

MEMORIA DE SESIÓN · GLOCAL EVALUATION WEEK 2026

Construcción de índices compuestos con IA: una práctica evaluativa desde El Salvador

Webinar presentado por Consultora Raíces — Investigación para el Desarrollo

Fecha	3 de junio de 2026
Modalidad	Webinar en línea, en el marco de Glocal Evaluation Week 2026
Organiza	Consultora Raíces — Investigación para el Desarrollo (El Salvador)
Ponente	Isis Mundo — Directora ejecutiva y especialista en monitoreo y evaluación
Moderación	Carlos Manuel Montano – Gerente de proyectos

Nota sobre alcance y confidencialidad

Esta memoria documenta el proceso metodológico expuesto durante la sesión, no los resultados del proyecto evaluado. Por acuerdos de protección y confidencialidad, no se identifica a la organización implementadora ni se incluyen datos del caso; los ejemplos se presentan únicamente con fines metodológicos.

1. Presentación de la sesión

La sesión compartió cómo Consultora Raíces utilizó la inteligencia artificial (IA) como asistente metodológico para construir un índice compuesto que permitiera medir, con criterios claros, un indicador de cierre que estaba formulado de forma ambigua. El caso provino de la evaluación final de un programa de mejoramiento de vida ejecutado en el departamento de Sonsonate, El Salvador.

El propósito fue socializar la ruta metodológica completa —desde el problema de medición hasta la interpretación de resultados— de modo que pudiera replicarse en otras evaluaciones. La presentación dejó claro, de principio a fin, que la IA apoya el trabajo técnico pero no sustituye el juicio del equipo evaluador.

Sobre Consultora Raíces

Consultora Raíces es una iniciativa empresarial liderada por mujeres salvadoreñas, comprometida con el fortalecimiento de capacidades, la generación de evidencia y la mejora de la eficacia de intervenciones sociales y humanitarias en El Salvador y la región. Acompaña a organizaciones de sociedad civil, instituciones públicas, organismos internacionales y empresas. Trabaja sobre cuatro componentes:

- **Evaluaciones externas** —finales, intermedias y ex post— según criterios del CAD/OCDE.
- **Construcción de índices compuestos, scorecards y sistemas de medición complejos.** Es una herramienta que la firma ya ha utilizado en múltiples procesos evaluativos.
- **Aplicación crítica de IA a la evaluación,** usándola como asistente metodológico.
- **Diseño de sistemas MEL:** marcos de resultados, indicadores, planes de monitoreo y planificación estratégica institucional.

2. Resumen ejecutivo

El indicador de cierre del programa exigía que al menos el 75% de los grupos familiares participantes hubiera “adoptado el modelo casa saludable”. El problema: no existía una definición operativa de qué significaba “adoptar”, ni ficha técnica, fórmula de cálculo, umbrales ni línea de base robusta. Había mucha información, pero dispersa y sin criterios analíticos de resultado.

La solución fue construir un índice compuesto —el Índice de Adopción del Modelo Casa Saludable (IAMCS)— que tradujo las cinco dimensiones del modelo en criterios observables, ponderados y estandarizados, y permitió valorar no solo si las familias adoptaban el modelo, sino con qué amplitud y profundidad lo hacían.

La IA (ChatGPT y Claude) sirvió para estructurar criterios, ordenar matrices, explorar opciones de estandarización y revisar coherencia lógica. Las decisiones críticas —definición del constructo, ponderaciones, umbrales e interpretación ética— permanecieron en manos del equipo humano.

3. El reto evaluativo

Al revisar el indicador, el equipo identificó que “medir lo que parecía claro” no lo era. Los principales riesgos metodológicos fueron:

- **Sin definición operativa:** no había fichas técnicas que precisaran qué significaba “adoptar el modelo casa saludable”, cuántas prácticas mínimas se requerían ni en qué profundidad.
- **Sin línea de base robusta:** solo existían fichas domiciliarias (registros de entrada), no un diagnóstico que fijara el valor de partida.
- **Información abundante pero dispersa:** mucha evidencia operativa, pero sin criterios analíticos de resultado —un “cementerio de datos” que no se usaba para decidir.
- **Verbos ambiguos:** “adoptar”, “mejorar”, “fortalecer” resultan atractivos conceptualmente pero débiles para una medición operativa.

4. La base conceptual: enfoque de mejoramiento de vida

El programa adoptaba el enfoque de mejoramiento de vida —ampliamente promovido por la cooperación japonesa— que la organización llamó “modelo casa saludable”. Este enfoque ya existía conceptualmente y se organiza en cinco dimensiones. El aporte de la consultora no fue crear el enfoque, sino construir la metodología para medir su adopción.

