Melbourne Institute of
Applied Economic and Social Research

Working Paper Series

Breaking Down Barriers Assistant: Leveraging AI to Conduct Policy Analyses with Complex Data

Ujjwal KC
Jan Kabatek

Working Paper 04/26

THE UNIVERSITY OF MELBOURNE
MELBOURNE INSTITUTE

Breaking Down Barriers Assistant: Leveraging AI to Conduct Policy Analyses with Complex Data

Ujjwal KC and Jan Kabatek
Melbourne Institute of Applied Economic and Social Research
The University of Melbourne

Melbourne Institute Working Paper No. 04/26
April 2026

We are grateful to the members of the Breaking Down Barriers Project and the AI Working Group within the Melbourne Institute for their continued support. This research was supported by grants from the Paul Ramsay Foundation under the Breaking Down Barriers Project. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Melbourne Institute of
Applied Economic and Social Research
The University of Melbourne
Victoria 3010 Australia
T +61 3 8344 2100
F +61 3 8344 2111
E melb-inst@unimelb.edu.au
W melbourneinstitute.unimelb.edu.au

Abstract

Evidence-based policy design increasingly relies on integrated administrative data, yet access to and effective use of these data remain constrained by technical and institutional barriers. This paper introduces the Breaking Down Barriers Assistant, an AI-powered analytical platform designed to democratise access to complex, linked administrative datasets for policy analysis. The system enables users to query secure, pre-aggregated Australian administrative and survey data using natural language and supports the full policy research cycle through three integrated tools: variable discovery, visual analytics, and automated reporting. Methodologically, the BDB Assistant combines Retrieval-Augmented Generation with controlled code generation and execution, ensuring that all analytical outputs are grounded in verified metadata, auditable Python code, and human-in-the-loop validation. Using benchmark comparisons with established platforms such as YouthView and the BDB Community Profiles, we demonstrate that the Assistant can accurately reproduce complex spatial and longitudinal policy analyses that traditionally require advanced technical expertise. The findings show that the system substantially reduces time-to-insight while maintaining strict standards of accuracy, transparency, and privacy. The BDB Assistant illustrates how responsibly governed generative AI can expand analytical capacity within government and the social sector, supporting more timely, rigorous, and locally targeted evidence-based policymaking.

JEL classification:

Keywords: Evidence-based policymaking, Artificial Intelligence, Generative AI, Large Language Models, AI-Assistants

1. Introduction

Reducing the extent of social problems such as deep-rooted intergenerational disadvantage is a primary goal for any modern society. Whether it is chronic unemployment, poor health, or low educational attainment, these issues are rarely isolated. Instead, they form a vicious circle of disadvantage, where interconnected problems are passed down through generations. To break this cycle, we need a shift toward Evidence-Based Policy (EBP). This means making decisions based on data that can track how different life factors, such as housing, education, and jobs, interact over time and respond to government policy (Eliadis et al. 2005; Howlett 2009).

The data needed for these applications already exist. Many nations now link robust statistical information sources, such as census records and employment statistics, to create holistic data snapshots of their communities. However, a major problem remains: the people who need these insights most, such as policymakers and community leaders, often do not know how to use these data (Kelman et al. 2002; Platt et al. 2020). These integrated datasets are massive, varied, and technically overwhelming (Cole et al. 2020). The corresponding codebooks are often filled with technical language that obfuscate underlying concepts[1] and even professional analysts can find these records indecipherable (Kelman et al. 2002). This creates a "bottleneck" where high-value evidence is effectively locked away because users lack the specialized coding skills in languages like Python, Stata, or R required to access it. This gap leads to high costs and long delays, preventing evidence from being used where it is needed the most (Howlett 2009).

In Australia, this challenge is particularly pressing. While the country has seen overall declines in poverty, more than 40 percent of Australian communities still face poverty rates higher than 12 percent (Payne and Samarage 2020). To alleviate this issue, the Melbourne Institute launched the Breaking Down Barriers (BDB) project. Its goal is to move beyond broad national averages and deeply understand the specific community and geographic factors that keep people in poverty (Ananyev et al. 2023). To achieve this, the project created the BDB Shared Data Environment (BDB-SDE), a secure, integrated data asset (Ananyev et al. 2025). This environment brings together high-value administrative data from a multitude of administrative and survey sources. This environment contains datasets on the many interconnected factors that

[1] For example, a simple concept like "home ownership" might be hidden under a cryptic variable name like home_own_mort.

contribute to disadvantage. For instance, the datasets can help identify the specific factors influencing high rates of "NEET" youth — those not in employment, education, or training in particular regions (Cole et al. 2020; Ujjwal et al. 2025). However, despite this immense potential, getting quick insights from the BDB-SDE still requires a high level of technical expertise. This gap leads to high costs and long delays, preventing evidence from being used when it is needed most.

Our solution to this problem is the BDB Assistant. This AI-powered application acts as a "knowledge translator" between complex data and the people who need to use it. The core innovation is a Retrieval-Augmented Generation (RAG) framework (Lewis et al. 2020). Instead of letting the AI guess an answer, RAG forces the system to look up specific and verified facts from the data's own documentation and metadata. This prevents the AI from making mistakes (termed "hallucinations"), which is essential in high-stakes government policy (Taeihagh 2021).

