

DOCUMENT OF THE INTER-AMERICAN DEVELOPMENT BANK

REGIONAL

**MAXIMIZING IMPACT: USING AI FOR EVALUATING DEVELOPMENT EFFECTIVENESS OF
IDB OPERATIONS AND DESIGNING**

(RG-T4841)

PROJECT DOCUMENT

This document was prepared by the project team consisting of: Team Leader: Scartascini, Carlos (RES/RES); Alternate Team Leader: Berlinski, Samuel (RES/RES); Attorney: Centeno Lappas, Monica Clara Angelica (LEG/SGO); Team Members: Busso, Matias (RES/RES); Escobar Genes, Myriam Helvecia (RES/RES); Latorre Salvador, Lucia (TTD/TTD); Mancilla, Elton Alexander (RES/RES); Marquez Guerra Karina Olenka Stella (RES/RES); Porrua Vigon, Miguel Angel (IFD/ICS); Sarrazin, Tom (RES/RES);

REGIONAL
MAXIMIZING IMPACT: USING AI FOR EVALUATING DEVELOPMENT EFFECTIVENESS OF IDB OPERATIONS AND DESIGNING
RG-T4841

PROJECT SUMMARY

Operation Type:	Technical Cooperation
Sector:	REFORM / MODERNIZATION OF THE STATE
Subsector:	MANAGEMENT FOR DEVELOPMENT RESULTS
TC Taxonomy:	Research and Dissemination
Project Number under the Operational Support Taxonomy:	N/A
Technical Responsible Unit:	RES-Research & Chief Economist
Unit with Disbursement Responsibility (UDR):	RES/RES-Research & Chief Economist
Executing Agency:	Inter-American Development Bank

PROJECT OBJECTIVE

To improve lives, the Inter-American Development Bank (IDB) has committed to increasing the impact and scale of our results across Latin America and the Caribbean. This Technical Cooperation will contribute to it by increasing access to impact evaluations on a scale and showing that our interventions actually lead to better outcomes for people, and how. Impact evaluations are not simply a technical exercise—they are essential for understanding whether our interventions actually lead to better outcomes for people, and how. Without credible evidence of impact, we risk investing in policies or programs that are ineffective or even counterproductive. The value of impact evaluations lies in their ability to generate causal evidence. By isolating the effects of a specific intervention from other confounding factors, these evaluations allow us to determine what works, for whom, and under what conditions. This is particularly important in a region marked by resource constraints and growing demands for transparency. Reliable impact evaluations inform better policy decisions; help allocate resources more efficiently and enhance the legitimacy of development institutions. However, producing robust evaluations requires that operations be designed with evaluation in mind from the outset. This involves identifying a clear theory of change, selecting measurable outcomes, considering appropriate comparison groups, and planning for data collection. Designing for evaluation also means being pragmatic—choosing interventions that allow for sufficient variation or randomization and aligning timelines with evaluation needs.

FINANCIAL INFORMATION

Financing Type	Fund	Amount in US\$
TCN - Nonreimbursable	W2F - OC SDP Window 2 - Economic Growth	200,000
Total IDB Financing		200,000
Counterpart Financing		0
Total Project Budget		200,000
Donors:	N/A	
Disbursement Period:	36 months	
Execution Period:	30 months	

