

Case studies

CONTENTS

[Some Case Studies](#)

[AI-assisted causal mapping. Uncovering causal pathways with INTRAC](#)

[Can AI accurately map causal claims - A validation study](#)

[DUOC UC. Evaluating Gender Equity in STEM with AI-Driven Interviews](#)

[Strengthening OH with causal mapping](#)

[Tree Aid - Empowering Communities Through Forest Management in Burkina Faso](#)

[Using AI to facilitate feedback on the learning experiences of doctoral students](#)

Some Case Studies

10 Dec 2025

Here are some examples of work with [Causal Map](#) and causal mapping, and also with Qualia interviews.

AI-assisted causal mapping. Uncovering causal pathways with INTRAC

8 Jan 2026

The partner

The partner

We partnered with INTRAC on a project for the SCC (Strengthening Civil Society) programme. This programme is led by a consortium of four organisations: PAX, ABAAD, DefendDefenders and Amnesty International Netherlands, with INTRAC serving as the MEL partner. INTRAC is a non-profit organisation with a deep commitment to understanding and improving the practice of civil society worldwide.

The challenge

The SCC programme is a complex, multi-year initiative spanning 13 countries. The INTRAC team was using Outcome Harvesting but found themselves struggling to get the 'big picture'. They needed a systematic, yet flexible, way to connect the dots in a large body of qualitative data to answer critical questions about the programme's overarching causal pathways of change.

The core challenge was translating a high volume of text into high-quality, actionable analysis. Alastair Spray, Senior Consultant at INTRAC, led the causal mapping component of the End Term Evaluation. He worked alongside a core team of four, supported by two other colleagues who provided input into the wider evaluation. As Alastair noted:

"The client had been pursuing usual evaluation approaches... and had found Outcome Harvesting useful, but they felt that they couldn't understand the big picture, for the whole programme but also for specific countries too."

The team needed to rigorously analyse the data to answer questions such as:

- How have the different levels of Lobby & Advocacy interventions fared (national, international, local)?
- What are the causal stories at the individual country level?

- Are there shifts in causal pathways over time, particularly when comparing the programme's earlier and later years?
- How do the causal pathways link directly to the four specific objectives of the SCC programme?

The Causal Map solution

We used Causal Map's consulting service and our AI-assisted approach to bring clarity and rigour to this complex qualitative data, creating a process that was both systematic and reproducible for INTRAC.

Given the sensitive nature of the SCC programme, we worked closely with the team to ensure all data was thoroughly anonymised before the mapping and analysis began.

