ImpactEngines

CASE STUDY · JULY 2026

Case Study: Using an AI Agent for Samoa Agriculture Survey 2025 Data Cleaning

Samoa Bureau of Statistics & Impact Engines

01 · INTRODUCTION

Impact Engines has been testing wide-ranging applications of agentic AI in national statistics. One of those areas is survey data cleaning. Data cleaning is not glamorous, but it is central to the quality of official statistics. It requires careful review of case status, duplicates, missing values, outliers, skip patterns, open-ended responses, coded classifications, and consistency across linked files. It also requires a high degree of rigour; every decision should be documented, every correction should be traceable, and the final dataset should be reproducible from the original raw data.

This combination makes data cleaning a useful test case for an AI agent. The work involves a large amount of coding, checking, summarising, and documentation, but the final decisions still need human judgement. An agent can write scripts, inspect data, prepare logs and run verification checks. Humans bring project context, field knowledge and accountability for what should actually happen to the data.

Impact Engines partnered with the Samoa Bureau of Statistics (SBS) on the Samoa Agriculture Survey (SAS) 2025 data cleaning process. SBS had completed extensive data collection and needed to move the microdata through a structured cleaning workflow while managing the usual time pressure and resource constraints within a national statistics office.

02 · WORKFLOW DESIGN

A human-in-the-loop workflow for the cleaning process was established based on best practices in survey cleaning methods. The goal was for AI agents to handle repetitive, high-volume tasks while ensuring that humans made the critical decisions.

The initial instruction prompt, which was approximately 1,500 words long, provided the AI agent with important information about the project context, file structure, roles, workflow, deliverables, coding rules, and review points. It included clear instructions for the agent to stop and ask for clarification whenever it was unsure of how to proceed. This clarification-seeking approach was designed to minimize unsupported assumptions and reduce the risk of AI hallucination.

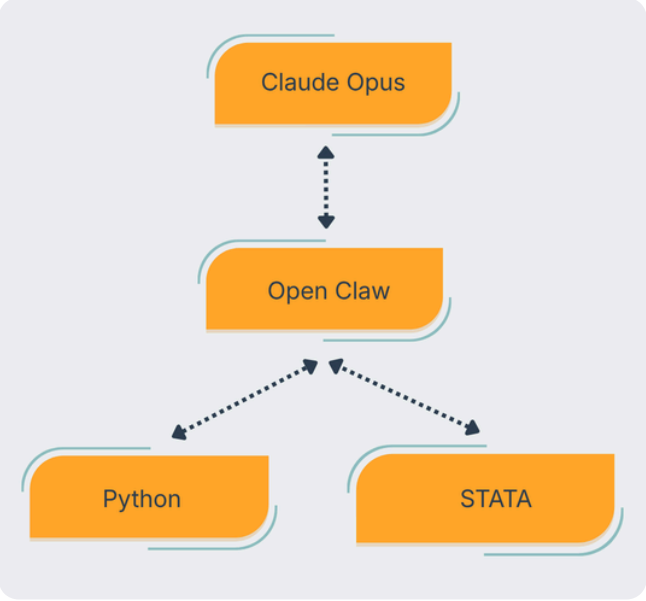
Steps in the workflow process

Step	Description
1. Review and diagnose	The AI agent mapped the data structure and developed Python scripts to run checks across the main file and rosters.
2. Recommend	The AI agent reviews the check outputs and creates a structured issue log in Excel. Each entry includes the variable information, a description of the problem, and a recommended action.
3. Decide	SBS reviewed each item in the issue log, then accepted, rejected, or modified recommendations based on their understanding of the fieldwork and local context.
4. Implement	The AI agent translated all decisions into modular, repeatable and documented Stata code, then ran and debugged the cleaning pipeline. The deliverable was a set of cleaned Stata microdata.
5. Verify and approve	SBS conducted human-led independent quality control of the cleaned microdata. Any actions not implemented correctly were returned to the AI agent for correction and updating of the cleaning pipeline and the deliverable.

03 · SOFTWARE AND TOOL SELECTION

To implement an agentic AI system, you first need software that creates a workspace for the AI agent to operate in. When installed locally, this software sets up a folder where the agent can create and access local files, such as scripts and data files. In this case, the open-source software "Open Claw" was used for this purpose.