ADDITIONAL FINANCIAL INFORMATION

N/A

I. JUSTIFICATION AND OBJECTIVE

- I.1 To improve lives, the Inter-American Development Bank (IDB) has committed to increasing the impact and scale of our results across Latin America and the Caribbean. Achieving this goal requires a rigorous and sustained effort to evaluate what we do. Impact evaluations are not simply a technical exercise—they are essential for understanding whether our interventions actually lead to better outcomes for people, and how. Without credible evidence of impact, we risk investing in policies or programs that are ineffective or even counterproductive.
- I.2 The value of impact evaluations lies in their ability to generate causal evidence. By isolating the effects of a specific intervention from other confounding factors, these evaluations allow us to determine what works, for whom, and under what conditions. This is particularly important in a region marked by resource constraints and growing demands for transparency. Reliable impact evaluations inform better policy decisions, help allocate resources more efficiently, and enhance the legitimacy of development institutions.
- I.3 However, producing robust evaluations requires that operations be designed with evaluation in mind from the outset. This involves identifying a clear theory of change, selecting measurable outcomes, considering appropriate comparison groups, and planning for data collection. Designing for evaluation also means being pragmatic—choosing interventions that allow for sufficient variation or randomization and aligning timelines with evaluation needs.
- I.4 Despite best efforts, not all operations are evaluable. Some projects are not suitable for impact evaluations due to their scale, objectives, or political and ethical constraints. Even where operations were designed to be evaluated, changes during implementation—such as protocol deviations, delays, or lack of a control group—may weaken causal inference. These realities often leave teams unable to test promising interventions before scaling or recommending them.
- I.5 One way to address these challenges is through Synthetic Randomized Controlled Trials (Synthetic RCTs). This emerging approach uses large language models (LLMs), such as GPT, to simulate the outcomes of randomized experiments. Before committing resources to costly and time-consuming field trials, researchers can construct synthetic treatment and control groups using AI-generated responses. These simulated respondents are exposed to interventions—such as policy messages or behavioral nudges—within a model environment, allowing for analysis of treatment effects using standard experimental techniques.
- I.6 This method builds on the statistical patterns embedded in LLMs, trained on extensive text corpora representing real human behavior and attitudes. Researchers such as Ray Duch have shown that synthetic RCTs can replicate classic findings in behavioral economics and political science, making them a valuable tool for pre-testing designs or vetting ideas before field deployment.
- I.7 The objective of this project is to build an internal capacity that will support IDB staff in designing and interpreting synthetic RCTs and allowing operational and knowledge teams to use synthetic RCTs to refine surveys instruments, test

- hypotheses, and prioritize interventions for full-scale evaluation. Better internal capacity and the support of the synthetic agents will lead to better project design that will increase effectiveness and efficiency of IDB operations in IDB member countries, such as Argentina, Chile, Colombia, and Uruguay and improve lives in the region.
- I.8 This project has been selected for the Development Effectiveness Intelligence Fund (DEIF) – which aims to support initiatives that generate knowledge or contribute to the reduction or strategic knowledge gaps in our understanding of Development Effectiveness – under a rigorous competitive process involving VPS, VPC, and SPD. The DEIF initiative supports projects that generate knowledge and address strategic gaps in development effectiveness through impact evaluations, analysis of execution factors, and effective enhancement initiatives. Following the OC-SDP framework, this initiative is aligned to GN-2819-14 Window 2, under the priority area Inclusive Economic Growth. This Technical Cooperation (TC) aligns with Window 2 Economic Growth priority area as it will address the challenges of social programs by building institutional capacity in the designing and interpreting synthetic RCTs that will allow operational and knowledge teams to use synthetic RCTs to refine survey instruments, test hypotheses, and prioritize interventions for full-scale evaluation.
- I.9 This TC also aligns with the "One Caribbean" initiative, "America en el Centro" initiative, the "Amazonia" initiative and all the other IDB initiatives by creating a Synthetic RCT platform that will allow IDB specialists to pre-test their interventions, the loan components, and components of impact evaluations before bringing them to the field. That way, the impact of IDB interventions will be enhanced, helping the institution to reach more efficiently and effectively the goals of the Institutional Strategy and all its initiatives, ensuring sustainable social progress and reduced inequality in the IDB member countries, consistent with regional development goals.