The assistant provides three integrated tools to simplify the policy research cycle:

1. Variable Information Tool: Provides instant, accurate definitions and context for any variable in the dataset using natural language queries.
2. Visual Analytics Tool: Identifies the right variables to answer a query. It lets the user guide the process by choosing chart types and levels of detail, then automatically merges the data to create interactive maps or charts.
3. Analyst Tool: Acts as a research partner for complex questions. It identifies the necessary data, writes the Python code needed for the analysis, and explains the statistical results in simple terms

This paper details the theoretical foundation, implementation, and governance of the BDB Assistant. The system leverages a specialized architecture that integrates Retrieval-Augmented Generation (RAG) with automated code execution to transform weeks of manual data exploration into minutes of interactive dialogue. This integration ensures strict factual grounding by anchoring every analytical claim in verified metadata and source code (Lewis et al. 2020; Zhao et al. 2023). Essential ethical safeguards reinforce the platform, specifically through a human-in-the-loop protocol and a policy of complete transparency that allows users to audit the exact logic and code used for every result. These mechanisms ensure that AI serves as a responsible instrument for developing effective social policy (Cole et al. 2020; Huang et al. 2025; Taeihagh 2021).

The remainder of this paper is organized to provide a comprehensive overview of the system and its implications. Section 2 establishes the theoretical underpinnings of integrated data and Retrieval-Augmented Generation. Section 3 details the technical architecture of the BDB Assistant and its specialized tools. Section 4 examines the policy impact and the ethical governance framework supporting the system. Finally, Section 5 offers concluding remarks and outlines directions for future work.

2. Literature Review and Theoretical Framework

2.1. The Academic Justification for Integrated Data in Policy Analysis

Major social policy issues, such as the cycle of persistent intergenerational disadvantage, are rarely driven by a single factor in isolation. Instead, they are fueled by a web of interconnected challenges, including chronic unemployment, health disparities, and low educational attainment (Connelly et al. 2016; Leigh 2025). Modern EBP frameworks rely on understanding these complex social structures, so that well-informed policymakers could move away from ad-hoc governance practices toward data-driven decisions and interventions.

However, this understanding is far from readily available. Historically, policy analysts and researchers have drawn insights from two data primary sources, each with significant limitations. Survey data offers rich personal detail, but it is frequently compromised by issues such as declining participation rates or difficulty remembering specific past events (Australian Bureau of Statistics 2025). Conversely, administrative data, including tax, health, and social service records, is vast and highly accurate, yet it often lacks vital context (Connelly et al. 2016; Lucke and Frank 2025; Schmeling et al. 2025). A tax record, for instance, might show a significant drop in income but cannot explain the household dynamics or local economic shifts that led to that outcome. Using these sources in isolation provides only a fragmented view of a person's life.

Integrated datasets provide the academic solution to this fragmentation. By definition, integrated datasets are the technical and probabilistic linking of diverse administrative and survey records to create a unified, longitudinal profile of an individual or cohort across different institutional touchpoints. This method links family census records with labor market outcomes and education history to create a holistic view of an individual's life course (Ujjwal et al. 2025). This integrated approach enhances analytical efficiency in EBP by allowing analysts to move beyond simple correlations to identify causal pathways.

2.2. The Policy Design Challenge: Breaking Down Silos

However, the existence of high-quality data does not automatically lead to effective policy. A second hurdle arises from entrenched organizational silos. Historically, public institutions have suffered from fragmentation, with complex and interrelated social problems being treated as isolated issues rather than as parts of a larger system (Eliadis et al. 2005). This often leads to instrument conflict, where a policy designed to achieve one goal accidentally undermines another. For instance, an education program meant to encourage study might be weakened by a social security rule that penalizes a student's part-time earnings (Howlett 2009).

The BDB Shared Data Environment (BDB-SDE) represents a strategic step toward mitigating this structural fragmentation by facilitating the technical convergence of evidence. By linking diverse datasets, such as the Census, the National Evidence Register and Outcomes (NERO), and the Internet Vacancy Index (IVI), the environment provides the foundational evidence required for a strategic shift in policy design (Leigh 2025). This integration allows users to analyze the intersections between policy domains, such as the link between regional labor demand and educational attainment, rather than viewing datasets in isolation.

However, a final barrier to effective EBP remains. Even within a shared data environment, obtaining quick insights still requires a high level of technical expertise in coding and data science. This creates a significant operational bottleneck, as policymakers must wait weeks for specialists to translate their questions into technical queries. This delay often prevents critical evidence from being used, leaving a gap between the availability of data and the actual delivery of policy solutions.

2.3. Large Language Models in EBP

The technical bottleneck to utilizing the potential of integrated data for decision-making can be significantly enhanced by the capabilities of large language models (LLMs). Built on the Transformer architecture and trained on extensive corpora, these models excel at interpreting human language and generating sophisticated responses. In the context of evidence-based policy, LLMs offer a computational aid for automating the synthesis of complex documentation and supporting data analysis by translating natural language into executable code (Lewis et al. 2020). This capacity is essential for unlocking holistic and cross-domain insights from the datasets within the BDB-SDE.

However, the deployment of generic LLMs in high-stakes policy environments presents significant challenges that preclude their unsupervised use. Because these models are designed

for probabilistic token prediction rather than deterministic database querying, they are prone to generating "hallucinations". Hallucinations are situations when LLM outputs are syntactically plausible but factually incorrect. Such errors are unacceptable in social policy, where decisions impact citizens and involve substantial public funding (Taeihagh 2021; Huang et al. 2025). Furthermore, many advanced models operate as opaque systems; if a policy analyst cannot inspect the underlying logic or the specific source of a result, they cannot validate the evidence or ensure procedural fairness. General AI also lacks the specialized knowledge needed to interpret unfamiliar data labels and complex rules inherent in administrative datasets. Without an intrinsic understanding of specific variable definitions or methodological caveats, the model remains at risk of producing flawed analysis (Huang et al. 2025; Zhao et al. 2024).