Dimensión	Ejemplos de prácticas / criterios observables (acotados a las actividades del proyecto)
Salud e higiene	Uso de biofiltro, manejo de agua segura, higiene del hogar, lavado de manos y prácticas de bienestar.
Medio ambiente	Manejo de desechos sólidos, reciclaje en el hogar, no quema de basura y protección de fuentes de agua.
Seguridad alimentaria	Huerto casero activo, diversidad dietética, acceso a alimentos, preparación segura e incorporación de nuevas recetas.
Economía solidaria	Emprendimiento activo, registro de finanzas del hogar, prácticas de ahorro, grupo de ahorro comunitario y mejora de ingresos.
Convivencia	Decisiones compartidas en el hogar, resolución de conflictos, equidad de género, distribución de tareas y plan de vida familiar.

¿Qué existía y qué se construyó?

Existía: el enfoque de mejoramiento de vida, las cinco dimensiones, el concepto de “casa saludable” como modelo, el indicador en el marco lógico y la información operativa.

Se construyó: la definición de criterios observables por dimensión, las escalas de medición, la ponderación explícita, la estandarización de escalas y el índice compuesto final.

5. ¿Qué es un índice compuesto y por qué usarlo?

Un índice compuesto es una medida sintética que combina múltiples dimensiones de un fenómeno en una sola puntuación. Permite resumir información compleja, heterogénea y multidimensional en un valor interpretable, comparable y defendible. En este caso integró las cinco dimensiones del modelo en una sola escala.

Ventajas observadas

- **Hace medible lo ambiguo:** convierte un constructo conceptual en criterios explícitos.
- **Permite ponderar y priorizar:** refleja de forma transparente que no todas las dimensiones pesan igual.
- **Estandariza información heterogénea:** integra datos binarios, ordinales y proporcionales en una escala común.

- **Valora amplitud y profundidad:** no solo si una práctica está presente, sino cuántas dimensiones y con qué intensidad.
- **Fortalece la trazabilidad:** cada dimensión, criterio y peso queda documentado y es auditable.
- **Mejora la interpretación evaluativa:** señala en qué áreas hay más fragilidad y en cuáles más solidez.

6. Construcción del IAMCS, paso a paso

1. **Identificar dimensiones y criterios observables.** Partir del constructo: ¿qué significa “adoptar el modelo” en términos observables? Cada una de las cinco dimensiones se acotó a las prácticas concretas que promovía el proyecto.
2. **Ponderar las dimensiones.** La ponderación es una decisión analítica del equipo evaluador. En este caso se decidió por consenso asignar el mismo peso (20% cada una) porque las actividades eran complementarias y estaban alineadas; en otros proyectos un componente podría pesar más que otro.
3. **Definir criterios observables y fuentes de verificación.** Para cada dimensión se especificaron criterios y sus fuentes. Como no todas las preguntas se miden igual, se combinaron tipos de datos (sí/no, frecuencia, intensidad) para lograr una escala más robusta.
4. **Convertir resultados en un índice final.** Se calculó el puntaje de cada dimensión, se llevó todo a una escala común de 0 a 100, se aplicó la suma ponderada (cada dimensión × su peso) y se obtuvo un índice global.
5. **Clasificar la adopción.** Se ubicó a cada hogar en niveles de adopción (bajo, medio, alto) y de profundidad (ausente, incipiente, consolidado), para distinguir una adopción superficial de una integral.

Dos miradas complementarias

Como el modelo era multicomponente, la evaluación combinó dos aproximaciones para no quedarse solo en el “sí/no” del indicador original:

Mirada	Qué mide	Niveles
Cobertura (amplitud)	Cuántas de las cinco dimensiones presentan evidencia de adopción (al menos una práctica).	Bajo: 0–2 dimensiones · Medio: 3 · Alto: 4–5.
Profundidad (intensidad)	Con qué consistencia se incorpora cada dimensión — cuántos criterios se cumplen dentro de ella.	Ausente · Incipiente (práctica ocasional) · Consolidado (práctica regular y sostenida).

Por qué importaba la doble mirada

Una familia podía “cumplir” el indicador adoptando una sola práctica. Medir solo la cobertura sobreestimaba el logro. La profundidad permitió diferenciar cambios mínimos de prácticas consolidadas y reportar en qué dimensiones la adopción era más sólida o más frágil.

7. El papel de la inteligencia artificial: asistente, no sustituto

Para este caso se usaron ChatGPT y Claude. La IA aceleró la construcción técnica del índice, pero el equipo evaluador conservó el control de todas las decisiones de fondo.