Additionally, the system requires a large language model (LLM) to serve as its "brain." For this project, Claude Opus 4.6 was chosen as it was the leading LLM at the time. It was connected to Open Claw through an application programming interface (API).



To complete an agentic setup, specific tools needed to be connected. Python was selected for the initial steps of data review and to create the issue log. It was selected because it is a multifunctional language that can effectively query the raw database and produce Excel documents. For the cleaning implementation, STATA was selected because it produced the desired output format, and the do-files are understandable and reviewable for the SBS team.

Key considerations in the selection of software and tools used for the AI agent:

- Claude Opus was the most capable model available - but also the most expensive. For this pilot, we purposefully wanted to optimise the functionality rather than trying to minimise model cost
- Stata was used as the cleaned data needed to be generated through a reproducible statistical workflow, not through ad hoc spreadsheet edits.
- OpenClaw was used to provide the agentic workspace; at the time, it was the most capable software of its type, since the pilot took place, there have been big developments in both OpenAI Codex and Claude Co-Work applications, which would make them simpler and equally as effective to use for future reproduction of this workflow.

04 · DATA CHECKING PROCESS

The agent applied rigorous quality-control checks across the micro data set. During testing, it became clear that if left to fully design the data-check categories, the AI agent would not suggest a list as thorough as required for national statistics, so human experts developed a comprehensive list of check categories for a national-level survey.

List of data checks run by the AI agent

01 duplicate records, interview status and case eligibility	**02** missing answers, distinguishing genuine non-response from valid skips	**03** numeric outliers and their surrounding context
04 validation errors remaining in the exported data	**05** open-ended responses requiring translation, grouping or back-coding	**06** occupation descriptions compared with assigned ISCO codes
07 cross-variable inconsistencies and skip-pattern errors	**08** questionnaire version changes affecting routing and apparent missingness	**09** alignment between multi-select parent variables and generated roster rows

After creating the scripts to check all categories, the AI agent created an error log document detailing all points that require checking. It also recommended how to handle each point. This error log contained 610 items, which SBS reviewed in detail. In just over 80% of cases, SBS accepted the agent's recommendation; in the remainder, SBS chose a different action based on field knowledge, questionnaire intent, or subject-matter expertise. That review step ensured the SBS team remained aware of, and responsible for, every change made to the data.

05 · IMPLEMENTATION

With these final decisions from SBS, the AI agent's next step was to write STATA do-files that would handle any required changes in the data. This was to comprise a final deliverable: a reproducible cleaning package that could regenerate the cleaned dataset from the raw, de-identified microdata.

The AI agent was configured to interact directly with STATA through STATA's command-line batch mode. This meant it could run its own do-files, review and verify the output, and make corrections accordingly.

The final cleaned sample was 2,902 households (a 99.2% response rate). The package included 47 cleaned Stata files, a manifest, modular cleaning do-files, the reviewed QC log, Python verification scripts, questionnaire reference material and run logs.

Across the project, the AI agent produced 32 Python scripts (5,815 lines) for checking, analysis and verification. The final package contained 11 Stata do-files (3,017 lines). These are physical line counts, including comments and spacing.

8,000 lines of Python & STATA code

47 cleaned Stata files in the package

06 - QUALITY CONTROL

One of the most critical parts of the pilot was the human review of the first deliverable. AI agents can be confident in their mistakes, so we planned a thorough human review step at this point in the workflow. Just as we would review and check a cleaning system developed by a colleague or external consultant, we checked the AI agent's work.

Human review identified several issues that needed correction. These included questions where "No" was not coded as zero, an aquaculture follow-up question with a different enabling condition from the rest of the aquaculture block, a missing Yes/No battery recode, a stray working variable that had leaked into the output, and new open-ended option columns that needed better labelling and ordering. Similar to the type of mistakes programmers may make on a complex survey cleaning job.

The quality result was strong, but not perfect. Against 610 review rows in the error log, the human review of the first deliverable identified six correction items. On that basis, around 99% of the reviewed items did not require correction. That figure should be interpreted cautiously because not every row in the review workbook represents an equal unit of work. Still, it gives a useful indication of scale: the agent got the overwhelming majority of the implementation right, while expert human QC still found important issues that needed fixing.

The human QC review reinforced an important lesson: an AI agent can move quickly, but survey cleaning still needs expert review. The best workflow is not full automation, but a structured process with a human-in-the-loop for decision-making and quality control.