Furthermore, a significant logical risk exists in assuming that an LLM's neutral tone equates to unbiased reasoning; the model's internal logic may still reflect biases present in its training data or the structural inequities of the administrative records it interprets. The safe application of AI to address systemic intergenerational disadvantage, therefore, requires a specialized solution that preserves linguistic power while upholding strict factual accuracy and transparency.

2.4. Retrieval-Augmented Generation (RAG) as a Trust and Accuracy Solution

The RAG framework serves as our primary defense against the factual errors and logical gaps found in standard AI. At its core, RAG is a technical architecture that provides a large language model with a verified external knowledge base to reference during the generation process. Instead of relying solely on the probabilistic patterns acquired during its initial training, the model is granted access to a specific, curated library of documentation to factually ground the generative output in administrative reality (Lewis et al. 2020; Huang et al. 2025). By decoupling this validated knowledge retrieval from the linguistic processing of the model, RAG ensures that the BDB Assistant remains a high-fidelity tool for rigorous policy analysis.

The process begins with Indexing, where all BDB data manuals and research papers are converted into a searchable vector database. When a user asks a question, the system performs a Retrieval step to find the exact sections of the documentation that hold the answer. Subsequently, in the Generation phase, these specific document chunks are provided to the AI as its only source of truth. The system is strictly instructed to answer using only the provided context, effectively tethering the AI to the factual information.

This architecture is fundamental to the system's design because it ensures that every analytical output is both accurate and verifiable. By forcing the model to operate within the formal boundaries of the BDB-SDE datasets documentation, the system prevents the fabrication of variable definitions or the invention of non-existent analytical procedures. For a policy analyst, this transition from a "black-box" generative process to a transparent and evidence-linked workflow provides a computational aid that maintains the academic rigor required for EBP. Furthermore, this process occurs within an isolated execution environment, ensuring that while the model generates logic based on metadata, it never gains direct access to protected unit-record data.

2.5. The Democratization and Capacity-Building Role of NLQ Systems

The BDB Assistant operates as a Natural Language Query (NLQ) system, serving as a computational interface that mitigates the technical barriers to entry in data-driven policy. Traditionally, the utility of integrated data has been constrained by a significant specialist or technology bottleneck. Accessing these resources required analysts to master the complex syntax of programming languages such as R, Python, or Stata, while simultaneously navigating intricate database schemas. This cognitive load created a substantial barrier to entry, ensuring that only a localized group of technical specialists could derive insights from the data (Howlett 2009; Schmeling et al. 2025). The introduction of an interface that allows users to interact with data through plain-language queries removes this obstacle, broadening institutional accessibility to EBP and increasing the efficiency of the research-to-action cycle (Lucke and Frank 2025).

The role of the tool as an embedded computational aid effectively replicates the procedural methodology of data scientists within an accessible interface. This transition from specialized manual coding to conversational interaction achieves more than mere temporal efficiency; the system builds the analytical capacity of the target audience, including policymakers and non-profit leaders. The acceleration of the "time-to-insight" allows decision-makers to prioritize the mitigation of systemic intergenerational disadvantage over the technicalities of software development. Ultimately, this architecture ensures that community leaders possess the same high-level analytical capacity as centralized data units, fostering a more inclusive and responsive evidence ecosystem.

3. The BDB Assistant: Architecture and Functionality

In this section, we detail the technical architecture and core functionalities of the BDB Assistant.

3.1. Data Foundation: The BDB Integrated Data Environment

The core utility of the BDB Assistant rests upon the BDB-SDE, a securely curated collection of linked administrative and survey datasets (Ananyev et al. 2025). This environment is structured to address systemic intergenerational disadvantage by providing longitudinal, high-resolution insights into demographics, education, and labor market outcomes of Australian communities. By linking regional data, such as Census records (DS016) and regional employment demand from the National Evidence Register and Outcomes (NERO) and the Internet Vacancy Index (DS007), the BDB-SDE allows researchers to analyze the longitudinal interactions between these factors to identify the drivers of community poverty or economic mobility (Ujjwal et al. 2025). All datasets currently integrated within the BDB-SDE are categorized in Table 1.

Table 1: Summary of BDB-SDE Datasets and Core Focus

Dataset Name	**Source/Composition**	**Core Focus**
Geographical Small-Area Measures (DS016)	ABS 100% Census data and ABS Person-Level Integrated Data Asset (PLIDA).	Poverty measures and various household demographics for place-based analysis (2011–2021).
Post-Secondary Data (DS011)	Higher Education Information Management System (HEIMS) from the Department of Education.	Institutional-level data on student enrolments, completion rates, and student composition (2001–2020).
Labour Market Data (DS007)	NERO, IVI, and DSS Jobseeker and YA benefits statistics from Job Skills Australia and Department of Social Services.	Regional measures of employment, occupational demand, and welfare benefits.

Youth Disadvantage and Transition Measures (DS012)	Measures built from Census data and ABS PLIDA, focusing on younger individuals.	Youth-level poverty and Not in Education, Employment and Training (NEET) rates.
Community Profiles Data (DS015)	ABS Census data and MI Taking the Pulse of the Nation Survey data.	Data used for the BDB Community Profiles at two levels of geographic disaggregation (SA2 and SA3).
Employment Service Provider Data (DS014)	Transition to Work (TTW) data from Department of Employment and Workplace Relations (DEWR).	Tracking of employment outcomes for clients of specific service programs.
COVID-19 Stay-at-home Orders (DS010)	Data derived from published stay-at-home orders.	Local Government Area-level data on the nature and duration of lockdowns (2020–2022).