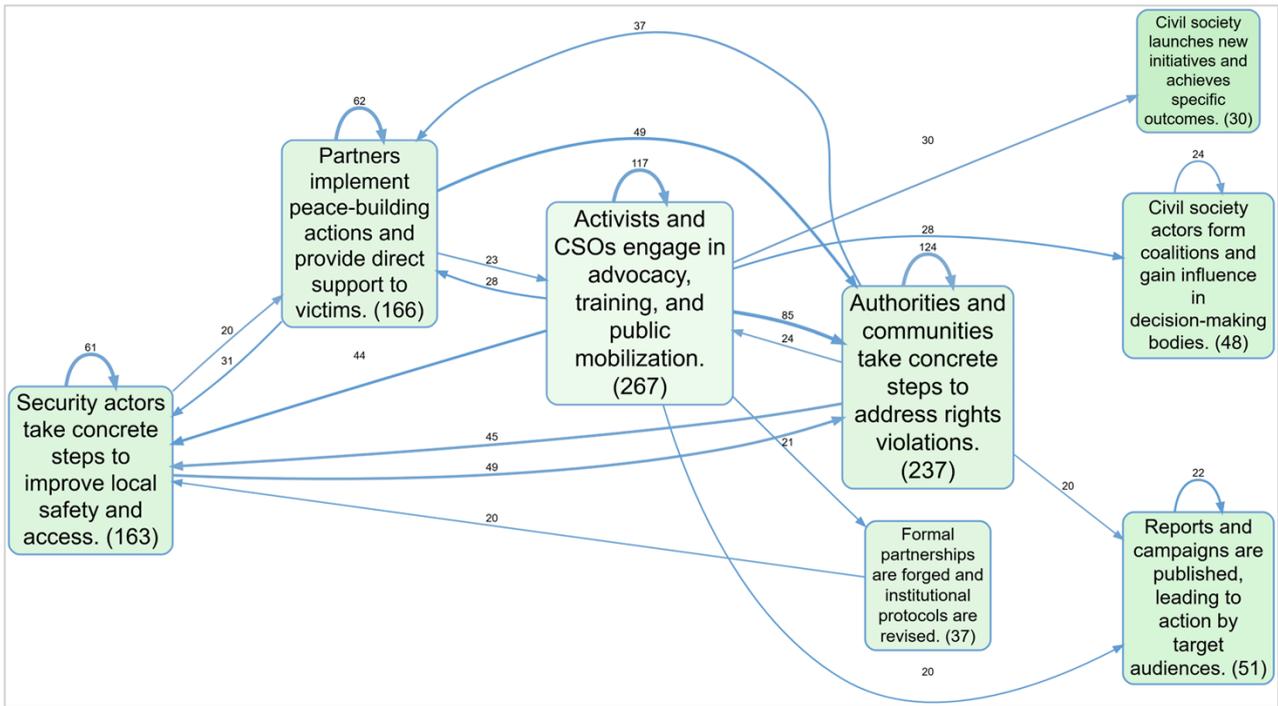
Our approach

1. **Data input:** We processed comprehensive data sources, including Harvested outcomes data from the 13 countries across four programme years, and Narrative Theory of Change Review documents (typically 8–15 pages long) from partner organisations.
2. **AI-Assisted coding:** The Causal Map team was responsible for the core coding. We used a [radical zero-shot approach](#) with our AI (Gemini 2.5 Pro), which was instructed to invent its own codes based purely on the text. This process quickly identified 5,430 causal claims from the respondents, which were also auto-coded for sentiment (positive/negative links).
3. **Collaborative Analysis:** We provided a preliminary analysis report and materials, along with specific training for Alastair on the cm4 platform. This allowed the INTRAC team to engage directly with the maps and dig deeper into the coding when writing their final report. This hands-on support was crucial for building confidence, which Alastair highlighted:

“The team (Steve and Gabriele) are incredibly helpful, and good at training you up on how you might want to use it to interrogate your data. My advice would be to learn by doing, try to get started and then ask for help as you need it, rather than worrying about getting it right from the start.”

Results

The use of Causal Map for this analysis generated a powerful set of **visual and analytical outputs**, allowing INTRAC to move from raw data to a better understanding of 'what causes what' in their programme.



- **Overall and country-specific maps:** A high-level causal map for the entire programme, along with 13 individual maps, one for each country, to capture localised dynamics.
- **Targeted analysis:** Alastair worked on creating filtered maps to specifically show outcomes linked to (1) Lobby & Advocacy activities and (2) local ownership, which was a key area of interest for the PAX Head of Programs.
- **Change over time:** By filtering the data into two distinct periods, the team could identify any shifts in causal pathways since the mid-term review.

The project provided INTRAC with a robust, visual evidence base that articulated how stakeholders perceived change in the SCC programme, helping them to move from raw data to a deeper understanding of 'what causes what' in their complex field.

See what Alastair has to say:

“Causal Map really allowed us to get that higher level of understanding, in both a clear visual but also high quality written analysis.”

On the Causal Map 4 app: “The app itself is really clear, intuitive (after a lesson of course), and the info/help icons are great.”

On the AI/Vignettes: “The vignettes function (using the in-built AI to generate written analysis of the selected causal map) is incredibly useful, and saves a lot of time! Once you get the hang of giving it better and better prompts you inevitably get better results too, [...] and the client clearly liked what it churned out too.”

