ASOCIACIÓN NACIONAL DE AVISADORES DE CHILE

Fecha: 19/10/2026 Audiencia: 15.000 Sección: ACTUALIDAD
Vpe: $986.541 Tirada: 5.000 Frecuencia: MENSUAL
Vpe pág: $1.210.000 Difusión: 5.000
Vpe portada: $1.210.000 Ocupación: 81,53%

Consumidores que no existen: Oportunidades y advertencias en el uso de *muestras sintéticas*



Cristóbal Barra
Profesor de Marketing
Universidad de Chile



Javiera Escobar
Estudiante de Doctorado
Universidad de Chile

Imagine que una empresa necesita entrevistar a cien consumidores sobre un nuevo producto. Tradicionalmente, esto exige definir la muestra, reclutar participantes, aplicar entrevistas o encuestas y analizar los datos. Hoy existe otra alternativa: pedir a un modelo de inteligencia artificial que adopte distintos perfiles de consumidores y responda por ellos. En minutos, genera cientos de respuestas, simula entrevistas e incluso resume los principales hallazgos.

Esta posibilidad explica el creciente interés por las muestras sintéticas: conjuntos de respuestas generadas por modelos de lenguaje para imitar a consumidores humanos. Sus ventajas son claras: bajos costos marginales, rapidez y acceso a perfiles difíciles de reclutar.

Sin embargo, estas pueden crear una peligrosa ilusión: asumir que una respuesta bien redactada es representativa del mercado. Un modelo no consume, paga, espera una entrega ni se arrepiente de una compra. Este solo produce respuestas a partir de patrones aprendidos en grandes volúmenes de texto, construyendo el discurso de un consumidor sin poseer su experiencia; diferencia central para comprender su valor y sus límites.

Parecer humano no es lo mismo que representar a los humanos

Las respuestas sintéticas suelen ser fluidas, detalladas y coherentes, e incluso pueden parecer más elaboradas que algunas respuestas humanas. Sin embargo, aún tienen dificultades para captar dimensiones propias del consumidor. Arora, Chakraborty y Nishimura (2025) muestran que los modelos pueden crear perfiles, simular entrevistados, formular preguntas de seguimiento e identificar temas, pero proponen un enfoque híbrido que combine capacidades humanas y artificiales.

En una entrevista exploratoria, una respuesta puede revelar hipótesis, segmentos u objeciones sin ser estadísticamente representativa. En cambio, para estimar intención de compra, disposición a pagar o participación esperada, debe reflejar la diversidad y distribución de la población.

El consumidor sintético suele ser demasiado ordenado

Las personas reales son contradictorias. Declaran preocuparse por el medioambiente y luego eligen el producto menos sustentable; dicen valorar la calidad, pero compran por precio; responden de manera distinta según su ánimo, contexto, información disponible o experiencia previa. Esa inconsistencia no es siempre un error de medición. Muchas veces forma parte del fenómeno que la investigación busca comprender.

Los modelos, en cambio, tienden a producir consumidores excesivamente consistentes. Pueden identificar correctamente la dirección general de una respuesta (por ejemplo, reconocer que un atributo favorable debería mejorar una evaluación), pero presentar menor variabilidad que una muestra humana. En el estudio cuantitativo de Arora et al. (2025), las respuestas sintéticas capturaron razonablemente la orientación promedio de distintas variables, aunque mostraron menor heterogeneidad y reprodujeron imperfectamente las correlaciones observadas entre las respuestas humanas. La calidad mejoró al proporcionar ejemplos previos y contexto adicional, pero no se volvió automáticamente equivalente a la de una muestra real.

Esta reducción artificial de la diversidad puede afectar decisiones relevantes. Una empresa podría concluir que existe un consenso

donde en realidad hay segmentos enfrentados, subestimar consumidores extremos o construir perfiles demasiado homogéneos. También podría obtener resultados estadísticamente "limpios" que ocultan la variabilidad que caracteriza al mercado.

No existe una única muestra sintética

En una investigación tradicional, modificar una palabra del cuestionario puede afectar los resultados. Con las muestras sintéticas, esa sensibilidad puede ser todavía mayor. Las respuestas dependen del modelo utilizado, su versión, las instrucciones entregadas, la definición de las personas ficticias, los ejemplos incluidos y los parámetros que regulan la variabilidad de las respuestas.