3.2. Architectural Components

The BDB Assistant architecture consists of interconnected components orchestrated to ensure secure, verifiable, and transparent data analysis. As illustrated in Figure 1, the system is centered around the BDB Backend Orchestrator, which manages the operational flow between the user interface, the RAG Engine, the inference engine (AWS Bedrock), and the isolated data processing environment.

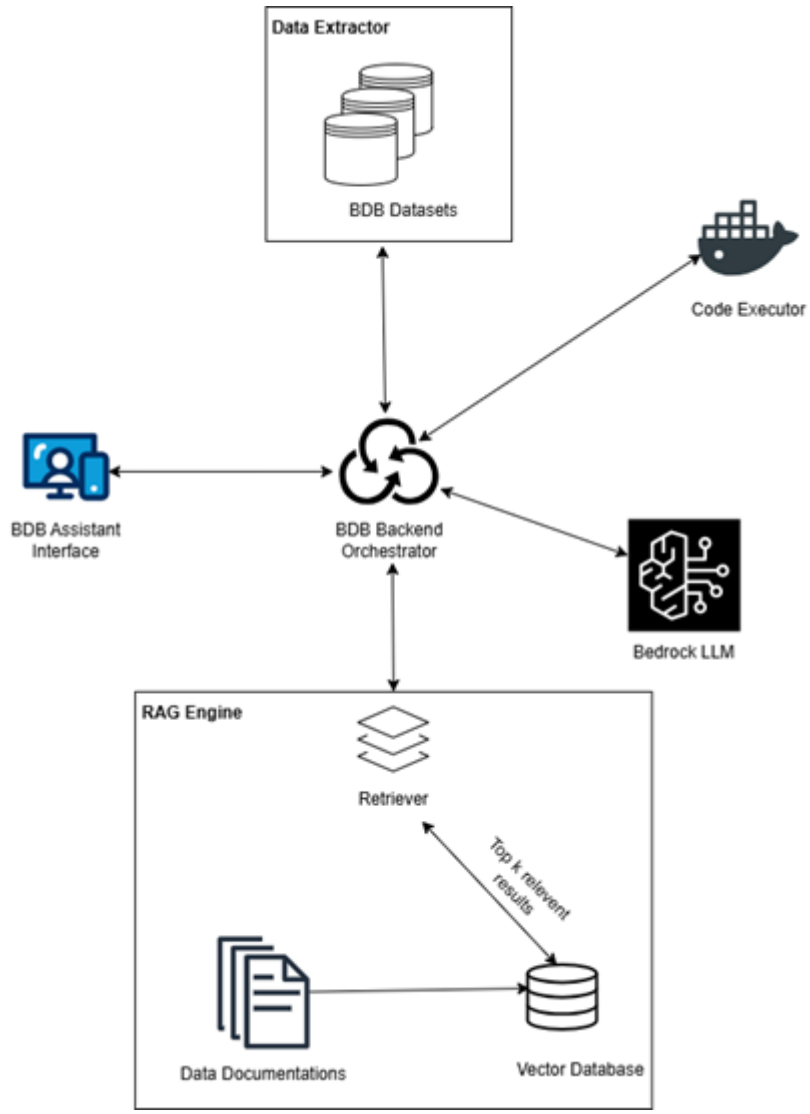

Figure 1: BDB Assistant's overview of components and their interactions

3.2.1 BDB Assistant Interface

This is the user-facing web application where researchers and policymakers submit natural language queries (NLQ). It serves as the primary input and output layer, providing users with structured reports, interactive visualizations, and downloadable analysis files. To maintain expert control, the interface requires the user to review and confirm the retrieved metadata and proposed analytical logic before any code is executed.

3.2.2 BDB Backend Orchestrator

This acts as the central control plane and workflow manager. It is responsible for the end-to-end process: routing queries to the RAG Engine, transmitting retrieved context to the LLM, validating generated code syntax, triggering the Data Extractor, and compiling the final results.

3.2.3 RAG Engine

The RAG Engine is the core knowledge mechanism responsible for generating contextually accurate and grounded responses. This is paramount for responsibly deploying Large Language Models in high-stakes policy domains, as it ensures the model operates strictly within the verified domain knowledge of the BDB-SDE. It operates through three integrated sub-components:

1. Data Documentations (The Knowledge Corpus): This is the definitive source of truth containing all official BDB codebooks, variable definitions, and methodological reports. Documentation undergoes semantically meaningful chunking to ensure that technical details and methodological caveats are preserved as discrete, retrievable units.
2. Hybrid Index (The Vector and Keyword Database): After chunking, each text segment is processed through two parallel indexing streams. First, an embedding model converts segments into high-dimensional vector representations for semantic similarity. Second, a traditional inverted index is created to support precise keyword matching (such as BM25). This hybrid approach ensures the system can capture both broad thematic intent and specific technical indicators.
3. Advanced Retriever (Hybrid Search and Ranking): The Retriever executes a multi-stage search, performing semantic similarity comparisons alongside keyword filters. Candidates are ranked to provide a precise list of variables and metadata for mandatory human validation, ensuring the subsequent generation phase is anchored to verified data assets.

3.2.4 Bedrock LLM

This component (utilizing models such as Anthropic Claude via AWS Bedrock[2]) is responsible for translation and synthesis. It receives the user query augmented by the RAG Engine's context and is strictly instructed to generate Python code or synthesize findings based exclusively on the provided documentation. This prevents the model from relying on its internal probabilistic training data for factual claims.

