

# ImpactAI: Embedding Generative AI into Evidence-Informed Development Decision-Making

Linxi Wang<sup>1</sup>  

<sup>1</sup>World Bank

## Abstract

This poster presents ImpactAI as an applied AI tool for evidence-informed development practice. We focus on the problem it addresses, the design choices behind source-grounded evidence synthesis, the metadata-driven ensemble architecture, evaluation framework, early lessons from institutional deployment, and priorities for future development. A worked example shows how metadata-aware work flow shapes a typical user workflow. We also use the poster format to invite feedback from researchers, practitioners, and potential users on how tools like ImpactAI can better support responsible and context-aware evidence use.

## 1. Introduction

Development practitioners must often make time-sensitive decisions while navigating a large and fragmented evidence base. Across sectors such as education, health, social protection, governance, jobs, and climate, rigorous impact evaluations and operational studies provide valuable insights into what works, for whom, and under what conditions. [1], [2], [3].

Yet this evidence is often difficult to access and synthesize at the point of decision-making. Program teams may need to quickly identify relevant interventions, compare evidence across approaches, understand likely mechanisms, and assess whether findings from one context may apply to another. These tasks remain highly resource-intensive, creating a persistent gap between the production of rigorous evidence and its use in development operations. Bridging this gap fundamentally poses a challenge within the domains of metadata architecture and knowledge organization. Specifically, it requires evidence to be documented, indexed, and retrieved in a manner that preserves both its originality and context.

**ImpactAI** is a generative AI-enabled evidence platform designed to help close this gap. Developed within the World Bank's Development Economics unit, ImpactAI allows users to ask natural-language questions about development interventions and receive structured, source-grounded responses. The platform is intended for operational teams, researchers, and development practitioners seeking to better use existing evidence in project design, policy dialogue, monitoring, evaluation, and learning.

## 2. Design Principle and Use Cases

ImpactAI is designed as an evidence-grounded decision-support tool, not a replacement for expert judgment. Unlike general-purpose chatbots, ImpactAI is built around curated

## Poster

Available Aug 01, 2026

Correspondence to  
Linxi Wang  
lincy.l.wang@gmail.com



Copyright © 2026 Wang. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International](https://creativecommons.org/licenses/by/4.0/) license, which enables reusers to distribute, remix, adapt, and build upon the material in any medium or format, so long as attribution is given to the creator.

development evidence and structured retrieval, with responses designed to remain traceable to underlying source materials. This design is especially important in development economics, where findings depend heavily on intervention design, target population, implementation quality, institutional setting, and context. By emphasizing source-grounded synthesis, ImpactAI aims to help users assess not only what the evidence says, but also how, where, and under what conditions it may apply.

The platform supports common operational workflows across project preparation, policy dialogue, monitoring, evaluation, and learning. Users can ask practical questions about development interventions, such as which approaches have improved girls' school attendance, how cash transfers affect women's empowerment, or what interventions have strengthened tax compliance in low- and middle-income countries. ImpactAI helps surface relevant studies, summarize key findings, compare intervention options, identify outcome measures, and highlight implementation considerations and evidence gaps. In doing so, it helps teams translate a fragmented evidence base into more structured inputs for operational decision-making.

### 3. System Architecture: Metadata-Driven Ensemble Approach

ImpactAI is built on a curated database of rigorous causal evidence called randomized controlled trials. Each study is integrated into a structured metadata record. The metadata combines bibliographic elements (title, author, date, publisher, identifier) with domain-specific attributes (sector, intervention type, outcome domains, geographic coverage, study design, target population, and quantitative results such as treatment effects). Functioning to facilitate scope retrieval, structural synthesis & comparison, and source attribution, this metadata layer directly underpins a five-stage workflow. During execution, the *query understanding* phase first parses user input to infer relevant metadata facets; the *retrieval* stage then operates over a paper-level index supporting hybrid dense and lexical search, with results filtered and re-ranked via the study metadata. Next, the *standardization* phase standardizes the retrieved effect sizes and runs a random-effect model to calculate the average effect, followed by *grounded generation* which leverages a large language model to synthesize answers strictly constrained to these retrieved papers. Finally, an *attribution* phase that embeds inline citations resolving to complete source records and persistent identifiers. Consequently, the system guarantees end-to-end traceability—from individual claims back to specific paper and bibliographic records. This shifts the evidential burden from parametric model memory to curated structured metadata, while providing a primary defense against hallucination by explicitly acknowledging when the retrieval phase yields insufficient evidence rather than extrapolating.