II. COMPONENTS

- 2.1 **Component 1. Designing and deploying of a synthetic RCT platform (\$150,000).**
- 2.2 This component will finance the design and deployment of a Synthetic RCT platform that enables IDB researchers to simulate randomized controlled trials using large language models (LLMs). This tool will allow for rapid pre-testing of interventions, helping to identify the most promising approaches before committing to field implementation.
- 2.3 The main activities include: 1) Design of the synthetic RCT platform and 2) Writing a report (to be published as a working paper) with the main instructions and lessons learned from the design of the platform.
- 2.4 Component 1 will finance a consultancy firm that will support the project team in the design and deployment of the synthetic RCT platform. The expected timeline for completion of Component 1 is 12 months. The IDB will have the intellectual rights of all products of this TC.
- 2.5 The expected outputs are: 1) discussion paper, and 2) working paper.
- 2.6 **Expected Results.** Designing and deploying the synthetic RCT platform and publishing a paper with the main instructions and lessons learned.
- 2.7 **Component 2: Validation of the platform (\$25,000).**
- 2.8 This component will validate the platform's outputs by replicating well-established experimental findings in development, behavioral economics, and political science. This step will ensure the reliability and credibility of synthetic RCTs for internal use at the IDB.
- 2.9 The main activities include validating the results from the platform by comparing the results of the synthetic RCTs and previous RCTs.
- 2.10 Component 2 will finance a consultancy firm that will support the project team in the validation of the synthetic RCT platform. The expected timeline for component 2 is 12 months, following the completion of component 1.
- 2.11 The expected outputs are: 1) blog post, and 2) research insight.
- 2.12 **Expected Results.** Validating the synthetic RCT platform and disseminating the results and the tool.
- 2.13 **Component 3. Internal dissemination of the platform (\$25,000).**

- 2.14 This component will build internal capacity through training sessions and documentation that support staff in designing and interpreting synthetic RCTs and allowing operational and knowledge teams to use synthetic RCTs to refine survey instruments, test hypotheses, and prioritize interventions for full-scale evaluation.
- 2.15 The main activities include: 1) publishing a blog post on the RES blog Ideas Matter, 2) In addition, we will disseminate the results and the tool through meetings with IDB specialists.
- 2.16 This component will finance the costs of dissemination, which could include brown bag lunches, venue reservations, etc. The expected timeline for component 3 is 6 months following completion of component 2.
- 2.17 The expected outputs are: 1) Blog post and 2) Research Insight.
- 2.18 **Expected Results.** Disseminating the results and policy implications to target policymakers, and dissemination of knowledge generated. The overall expected results of this TC include improved lives in the region through more efficient and effective public policies and IDB projects in the IDB member countries. The design and deployment of the synthetic RCT platform will provide IDB specialists with the tools to pre-test policy and project design as well as impact evaluations. This platform will become an essential tool in the IDB work cycle and the relationship with our member countries.
- 2.19 **Beneficiaries.** This TC will benefit numerous stakeholders. First, the results of the project will directly inform the design of IDB operations in the IDB member countries such as Argentina, Chile, Colombia, and Uruguay, and public policies. Second, it will provide valuable lessons for policymakers in the region about how to design better operations. Third, it will inform the evaluation of operations and public policies, feeding into the loop of development effectiveness.

III. BUDGET

- 3.1 The funds that will provide financing to this TC is Window 2 under the Inclusive Economic Growth Priority Area (W2F) for the total of US\$200,000.00.

Budget in US\$			
Components	W2F	Local	Total
Component 1: Design and deployment of a Synthetic RCT platform	150,000	0	150,000
Component 2: Validation of the platform	25,000	0	25,000
Component 3: Internal dissemination of the platform	25,000	0	25,000

Direct operating expenses	0	0	0
Final evaluation and audit	0	0	0
Total	200,000	0	200,000

IV. EXECUTION STRUCTURE

- 4.1 This TC will be executed by the Inter-American Development Bank through the Research Department (RES/RES) based on a request by the beneficiaries, in accordance with the Bank's Technical Cooperation Policy (GN-2470-2) and the Procedures for the Processing of Technical Cooperation Operations and Related Matters (OP-619-4). The research has complementarities with Bank operations and research. RES/RES will be responsible for the preparation, execution, and supervision of Components 1, 2 and 3 following the policies established by the Bank. The team leader will be responsible for the execution and monitoring of the operation in collaboration with the alternate team leader and the rest of the team members. The team leader will directly supervise consultants and track that deliverables are completed according to the project's planned timeline with the support of the team members of the project.
- 4.2 The principal reason for this execution structure is that the IDB has expertise, capacity, and experience in similar research projects. A second reason is that the Bank has the capacity and channels to disseminate the knowledge generated by this TC to target audiences in the beneficiary countries and in other countries in the region.
- 4.3 There will be no supervision by the IDB's Country Offices for the execution of the TC. However, the IDB Country Offices of Argentina, Chile, Colombia, and Uruguay will be advised of the results and policy implications. They will be informed and consulted about missions and the main activities related to the implementation of the TC.
- 4.4 **Procurement.** All procurement to be executed under this TC have been included in the Procurement Plan (Annex IV) and will be hired in compliance with the applicable Bank policies and regulations as follows: (a) Hiring of individual consultants, as established in the regulation on Complementary Workforce (AM-650) and (b) Contracting of services provided by consulting firms in accordance with the Corporate procurement Policy (GN-2303-33) and its Guidelines. All procurement to be executed under this Technical Cooperation have been included in the Procurement Plan (Annex IV) and will be executed in accordance with the applicable policies and procedures for the Policies for the Procurement of Goods and Works Financed by the IDB (GN-2349-15) and the Policies for the Selection and Contracting of Consultants Financed by the IDB (GN-2350-15).
- 4.5 Responsibilities for supervision and monitoring this operation will fall on RES/RES. The UDR for this TC will be RES/RES. The Team Leader should