[2] Other AI models such as ChatGPT, and Gemini were also used for experimentation

3.2.5 Data Extractor

This secure module manages programmatic access to the underlying BDB datasets (e.g., DS007, DS016). Its role is to retrieve variable information and prepare an integrated dataset required for the analysis. This ensures that only the minimal necessary data, filtered and merged according to the valid analytical code, is provided for processing.

3.2.6 Code Executor

This is a secure, isolated, and containerized computing environment using Docker. It contains all necessary libraries to safely execute the generated Python code against the extracted data. This data-masking middleware architecture ensures that the LLM never gains direct, unmediated access to raw unit-record data. The executor returns only statistical logs, aggregate tables, and charts to the Orchestrator, upholding rigorous privacy and governance standards.

3.3. Specialized Tools and Operational Flow

The BDB Assistant architecture supports three specialized tools, each engineered to serve as a computational aid at a specific stage of the policy research cycle, ranging from initial data discovery to comprehensive statistical reporting.

3.3.1 Variable Information Tool

The Variable Information Tool functions as a comprehensive guide to the BDB-SDE data universe. Its primary purpose is to broaden access to technical documentation by providing precise answers regarding variable definitions, measurement units, and spatial or temporal availability. By allowing users to submit natural language queries, such as inquiring about specific measures of unemployment or the range of available poverty data, the tool significantly reduces the high barrier to entry associated with complex integrated datasets.

The workflow sequence for the Variable Information tool is as follows.

1. Semantic Search: The user's query is converted into a vector embedding, and the RAG engine executes a similarity search to retrieve the most relevant documentation segments.
2. Contextual Response: The Bedrock LLM utilizes these retrieved segments to generate detailed and accurate information, often organizing variables into thematic categories for better clarity.

3.3.2 Visual Analytics Tool

The Visual Analytics Tool facilitates interactive data exploration by translating complex datasets into comprehensible visual insights. This tool accelerates the initial stages of policy analysis, allowing users to rapidly identify patterns and trends across time or geography. For instance, a user can request a trend analysis of youth not in employment, education, or training for a specific region over a ten-year period.

The workflow sequence for the Visual Analytics Tool is as follows.

1. Identification: The user submits a visualization query, which the RAG Engine uses to identify the necessary variables within the data environment.
2. Human-in-the-Loop Selection: The tool presents the identified variables to the user, who then selects the specific items of interest and defines parameters such as chart type and level of aggregation.
3. Data Integration: The Data Extractor dynamically merges and filters the underlying records to prepare a customized dataset.
4. Rendering: The BDB Assistant Interface generates interactive charts or maps based on the user's specific technical requirements.

3.3.3 Analyst Tool

The Analyst Tool serves as an automated research partner, converting open-ended queries into rigorous statistical reports. It supports high-level inference and multi-step analyses, enabling users to receive interpreted results without requiring advanced programming expertise. The tool produces audit-ready documentation that includes both synthesized findings and the underlying analytical code.

The workflow sequence for the Analyst Tool is as follows.

1. Knowledge Retrieval: Upon receiving a complex query, the RAG Engine identifies the relevant knowledge segments and the most suitable variables.
2. Validation: Through a human-in-the-loop protocol, the user validates the suggested variables before the analysis proceeds.
3. Secure Preparation: The Data Extractor prepares the integrated dataset required for the specific statistical task.
4. Code Generation and Execution: The LLM generates bespoke Python code based strictly on the retrieved documentation. This code is executed within the isolated Code Executor environment.

5. Interpretation and Reporting: The Orchestrator collects the resulting statistical logs and charts. The Bedrock LLM then interprets these aggregated results, without ever accessing raw individual data records, to produce a context-rich report. The final output includes the automated code and the integrated data used, ensuring transparency and allowing for independent validation.

4. Policy Impact, Ethical Governance, and Validation

4.1. Impact on the Evidence-Based Policy Cycle

The architecture of the BDB Assistant intervenes at critical stages of the policy cycle, accelerating the transition from data awareness to informed action (Sempe et al., 2025; Treasury Evaluation, 2025). By providing a seamless interface between natural language and integrated datasets, the system serves as a computational aid for analysts (Verian Group, 2025).

4.1.1 Agenda Setting and Problem Definition

Policymaking requires a precise scoping of issues to reveal geographic or demographic concentrations of need (Column Content, 2025). Using the Variable Information and Visual Analytics tools, advisors can quickly define the scale of regional distress. For example, a user seeking to address youth vulnerability can instantly query the geographic distribution of poverty and overlay this with the regional incidence of youth not in employment, education, or training. This ability to segment populations by age and cross-reference them through disadvantage indicators identifies whether a problem is a localized phenomenon or a systemic (and potentially inter-generational) trend.

4.1.2 Policy Formulation and Option Generation

Evaluating intervention options requires moving from problem identification to causal reasoning and forecasting. The Analyst Tool facilitates this by allowing for the rapid generation of disaggregated analyses and policy scenarios. For example, users can compare the labor market dynamics corresponding to specific occupational sectors, filtering by income and skill levels. With this level of disaggregation, a user can evaluate whether regional labor shortages result from an overall lack of job candidates or a skills mismatch between the available job candidates and the offered job postings.

4.1.3 Monitoring, Evaluation, and Learning

Effective governance requires continuous tracking to ensure policies remain responsive (Department of Education, 2025). The Analyst Tool's ability to visualize time-series data allows for longitudinal performance tracking. Analysts can monitor jobseeker or youth allowance payments from 2016 to 2024 to detect policy drift or program failure early. Because the assistant provides the exact Python code used for these aggregations, the process remains transparent to auditors, fostering a culture of learning where policymakers understand precisely how metrics and statistics were derived.