Dónde ayudó la IA	Qué permaneció en el juicio humano
• Traducir dimensiones conceptuales en criterios observables. • Ordenar criterios en matrices de trabajo. • Explorar opciones de estandarización. • Revisar coherencia entre criterios, pesos y umbrales. • Identificar vacíos, redundancias y riesgos de sobreinterpretación. • Organizar y clasificar información dispersa, y apoyar la redacción técnica.	• Definir qué significaba “adopción” en el contexto. • Validar la pertinencia de los criterios con la realidad territorial. • Decidir y justificar las ponderaciones finales. • Definir los umbrales de clasificación. • Interpretar vacíos y contradicciones, evitando sobreinterpretar. • Sostener la lectura ética y la autoría metodológica del diseño.

Flujo de trabajo con IA

1. **Definir el problema** —un indicador complejo con dimensiones dadas pero sin criterios de medición.
2. **Pedir opciones de estructuración** —convertir dimensiones en criterios, ordenar matrices, explorar ponderaciones.
3. **Contrastar respuestas entre dos modelos de IA** —comparar sugerencias y no depender de una sola salida.
4. **Filtrar y validar humanamente** —revisar pertinencia y consistencia metodológica.
5. **Documentar la lógica final** —redactar y ordenar la explicación técnica del índice.

Tipos de prompts utilizados

Ejemplos de las instrucciones empleadas en cada etapa:

Etapa	Ejemplo de prompt
Estructuración	«Con base en estas cinco dimensiones, propone criterios observables para medir el nivel de adopción en una evaluación final, distinguiendo entre evidencia mínima, intermedia y robusta.»
Organización en matrices	«Organiza esas dimensiones en una matriz con columnas para dimensión, criterio observable, tipo de evidencia y escala de puntuación.»
Ponderación	«Propón tres alternativas de ponderación para estas cinco dimensiones y explica ventajas y riesgos metodológicos de cada una.»
Estandarización	«Sugiere opciones para llevar estos puntajes heterogéneos a una escala común de 0 a 100, manteniendo comparabilidad y transparencia.»
Validación crítica	«Revisa esta propuesta de índice e identifica posibles riesgos, redundancias, vacíos de medición o riesgos de sobreinterpretación.»
Redacción metodológica	«Redacta de forma técnica y clara la explicación metodológica de un índice compuesto que mide niveles de profundidad y adopción.»
Organización de información	«Revisa los documentos de esta carpeta y elabora, por cada uno: nombre, propósito, conceptos clave, dimensiones mencionadas y posibles aportes para el índice. No inventes información y señala cuando un elemento no esté claro.»

8. Ética, confidencialidad y control de calidad

- **Anonimización:** no se ingresó información sensible o identificable del cliente. El caso se trabajó anonimizado, con una clasificación previa de la información.
- **La IA no completó vacíos:** no se usó para inventar datos, inferir respuestas faltantes ni reconstruir evidencia inexistente. La organización inicial de la información dispersa se apoyó en Gemini, dentro del entorno corporativo de Google.
- **Control de sesgos y alucinaciones:** toda salida pasó por una etapa de validación humana para detectar simplificaciones, categorías inventadas o mezcla de planos conceptuales.
- **Decisiones críticas no delegables:** definición del constructo, umbrales de clasificación e interpretación de contradicciones quedaron fuera del alcance de la IA.
- **Separar apoyo técnico de autoría:** la IA apoya el proceso, pero la autoría del diseño metodológico es del equipo técnico.
- **Adecuación del tipo de información:** testimonios, relatos y datos personales requieren mayor cautela o se excluyen del flujo con IA.

9. Consideraciones metodológicas clave

- **La ponderación es una decisión analítica**, no objetiva: debe documentarse, justificarse y ser transparente.
- **El índice sintetiza, pero no agota la complejidad**: se lee junto con evidencia cualitativa y el contexto de implementación.
- **Sin línea de base no se afirma mejora longitudinal**: solo puede reportarse la adopción observada al cierre del proyecto.
- **La calidad depende de la evidencia disponible**: si es limitada o inconsistente, el índice también tendrá límites. Triangular es clave.
- **Cuidado con la sobreinterpretación**: un puntaje alto no es éxito automático ni uno bajo es fracaso total.
- **La trazabilidad sostiene la credibilidad**: cada criterio, peso, umbral y regla de agregación debe quedar documentado.
- **Las restricciones operativas importan**: tiempo, presupuesto y acceso condicionan qué puede medirse y verificarse.
- **El autorreporte tiene límites**: variables como ingresos o convivencia se miden por percepción de la persona.
- **La selección de criterios define el índice**: importa qué entra y qué queda fuera, no solo cuánto pesa cada dimensión.
- **Sensibilidad a cambios pequeños**: mover un criterio o un peso puede alterar el resultado; conviene revisarlo.

