultural values? An exploration. *Medium*. <https://medium.com/@ed.flahavan/can-ai-accurately-simulate-cross-cultural-values-an-exploration-e4637bfd8901>.
- Handa, K., Tamkin, A., McCain, M., Huang, S., Durmus, E., Heck, S., Mueller, J., Hong, J., Ritchie, S., Belonax, T., Troy, K. K., Amodei, D., Kaplan, J., Clark, J., & Ganguli, D. (2025). Which economic tasks are performed with AI? Evidence from millions of Claude conversations. *arXiv Preprint*. <https://arxiv.org/pdf/2503.04761>.
- Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences*, 33(2–3), 61–135. <https://doi.org/10.1017/S0140525X0999152X>.
- Hewitt, L., Ashokkumar, A., Ghezze, I., & Willer, R. (2024). Predicting results of social science experiments using large language models. *Preprint*. Stanford University & New York University.
- Horton, J. J. (2023). Large language models as simulated economic agents: What can we learn from Homo Silicus? *National Bureau of Economic Research*. <https://www.nber.org/papers/w31122>.
- Kahneman, D., Sibony, O., & Sunstein, C. R. (2021). *Noise: A flaw in human judgment*. HarperCollins.
- Kemper, J. (2024). Falcon 3 series sets new benchmarks for open-source LLMs on a single GPU. *The Decoder*. <https://the-decoder.com/abu-dha-bis-tii-releases-falcon-3-setting-new-bench-marks-for-efficient-open-source-ai-models/>.
- Lewin, K. (1951). *Field theory in social science: Selected theoretical papers* (D. Cartwright, Ed.). Harpers.

REFERENCIAS:

- Muthukrishna, M., Bell, A.V., Henrich, J., Curtin, C.M., Gedranovich, A., McInerney, J., & Thue, B. (2020). Beyond Western, Educated, Industrial, Rich, and Democratic (WEIRD) psychology: Measuring and mapping scales of cultural and psychological distance. *Psychological Science*, 31(6), 678-701. <https://doi.org/10.1177/0956797620916782>.
- Neekhra, B., Kapoor, K., & Gupta, D. (2023). SYNTHPOP++: A hybrid framework for generating a country-scale synthetic population. *arXiv Preprint*. <https://arxiv.org/abs/2304.12284>.
- Santurkar, S., Durmus, E., Ladhak, F., Lee, C., Liang, P., & Hashimoto, T. (2023). Whose opinions do language models reflect? *arXiv Preprint*. <https://arxiv.org/abs/2303.17548>.
- Shrestha, P., Krpan, D., Koaik, F., Schnider, R., Sayess, D., & Binbaz, M. S. (2025). Beyond WEIRD: Can synthetic survey participants substitute for humans in global policy research? *Behavioral Science & Policy*. <https://doi.org/10.1177/237946072413117>.
- UNICEF. (2023). New data indicates declining confidence in childhood vaccines of up to 44 percent- age points in some countries during the COVID-19 pandemic. *UNICEF Press Release*. https://www.unicef.org/press-releases/sowc_2023_immunization.
- von der Heyde, L., Haensch, A.-C., & Wenz, A. (2024). Vox Populi, Vox AI? Using language models to estimate German public opinion. *arXiv Preprint*. <https://doi.org/10.48550/arXiv.2407.08563>.