approve payments, send yearly reports and supervise the deliverables from firms and consultants.

- 4.6 **Intellectual Property.** Knowledge products generated from Bank-executed activities within this TC will be property of the Bank and may be made available to the public under a Creative Commons license. However, at the request of the beneficiaries, following the provisions of AM-331, the intellectual property of said products may also be licensed through specific contractual commitments that shall be prepared with the advice of the Legal Department and disseminated in accordance with AM-325. If the Bank receives, manages, or uses information that may contain personal data or sensitive information, the Bank's Personal Data Privacy Policy (GN-3030) will be followed in coordination with the IDB Privacy Team.
- 4.7 **Execution and Disbursement Period.** The execution period is thirty-six (36) months from the day of the approval of the TC eligibility, and the disbursement period is thirty (30) months.
- 4.8 **Monitoring, Reporting, and Supervision.** The UDR for this TC will be RES/RES. The Team Leader is Carlos Scartascini (RES/RES) and the Alternate Team Leader is Samuel Berlinski (RES/RES) who are responsible for the monitoring, reporting and supervision of the TC.

V. POTENTIAL RISKS

- 5.1 The main risk to successful and timely execution of this TC is Limited Validity of Synthetic Results. The primary concern with using Synthetic RCTs is whether the responses generated by LLMs accurately reflect real human behavior and decision-making. If not properly validated, this could lead to misleading conclusions or poor policy design. To minimize that risk, we will conduct validation exercises comparing synthetic results to real-world RCTs and survey experiments. We will replicate well-established findings to build trust in the method and will document limitations clearly and position synthetic RCTs as pre-testing and exploratory tool.
- 5.2 Another risk is resistance to adoption. Staff may be unfamiliar with synthetic RCTs or skeptical about their usefulness, which could limit the uptake across departments. To minimize this risk, we will provide hands-on training, use cases, and technical support to build user confidence. We will also showcase early success stories through pilots in coordination with operational teams and will position the platform as a complement to existing evaluation tools rather than a replacement.
- 5.3 A final risk is the rapid evolution of AI tools. The fast-changing nature of AI technology could make the chosen models or platforms obsolete within a short period. To minimize the risk, we will use flexible, modular architecture that allows integration with different LLM providers (e.g., open-source and commercial models).

- 5.4 It is suggested that the project incorporate early pilot applications in collaboration with operational teams, identify departmental champions to foster uptake, and allocate resources for ongoing platform updates given the rapid evolution of AI tools. These measures would help secure institutional buy-in and ensure that the platform is not a one-off experiment but a sustainable component of Banks evaluation toolkit.

VI. EXCEPTIONS TO BANK POLICIES

- 6.1 This TC does not involve exceptions to Bank policy.

VII. ENVIRONMENTAL AND SOCIAL ASPECTS

- 7.1 This Technical Cooperation is not intended to finance pre-feasibility or feasibility studies of specific investment projects or environmental and social studies associated with them; therefore, this TC does not have applicable requirements of the Bank's Environmental and Social Policy Framework (ESPF).

REQUIRED ANNEXES:

- Annex I: Results Matrix
 - Annex II: Terms of Reference
 - Annex III: Procurement Plan
-