4.2. Case Studies: Validation and Application

The validation of the BDB Assistant is based on its ability to accurately replicate complex findings established by human-led research platforms, such as YouthView and the BDB Community Profiles. By using these platforms as an empirical benchmark, we demonstrate that the RAG-to-code pipeline achieves technical parity with traditional manual analysis. Validation is thus confirmed when the Assistant's automated outputs align perfectly with the established ground truth of the BDB-SDE data assets in these platforms.

4.2.1 Case Study: Labor Demand-Supply Mismatch

The identification of job deserts versus skills shortages requires integrating demand-side vacancy data with supply-side labor force statistics. This provides a rigorous test of the Assistant's ability to handle derived ratios and multi-dataset joins.

- Policy Question: Which SA4 regions have the highest ratio of vacancies to unemployed people for the 2022 period?
- BDB Assistant Action and Validation: Upon receiving this natural language query, the RAG-to-code engine retrieves the specific metadata for the derived ratio (ratio_vac_unempld) and the regional identifiers (sa4_name). The system identifies the intersection of available temporal data and generates the Python code required to filter the 2022 dataset, rank the regions, and render a comparative bar chart. The output was validated by comparing the regions against the analyses from YouthView Dashboard.
- Significance: The system replicates a complex labor market analysis, which historically required manual data merging and rigorous schema validation, within minutes. By providing the exact formula and the underlying Python code

used for the ranking, the Assistant ensures the resulting list of high-demand regions is factually traceable.

4.2.2 Case Study: Spatial Persistence of Disadvantage

Community-level poverty rates are often geographically entrenched, requiring hyper-localized longitudinal analysis to identify where disadvantage is persisting (Payne and Samarage 2020). This case tests the Assistant's geospatial and temporal calculation accuracy.

- Policy Question: Map the SA2 areas in Sydney that have experienced a decrease in the number of individuals observed in poverty between 2016 and 2021.
- BDB Assistant Action and Validation: The RAG component retrieves metadata for poverty counts and SA2 identifiers across a five-year window. The engine generates code to calculate the difference between the two periods and executes a geospatial mapping script in the isolated execution environment. The outputs were compared against the outputs of the BDB Profiles.
- Significance: This high-resolution spatial insight allows for the hyper-localization of social services. The validation proves the Assistant can reliably move beyond broad metropolitan averages to address specific pockets of systemic intergenerational disadvantage, providing a trustworthy tool for infrastructure and service placement.

4.3. Ethical Governance and Transparency in AI-Aided EBP

The integration of Generative AI into EBP demands more than technical precision; it requires a clear framework that ensures humans remain in control of the results. The BDB Assistant follows specific protocols designed to protect data integrity, prevent AI errors, and ensure the user is the final judge of the truth (Taeihagh 2021). By keeping the human researcher at the center of key decisions during analyses, the system moves away from autonomous decision-making toward a model of collaborative verification.

A primary risk in this field is the potential for automation bias, where users might accept an AI's answer without checking its validity. To prevent this, the BDB Assistant uses a Human-in-the-Loop approach that makes every step of the AI's logic visible to the user (Busuioc 2021; Huang et al. 2025). This transparency is achieved through three progressive layers of verification:

- Result Verification: Every analytical output is accompanied by the exact Python code used for the calculation and the specific data retrieved from the BDB-SDE. This allows the user to act as a gatekeeper, verifying that the AI's interpretation remains strictly aligned with the underlying statistical reality (Cheong 2024).
- Reasoning Disclosure: By requiring the Assistant to explain its internal reasoning, such as why it chose a certain variable, the system ensures the researcher stays actively engaged with the data. This setup ensures the AI handles the high-volume task of data aggregation, while the human researcher retains sole responsibility for the final policy interpretation.
- Methodological Openness: The system uses a strict discovery protocol to show exactly where information comes from. This ensures that the transition from a general concept, such as job insecurity, to a technical measure is easy to follow and check. If the system suggests a specific rate or metric, it is architecturally required to disclose the exact mathematical formula used. This openness ensures that every insight is defensible, can be repeated by others, and meets high evidentiary standards (Ada Lovelace Institute et al. 2021).

4.4. Data Security and Privacy Compliance

To ensure the BDB Assistant meets the stringent security requirements of high-stakes governance, its architecture is predicated on the principle that privacy protection must be an inherent property of the data infrastructure rather than a secondary feature of the AI layer (Lucke and Frank 2025). This is operationalized through a multi-layered security framework that maintains a strict air gap between the generative model and sensitive information.

- Architectural Separation of Data and LLM: The Large Language Model, accessed via secure services like AWS Bedrock, never interacts directly with raw data files. Instead, it operates exclusively within a symbolic reasoning layer, utilizing only vectorized documentation and metadata to generate analytical plans and interpret results. This ensures that the model's generative processes are informed by the structure of the data without ever having exposure to its specific records.
- Pre-Aggregated Data Infrastructure: The entire BDB-SDE environment is built upon datasets that have been pre-aggregated at regional scales, such as

SA2 or SA4 levels. This foundational step ensures strict compliance with Australian Bureau of Statistics (ABS) re-identification guidelines and vetting requirements. Consequently, the Assistant functions as a natural language interface for querying already-vetted summary-level statistics. such as unemployment counts or income averages, rather than individual unit records.