On the support: “A small point, but important one, you were both very good at building my confidence in using it! Questions I asked weren’t considered obvious or basic, and the initial attempts at making maps or analysis were met with enthusiasm / encouragement, which on reflection I’m sure made me more keen to use the app. I think it would be easy for a non-tech savy user to get into their own head that ‘they are bad at it’ and that would inevitably affect how (and how much) they actually use it.”

On the results: “The analysis I produced from using Causal Map was received very well by the client. It was very useful in validating their approach and theory of change, and they appreciated how clearly the combination of the maps and vignettes revealed exactly what the outcomes were and what had (and hadn’t) caused them.”

— Alastair Spray (Senior Consultant, INTRAC)

Can AI accurately map causal claims - A validation study

3 Mar 2026

_Summarised from "[AI-assisted causal mapping: a validation study](#)" by Steve Powell and Gabriele Caldas Cabral.

Extracting causal claims from qualitative interviews is a notoriously slow process. Evaluators spend hours combing through transcripts to work out what respondents think influences what. Generative AI offers a solution, but a critical question remains. Is an untrained AI assistant actually any good at identifying and labelling these causal links compared to human experts?

We conducted a validation study to find out. The results show that treating AI as a tireless, low-level coding assistant works surprisingly well.

The approach: naive causal coding

We used a pragmatic, minimalist approach to causation. We did not ask the AI to model complex systems, determine the strength of causal effects, or make predictions. We simply asked it to identify evidence of causal influence within the text.

The task was reduced to a clear set of instructions: read the text, find the causal claims, and tell us what influences what. By asking the AI only to extract claims rather than interpret their ultimate validity, we keep the human evaluator firmly in charge of the actual analysis.

The experiment

We used a dataset from a Qualitative Impact Protocol (QuIP) study evaluating an agriculture and nutrition programme. Expert human analysts had already coded these transcripts by hand. This provided our benchmark.

We took a subset of the data (163 statements from three sources) and instructed GPT-4.0 to code it. The AI processed the text statement by statement, extracting the cause, the effect, and the verbatim quote proving the claim. We ran the test in two ways:

1. Radical zero-shot: The AI was given research context but no codebook or suggested labels. It had to invent its own factor labels.

2. Closed coding: The AI was provided with a basic codebook of 29 top-level labels derived from the prior human coding.**

The Results: high precision and high recall

The AI performed at a level comparable to a human assistant. We evaluated the results on both precision (were the links correct?) and recall (did it find all the links?).

- Precision: Without a codebook, 84% of the causal links identified by the AI were rated as perfectly correct across four strict criteria. When given a basic codebook, this precision rate rose to 87%. The errors the AI did make were usually minor labelling discrepancies or ‘noise’ rather than complete misunderstandings of the text.
- Recall: The AI successfully identified approximately as many valid causal links as the human experts. In some transcripts, it actually found valid links that the human coders had missed.

The bigger picture: comparing the maps

In qualitative data analysis, two human coders rarely produce the exact same set of detailed labels. The AI was no different. The raw, detailed maps produced by the AI and the humans differed in specific terminology.

However, evaluation rarely relies on looking at every single microscopic link. When we zoomed out to create high-level overview maps showing the most frequently mentioned factors and links, the structures were very similar to each other.

Figure 1: Overview causal map generated from AI coding.