### 4. Illustrative Workflow

Consider a task team preparing a social protection operation who asks: “What is the evidence on cash transfers and women’s empowerment in South Asia?” ImpactAI infers the metadata facets implied by the question (intervention: cash transfers; outcome domain:

women's empowerment; coverage: South Asia), and retrieves from studies whose records match those facets. It returns a sourced synthesis organized by intervention variant (e.g., conditional versus unconditional transfers), reporting typical outcome measures, implementation considerations, and explicitly flagged evidence gaps. Every claim links to its source study, whose record displays country, design, population, and identifiers, so the team can open the underlying paper, look into contextual fit, and carry a fully attributed summary into project documentation. Relative to manual search across multiple repositories, metadata-aware retrieval narrows the space to contextually relevant studies while keeping their originality—and their limits—in view.

## 5. Evaluation Framework

ImpactAI is evaluated in three dimensions, including the *retrieval quality* such as precision, recall, and coverage of evidence, second, *accuracy* which evaluates the citations counts, treatment-outcomes coverage, effect sizes & completion of task and finally the standardization which standardizes the extracted effect sizes using a very specific standardization protocol and estimation of average effect using an econometric model.

## 6. Progress and Early Lessons

In 2025, ImpactAI moved from prototype development toward broader institutional deployment. The team worked across evidence curation, system design, product development, user testing, governance, and risk review to prepare the tool for use in a high-stakes institutional setting. ImpactAI has now launched its first version for external user registration, marking a transition from internal experimentation to wider engagement with researchers, operational staff, and development partners.

The development process has generated several lessons. First, deploying generative AI in development operations is not only a technical challenge; it also requires institutional alignment, responsible AI review, user education, and expectation management. Second, demand for AI-enabled evidence tools is substantial, but must be balanced against evidence quality, transparency, and sustainability. Third, the value of ImpactAI lies not simply in generating fluent answers, but in structuring evidence in ways that are useful for real operational decisions. Fourth, structured metadata is a core of ImpactAI, given that overall performance, system authenticity, and user trust depend fundamentally on high-quality, consistent evidence curation.

## 7. Looking forward

ImpactAI remains under active development, with next steps focused on expanding evidence coverage, strengthening source-grounding and quality assurance, and improving evaluation, user guidance, and support for questions on external validity, implementation constraints, and evidence gaps.

## Acknowledgments

ImpactAI has been a tremendous team effort. We would like to extend our sincere thanks to all past team members whose work over the years has made this initiative possible.

Luis Eduardo San Martin, Tomáš Dulka, Pedro Vitor Quinta de Castro, Mona Schirmer, John Ngan Pougué Biyong, Arnault Gombert, Ana Arieas, Nadir Khan, Rohit Tripathi, Sabrina Yuseoff, Harshali Ranjan, Jan Willem Tulp, Kushnoor Pathan, Joachim Asare, Dami Rosanwo, Ishaan Bansal, and Ayush Shukla.

## References

- [1] A. V. Banerjee, E. Duflo, and M. Kremer, “The influence of randomized controlled trials on development economics research and on development policy,” *The State of Economics, the State of the World*, pp. 482–488, 2016.
- [2] D. B. Cameron, A. Mishra, and A. N. Brown, “The growth of impact evaluation for international development: How much have we learned?,” *Journal of Development Effectiveness*, vol. 8, no. 1, pp. 1–21, 2016.
- [3] H. White, “The twenty-first century experimenting society: The four waves of the evidence revolution,” *Palgrave Communications*, vol. 5, no. 1, 2019.