- Sandboxed Execution: All statistical code generated by the LLM is executed within a highly controlled, isolated, and non-persistent environment. This sandbox is specifically pre-configured to prevent any unauthorized data reconstruction or direct access to secure endpoints. By isolating the computation in this manner, the Assistant ensures that individual privacy is maintained throughout the entire analytical workflow, from initial query to final report.

4.5. Mitigation of Analytical and Systematic Bias

The BDB Assistant incorporates specialized safeguards designed to maintain analytical integrity and address the systemic biases inherent in administrative data collection (Connelly et al. 2016). These guardrails are operationalized through a combination of system instructions, constrained prompting, and automated retrieval protocols:

- Mandatory Contextual Retrieval (Data Bias): Administrative datasets frequently suffer from systematic omissions, such as the missing middle—individuals who are eligible for support but do not claim it. To prevent the over-generalization of findings, the system utilizes a hard-coded retrieval instruction. Whenever a user queries a variable, the assistant is programmed to prioritize the retrieval of the methodological caveats and notes fields from the BDB-SDE metadata. By presenting these definitions alongside the statistical output, the system compels the analyst to confront the limitations of the data, such as specific exclusion criteria or historical shifts in measurement.
- Neutrality Protocol (Algorithmic Bias): To prevent Large Language Models from interpreting neutral statistics through sensationalized or deficit-based narratives found in their training data, the assistant is governed by a discursive straitjacket of system-level constraints. These instructions

explicitly forbid the use of emotive adjectives, mandating that insights be reported in frequentist terms, such as a 2 percentage point increase.

- Reference Anchoring (Causal Fallacy): To resolve potential causal fallacies, the assistant is subject to a reference anchoring protocol. Every narrative claim must be anchored to a specific generated statistic. If the retrieved documentation does not provide explicit causal evidence, the system instructions force the model to rephrase the insight as a spatial pattern or a correlation. This architectural constraint ensures the AI remains within the empirical boundaries of the data, preventing speculative claims that could misinform policy decisions.

5. Conclusion and Future Work

5.1. Summary of Achievements and Contributions

The BDB Assistant represents a significant shift in the accessibility of integrated administrative data for evidence-based policy. This tool broadens institutional accessibility to sophisticated spatial and longitudinal analysis by bridging the gap between natural language inquiry and complex data assets through a RAG-to-code architecture. A primary contribution of this work is the development of a high-precision retrieval framework that avoids the common pitfalls of generative models. By utilizing a hybrid search strategy that combines semantic embeddings with exact keyword matching, the system ensures that the discovery of variables is technically grounded in verified metadata rather than interpretative generation.

Central to this achievement is the formalization of a Human-in-the-Loop verification gateway. The assistant functions by design as a high-fidelity query interface that presents its retrieval candidates for expert validation, ensuring that the human researcher remains the final arbiter of truth. This design ensures that every generated Python script is anchored to the precise data identifiers selected by the user. Furthermore, the core architecture demonstrates a robust security posture defined by the architectural isolation of the large language model from data storage and the use of isolated and non-persistent execution sandboxes. This framework provides a scalable model for applying generative AI to integrated data while maintaining the rigorous privacy and accuracy standards essential for governance.

5.2. Limitations and Challenges

The BDB Assistant faces inherent limitations rooted in the nature of administrative data, the economics of cloud-native AI, and the linguistic complexities of social science research, despite the advancements in automated discovery.

1. Computational and Resource Constraints: High-performance infrastructure is currently required for both the vector database and LLM orchestration through services such as AWS Bedrock. The computational overhead of generating high-fidelity embeddings and executing multi-step statistical code remains a significant operational cost lever. Achieving the near-instantaneous latency expected in modern decision-support systems remains a primary optimization goal, although the current execution time, averaging up to a few minutes, represents enhanced efficiency relative to manual analysis.
2. Knowledge Base Boundaries: Retrieval-augmented generation accuracy is strictly bounded by the scope of vectorized documentation. The model cannot autonomously integrate new domains without a rigorous process of collection, chunking, and semantic indexing because the corpus is currently limited to labor, youth disadvantage, and census data. This restriction underscores the role of the assistant as a computational aid and curated expert tool rather than a general intelligence.
3. The Interpretation Gap and Accountability: Precise reporting requirements prevent the AI from providing deep contextual storytelling, even as the neutrality protocol prevents sensationalist claims. Nuanced contextualization remains a task that only the human researcher can complete. This gap reinforces the necessity of the human-in-the-loop framework, as responsibility for the final truth and the socio-economic consequences of a policy decision must remain with the human agent.

5.3. Roadmap for Future Work

The evolution of the BDB Assistant will focus on expanding analytical depth and inter-domain capabilities to better address the complexity of modern social challenges.

1. Transition to Unit-Record Analysis: Pilot programs within highly secure environments for individual-level unit-record analysis represent a primary roadmap item. The RAG-to-code logic could generate sophisticated econometric scripts that run within isolated execution environments against raw records to produce

aggregated, non-identifiable results, significantly increasing the precision of policy evaluations.

2. Multi-Domain Cross-Querying: Future iterations aim to integrate a broader array of social determinants by linking labor market vacancies with regional housing affordability and healthcare accessibility metrics.
3. Inter-Regional Comparison Logic: Refinements to the semantic reasoning engine will better support comparative studies between non-contiguous regions that share similar socio-economic profiles.
4. Enhanced Synthesis Integration: Future updates will enable the assistant to draft templated policy briefs that include a mandatory review and override stage. This will allow human researchers to manually adjust the narrative synthesis to ensure it aligns with broader policy context and institutional knowledge.