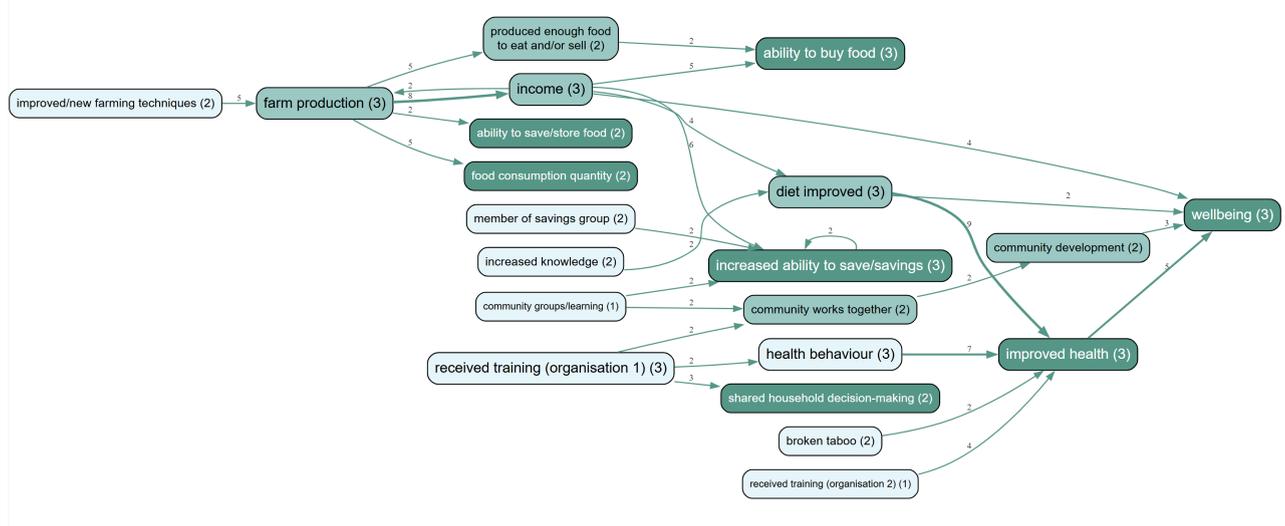
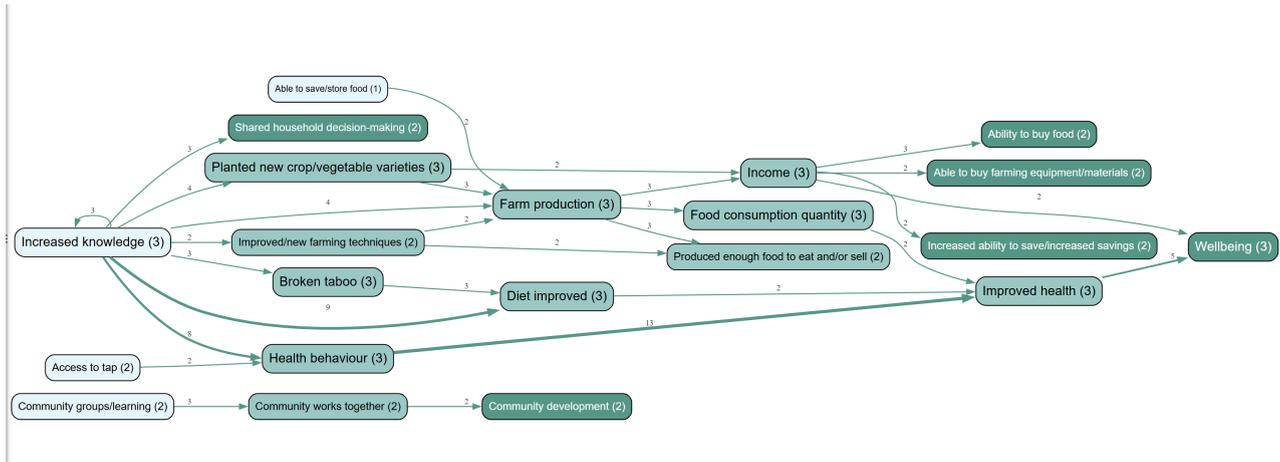


Figure 2: Overview causal map generated from human expert coding.



The AI map and the human map told the same overarching story. The AI tended to prefer labels like "received training", whereas the human coders preferred "increased knowledge", but both were perfectly reasonable interpretations of the underlying text.

The verdict

Generative AI is highly capable of extracting causal claims from qualitative data. Using a naive approach to causal coding, AI achieved a precision rate well over 80% GPT-4.0, which is now an outdated model.

This does not turn the AI into a black box that replaces the evaluator. Instead, the AI functions as a transparent assistant. It does the heavy lifting of initial extraction, reading through pages of text to pull out claims and attach the exact quotes that prove them. This allows evaluators to process qualitative data rapidly and at scale, freeing up time for the human judgment and synthesis that actually matters.