07 ·PRIVACY AND SECURITY

The data was de-identified before being placed in the agent's workspace. Although the full database was not sent to the model, de-identification reduced the risk that respondent-identifying details could be exposed through prompts, logs, code snippets, summaries, or human error. The agent worked on the statistical content needed for cleaning, while direct respondent identifiers were kept outside the system.

The agent ran on its own secure PC, with outbound connections used only to message the project team and connect to the model API.

API access was used rather than a consumer chatbot interface because it provided better control over cost, project separation and data retention settings.

08 ·COST

The cost of LLM tokens for the pilot was USD 114.21. Most of this cost came from the code-heavy implementation phase, where the agent created the Stata cleaning system, ran it, debugged it and carried out internal QC.

Breakdown of token cost for each task

Component	Token cost (USD)
Initial ingestion of prompt and project setup	4.68
Data checks and production of issue log	19.78
Develop clean data	82.00
Revision of deliverable following human QC	7.75
Total	**114.21**

The time saving was substantial. According to the SBS estimate, the agent's work would **have taken up to 20 days of a data processing specialist's time**. The pilot did not remove the need for expert human time, but it shifted that time toward review, decision-making and quality assurance rather than repetitive coding and manual checking.

09 ·REFLECTIONS

Using the cleaned data delivered by the AI agent, SBS was able to proceed through the steps of weighting, tabulations, and report writing without needing to return to resolve further discrepancies. This shows that the output quality from this human-in-the-loop and AI workflow was high. Undertaking this type of pilot activity using agricultural microdata was ambitious, given the complexity of agricultural data and represents a clear proof of concept that AI agents have advanced to a place where they can now be safely and efficiently used for national statistics work to produce high-quality outputs.

The AI agent excelled in environments where the tasks were structured, repetitive, and code-intensive. It could swiftly inspect numerous variables, write scripts, generate logs, check outputs, rerun Stata, and document results much faster than a human analyst working manually. However, the agent struggled with tasks that required judgment about field reality, questionnaire intent, or institutional preferences. For instance, decisions regarding whether to drop a case, the plausibility of an outlier, or how to handle an ambiguous "other specify" response still required human insight.

A key lesson learned was the importance of detailed prompting. The agent needed a clear workflow, well-defined responsibilities, and very specific instructions regarding the structure of the quality control (QC) log. It also required the ability to ask questions; AI systems can be overly eager to please and may rush into decisions if the workflow does not explicitly allow for clarification in both initial prompts and ongoing instructions. The system functioned effectively because it adhered to this workflow. The agent did not replace the statistician or data manager; instead, it acted more like a disciplined analytical operator that prepared evidence, wrote code, implemented approved decisions, and ensured that the output complied with established rules.

09 · CONCLUSION

For Impact Engines, the pilot served as a compelling proof of concept. A complex national survey database was cleaned in significantly fewer human hours than typically required, all while maintaining a transparent record of decisions and producing high-quality, reproducible code. The human-in-the-loop workflow was essential; it kept the SBS and the data cleaning team in control of decision-making while allowing the AI agent to handle the repetitive, technical, and code-intensive tasks quickly.

This approach presents a promising model for official statistics and international development. When used thoughtfully, agent-based AI can alleviate the burden of tedious technical work while ensuring that critical decisions remain with the individuals who understand the survey, the country context, and the statistical goals of the data.

Impact Engines will continue to explore the area of agentic AI in relation to development statistics. Data cleaning is one aspect of the opportunity, but the same principles can be applied to other workflows where code, documentation, quality control, and human judgment must work together.

For example, in-field survey quality control can benefit from agents that review incoming paradata and response patterns during data collection. Additionally, in tabulation and analysis workflows, agents can assist in producing reproducible statistical tables, analysis scripts, and publication-ready outputs under expert supervision.

The goal is not to eliminate statisticians from the process; instead, it is to provide them with better tools to conduct careful work at the scale and pace demanded by modern survey programs.

The pilot of the Samoa Agriculture Survey 2025 demonstrated that an AI agent can effectively support national survey data workflows in a practical and accountable manner. The key benefit was not merely the automation of processes without oversight, but rather the integration of local data processing, structured review logs, reproducible Stata code, human decision-making, and repeated verification.

Written by Lachlan Bruce - lbruce@impactengines.ai