10. Lecciones aprendidas

1. Un índice no empieza por la fórmula, sino por la definición del constructo: primero hay que aclarar qué se quiere medir y qué cuenta como adopción.
2. Las dimensiones existentes no bastan si no se traducen en criterios observables y comparables.
3. La IA es útil para estructurar, pero no para decidir: no sustituye la definición conceptual, la validación, la interpretación ni el juicio ético.
4. Construir un índice es tomar decisiones metodológicas explícitas, transparentes y éticamente defendibles.

11. Preguntas y respuestas (síntesis)

P1. ¿Cuál fue la principal lección al medir un enfoque multidimensional?

Que el enfoque no se vuelve medible solo por tener muchas dimensiones. La diferencia está en traducir esas dimensiones en criterios observables, comparables y claros, y en justificar por qué se eligen unos criterios y se descartan otros, según la realidad del proyecto.

P2. ¿Qué recomendarían a equipos que parten de indicadores poco definidos?

Empezar por el constructo y no por la fórmula: definir qué significa realmente el indicador, cómo se expresa en la práctica y qué evidencia existe. Asumir que no todo debe entrar en el índice —seleccionar criterios pertinentes para evitar redundancias— y documentar todo el proceso para poder explicarlo y defenderlo.

P3. ¿Qué parte del proceso es replicable y qué fue específico del caso?

La ruta metodológica completa es altamente replicable: desagregar el constructo en dimensiones y criterios observables, explorar opciones de operativización y apoyarse en IA para organizar información. Lo específico fue el contenido sustantivo: las dimensiones del enfoque de mejoramiento de vida y los niveles de adopción definidos para este modelo.

P4. ¿Cuándo conviene usar un índice compuesto y cuándo no?

Conviene cuando el fenómeno es multidimensional y un indicador simple de sí/no resulta insuficiente (por ejemplo, “adoptar” o “fortalecer”). No conviene cuando el indicador ya es directo y claro, cuando no hay evidencia suficiente para triangular, o cuando la información es demasiado débil o dispersa para construir una escala válida.

P5. Para homologar indicadores de distinta naturaleza, ¿qué proceso se realiza?

Una estandarización estadística: asignar valores (por ejemplo 0/1 según exista o no la práctica), ponderar y normalizar todas las escalas a un rango común de 0 a 100, de modo que criterios heterogéneos sean comparables dentro del índice global.

P6. ¿Qué modelos usaron y qué harían distinto en otro caso?

Usaron ChatGPT (una versión extendida) y Claude; el modelo Opus de Claude ofreció mejores resultados que opciones más básicas. En un próximo caso reducirían el número de subcriterios: con cinco dimensiones que se ramifican en 20–30 subcriterios, el procesamiento se vuelve complejo y consume mucho tiempo.

P7. ¿Qué fue más difícil: construir el índice o justificarlo metodológicamente?

Justificarlo. La parte técnica —ordenar dimensiones, criterios, pesos y puntajes— se resuelve. El verdadero reto fue demostrar por qué la estructura era pertinente y hasta dónde podía interpretarse, sobre todo porque el enfoque no estaba operacionalizado ni contaba con sistema de monitoreo previo.

P8. ¿Cómo reaccionó la organización ante el uso de IA?

De forma positiva. En lugar de resistencia, vieron que la propuesta visibilizó una debilidad institucional no advertida y abrió oportunidades para robustecer sus sistemas de monitoreo en futuras fases del programa.

P9. ¿La metodología quedó por escrito y será accesible?

Sí. La ruta es aplicable a cualquier tema que requiera medirse y se publicarán en la plataforma del evento los insumos metodológicos y los prompts utilizados, para consulta pública.

12. Mensaje de cierre

El uso de IA en monitoreo y evaluación está en auge y vale la pena incorporarlo, siempre que se preserve la ética profesional y el juicio evaluador: la IA funciona como herramienta y asistente, pero no sustituye la responsabilidad humana. Bien usada, puede acercar formas rigurosas de evaluar incluso a organizaciones con recursos limitados, fortaleciendo la toma de decisiones para la transformación social.

Consultora Raíces — Investigación para el Desarrollo

Evaluación técnica de proyectos para la transformación social. Puede seguir su trabajo en LinkedIn, Facebook e Instagram.

La presentación, los insumos metodológicos y los prompts utilizados estarán disponibles en la plataforma de Glocal Evaluation Week 2026.