DUOC UC. Evaluating Gender Equity in STEM with AI-Driven Interviews

15 Jan 2026

06/08/2024

The partner

In 2024, we collaborated with researchers at **DuocUC**, a leading higher education institution in Chile. The project was led by **Felipe Rivera**, Head of Academic Quality Evaluation and **Javiera Cienfuegos**, a Senior Researcher at DuocUC and they focused on understanding and improving the educational landscape for women in technical and professional fields.

The challenge

DuocUC needed to conduct a qualitative, explorative evaluation of the gender gaps faced by women pursuing STEM (Science, Technology, Engineering, and Mathematics) careers at the university.

Traditional qualitative methods often struggle to scale while maintaining the depth required to uncover complex social issues. The team needed a way to collect authentic, personal narratives from a large group of students and then rigorously analyse that data to understand the underlying causal drivers of gender discrimination.

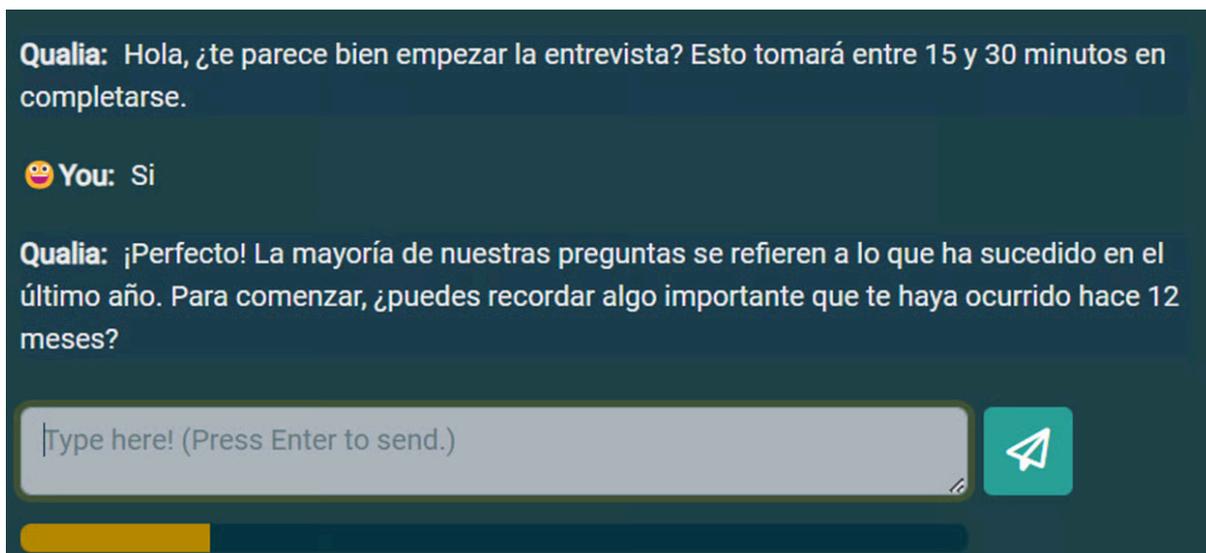
The Causal Map solution

We deployed our "[Seamless Stories](#)" workflow—an end-to-end, AI-supported process that combines [QualiaInterviews](#) for data collection and [Causal Map](#) for systematic analysis. This approach allowed the university to move from research questions to high-quality visual maps much faster than traditional manual coding, without sacrificing the nuance of the students' lived experiences.

Our approach

We had a first meeting to understand what they wanted to find out, their research questions and the scope of the study and to determine the domains in which the interviews would be conducted.

1. **Automated Data Collection (Qualia):** We designed a responsive, chat-based interview in Spanish using QuIP-style questions. Our AI interviewer, Qualia, powered by GPT-4o, was instructed to explore three key domains: educational experiences, professional development, and relationship dynamics. Out of 50 students invited, 32 completed the conversational interviews.