La revisión de Sarstedt et al. (2024) muestra precisamente que las comparaciones entre muestras humanas y sintéticas producen resultados mixtos. Los modelos han reproducido algunos efectos relacionados con personalidad, encuadre o preferencias políticas, pero han fallado en otros fenómenos conductuales ampliamente documentados, incluyendo ciertos patrones de riesgo, contabilidad mental y efectos asociados a procesos implícitos. Además, algunas versiones entregan respuestas casi uniformes o demasiado "correctas", en lugar de imitar los errores y sesgos habituales de las personas.

Esto conduce a una primera advertencia: El realismo de una persona sintética no debe evaluarse únicamente por la coherencia de su descripción ni por la calidad literaria de sus respuestas. Debe contrastarse empíricamente con datos humanos relevantes para el fenómeno y el mercado estudiado.

Una muestra más grande también puede producir un error más grande

Generar miles de respuestas sintéticas no corrige un sesgo sistemático. Si el modelo representa incorrectamente las preferencias de los consumidores, repetir esa interpretación solo puede hacer que una estimación equivocada parezca más precisa.

Wang, Zhang y Zhang (2026) muestran este riesgo en estudios de análisis conjunto: combinar directamente datos humanos y sintéticos no necesariamente reduce el error y, en algunos casos, agregar más observaciones artificiales incluso lo aumenta. Su propuesta es utilizar una cantidad de datos humanos para calibrar y corregir la fuente sintética.

La conclusión es clara: un "n" elevado no equivale automáticamente a mejor evidencia. Los datos sintéticos pueden complementar una muestra humana, pero no reemplazarla de forma directa ni tratarse como miles de experiencias independientes.

¿Dónde aportan más valor?

Las muestras sintéticas resultan especialmente útiles en etapas tempranas de la investigación: permiten revisar cuestionarios, detectar ambigüedades, ensayar guías de entrevista, explorar respuestas y anticipar problemas de diseño. También pueden ampliar la búsqueda cualitativa, sugiriendo segmentos o perspectivas que el equipo no había considerado, como ocurrió en el estudio de Arora et al. (2025).

Su principal valor está en enriquecer la reflexión y mejorar el diseño del estudio. La cautela debe aumentar cuando se busca cuantificar preferencias, disposición a pagar o tamaño de segmentos, porque en esos casos la validación con datos humanos es indispensable.

Cinco warnings antes de utilizar consumidores sintéticos

- No confundir plausibilidad con validez. Una respuesta puede sonar auténtica y, aún así, no corresponder al comportamiento de consumidores reales.
- No tratar las observaciones sintéticas como individuos independientes. Todas provienen de un modelo común y pueden compartir sesgos, asociaciones y omisiones.
- No ignorar la falta de heterogeneidad. Promedios razonables pueden coexistir con varianzas, correlaciones y segmentos incorrectamente representados.
- No mezclar automáticamente datos humanos y sintéticos. La combinación requiere validación y procedimientos que reconozcan explícitamente las diferencias entre ambas fuentes.
- No ocultar cómo se produjeron y validaron los datos. El estudio debería reportar el modelo y su versión, los prompts, los parámetros, la construcción de las personas y los controles de calidad. También debe justificar los procedimientos analíticos, contrastar los resultados con datos humanos y evitar ajustes que solo reproduzcan una muestra particular.

El futuro no es humano o sintético

La discusión no debería centrarse en elegir entre muestras humanas o sintéticas, sino en determinar qué función puede cumplir cada una. Utilizadas con cautela, las muestras sintéticas pueden reducir tiempos, mejorar instrumentos y ampliar la exploración; utilizadas como reemplazo directo del mercado, pueden generar una falsa sensación de precisión.

El escenario más prometedor es, por tanto, una investigación híbrida: personas que definen los problemas, aportan experiencia y validan conclusiones, junto con modelos que amplían alternativas y agilizan el análisis. Las muestras sintéticas no eliminan la necesidad de escuchar al consumidor; obligan a hacerlo con mayor rigor.

REFERENCIAS

Arora, N., Chakraborty, I. and Nishimura, Y. (2025), "AI_Human Hybryds for Marketing Research: Leveraging Large Language Models (LLM's) as Collaborators", Journal of Marketing, 89(2), 43-70.

Sardstedt, M., Adler, S., Rau, L. and Schmitt, B. (2024), "Using Large Language Models to Generate Silicon Samples in Consumer and Marketing Research: Challenges, Opportunities, and Guidelines", Psychology & Marketing, 41, 1254-1270.

Wang, M., Zhang, D.J. and Zhang, H. (2026), "Large Language Models for Market Research: A Data-Augmentation Approach", Marketing Science, 45(4), 728-751.