This work has successfully pioneered a RAG-to-code architecture in a high-stakes and knowledge-intensive policy domain. The BDB Assistant provides a robust, scalable model for the secure and transparent application of generative AI in the public sector by bridging the gap between natural language inquiry and complex administrative datasets. The design broadens institutional accessibility to integrated data while maintaining the rigorous privacy standards essential for governance through radical transparency, human-in-the-loop oversight, and an isolated processing model. Ultimately, this tool empowers a diverse range of stakeholders to move beyond data discovery toward high-precision, evidence-based interventions to address systemic intergenerational disadvantage.

References

Ada Lovelace Institute, AI Now Institute, and Open Government Partnership. 2021. Algorithmic Accountability for the Public Sector. Open Government Partnership. https://www.opengovpartnership.org/documents/algorithmic-accountability-public-sector/.

Ananyev, M., A. Gamarra Rondinel, U. KC, S. Marchand, A. A. Payne, and C. R. Samarage. 2025. Breaking down Barriers - Shared Data Environment Poverty and Disadvantage Datasets. Dataset; Melbourne Institute of Applied Economic; Social Research, The University of Melbourne.

Ananyev, M., U. KC, A. A. Payne, and C. R. Samarage. 2023. Breaking down Barriers Community Profiles. Data visualisation; Melbourne Institute: Applied Economic & Social Research, The University of Melbourne. https://bdbprofiles.melbourneinstitute.unimelb.edu.au/.

Australian Bureau of Statistics. 2025. Person-Level Integrated Data Asset (PLIDA): User Guide. Australian Government. https://www.abs.gov.au/about/data-services/data-integration/integrated-data/person-level-integrated-data-asset-plida/plida-data.

Busuioc, Madalina. 2021. "Accountable Artificial Intelligence: Holding Algorithms to Account." Public Administration Review 81 (5): 825–36.

Cheong, Ben Chester. 2024. "Transparency and Accountability in AI Systems: Safeguarding Wellbeing in the Age of Algorithmic Decision-Making." Frontiers in Human Dynamics 6: 1421273.

Cole, Shawn, Iqbal Dhaliwal, Anja Sautmann, and Lars Vilhuber, eds. 2020. Handbook on Using Administrative Data for Research and Evidence-Based Policy. Abdul Latif Jameel Poverty Action Lab.

Connelly, Roxanne, Christopher J Playford, Vernon Gayle, and Chris Dibben. 2016. "The Role of Administrative Data in the Big Data Revolution in Social Science Research." Social Science Research 59: 1–12.

Eliadis, Pearl, Margaret M Hill, and Michael Howlett. 2005. Designing Government: From Instruments to Governance. McGill-Queen's Press-MQUP.

Howlett, Michael. 2009. "Policy Analytical Capacity and Evidence-Based Policy-Making: Lessons from Canada." Canadian Public Administration 52 (2): 153–75.

Huang, Lei, Weijiang Yu, Weitao Ma, et al. 2025. “A Survey on Hallucination in Large Language Models: Principles, Taxonomy, Challenges, and Open Questions.” ACM Transactions on Information Systems 43 (2): 1–55.

Kelman, Christopher W, A John Bass, and Cashel DJ Holman. 2002. “Research Use of Linked Health Data—a Best Practice Protocol.” Australian and New Zealand Journal of Public Health 26 (3): 251–55.

Leigh, A. 2025. Unlocking the Value of Better Use of Integrated Government Data for Evidence-Based Policy. Speech to the Australian Bureau of Statistics; The Treasury, Australian Government. https://ministers.treasury.gov.au/ministers/andrew-leigh-2025/speeches/address-unlocking-value-better-use-integrated-government-data-evidence-based-policy.

Lewis, Patrick, Ethan Perez, Aleksandra Piktus, et al. 2020. “Retrieval-Augmented Generation for Knowledge-Intensive Nlp Tasks.” Advances in Neural Information Processing Systems 33: 9459–74.

Lucke, Jörn von, and Sander Frank. 2025. “Potentials of Administrative Informatics for the Analysis of Policymaking: Notes on the Integration of Administrative Informatics into the Policy Cycle.” Data & Policy 7: e47.

Payne, Abigail, and Rajeev Samarage. 2020. “Spatial and Community Dimensions of Income Poverty.” Melbourne Institute of Applied Economic and Social Research, University of Melbourne.

Platt, Lucinda, Gundi Knies, Renee Luthra, Alita Nandi, and Michaela Benzeval. 2020. “Understanding Society at 10 Years.” European Sociological Review 36 (6): 976–88.

Schmeling, Juliane, Sami Al Dakruni, and Ines Mergel. 2025. “Data Collaboration in Digital Government Research: A Literature Review and Research Agenda.” Government Information Quarterly 42 (3): 102063.

Taeihagh, Araz. 2021. “Governance of Artificial Intelligence.” Policy and Society 40 (2): 137–57.

Ujjwal, KC, Steeve Marchand, and A Abigail Payne. 2025. “YouthView: A Platform for Interactive Visualizations to Explore Youth Disadvantage.” Data & Policy 7: e69.

Zhao, Siyun, Yuqing Yang, Zilong Wang, Zhiyuan He, Luna K Qiu, and Lili Qiu. 2024. “Retrieval Augmented Generation (Rag) and Beyond: A Comprehensive Survey on How to Make Your Llms Use External Data More Wisely.” arXiv Preprint arXiv:2409.14924.

Zhao, Wayne Xin, Kun Zhou, Junyi Li, et al. 2023. “A Survey of Large Language Models.” arXiv Preprint arXiv:2303.18223 1 (2): 1–124.

THE UNIVERSITY OF
MELBOURNE

MELBOURNE
INSTITUTE