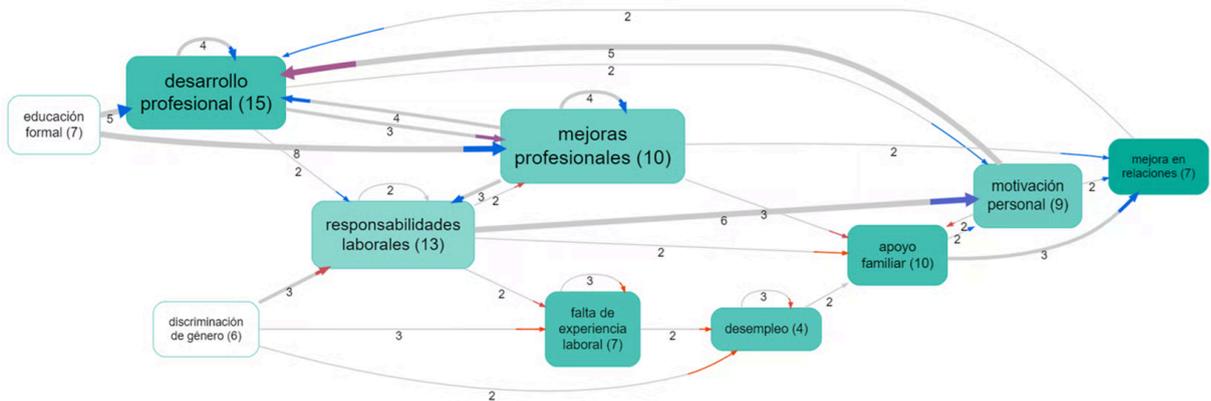
2. **Radical Zero-Shot AI Coding:** Once the interviews were completed, the transcripts were uploaded to Causal Map, and we used a ["radical zero-shot" approach](#) to code the data using GPT-4o. The AI was given no pre-defined codebook; instead, it was instructed to "invent" its own codes based purely on the students' narratives (in Spanish). This process identified 251 distinct causal links.
3. **Targeted Research Analysis:** We used Causal Map's filtering tools to interrogate the data and answer specific research questions, such as "What is the immediate impact of gender discrimination?" and "What are the most mentioned causal factors?" We also used the "AI Answers" feature to generate summaries of the broader text independently of the causal coding.

Results

The project provided DuocUC with a visual, evidence-based understanding of the "causal network" surrounding gender discrimination in their STEM programmes.

- **Sentiment-coded maps:** We auto-coded the sentiment of every link, using blue arrows for positive contributions and red for negative ones, allowing the researchers to see exactly where the "friction points" were in the students' journeys.
- **Synthesis:** The workflow transformed 32 in-depth interviews into manageable causal maps, making it easy to identify the most frequent factors mentioned by respondents.

- **Actionable insights:** By filtering the maps, the team could isolate specific pathways of change, helping DuocUC to validate where their current interventions were working and where new support was needed.



See what Javiera has to say:

"The type of questions that were asked "what causes what", were equally linked to methodological innovation. The results were able to portray how gender barriers are intertwined in domains ranging from higher STEM education to the performance of new professionals and technicians once they enter the labour market, reaching deeper explanations and social impact." — Javiera Cienfuegos (Senior Researcher, DuocUC)

Strengthening OH with causal mapping

8 Jan 2026

STRENGTHENING OUTCOME HARVESTING ANALYSIS WITH AI-ASSISTED CAUSAL MAPPING

Written by:



Heather
Britt



Steve
Powell



Gabriele
Caldas Cabral

**CAUSAL
MAP**
MAKING CONNECTIONS



Here is a version of the final text. Check the official publication [here](#)

Strengthening Outcome Harvesting Analysis with AI-Assisted Causal Mapping - shortened full version

Written by: Heather Britt, Steve Powell, Gabriele Caldas Cabral

Summary

This case study explores how AI-assisted causal mapping can enhance Outcome Harvesting (OH) analysis by revealing interrelationships between outcomes and identifying new actors contributing to change. The pilot demonstrates how this approach provides actionable insights and strengthens causal relationship analysis in OH. It emphasizes the importance of a principle-led analysis plan and human expertise in guiding the AI process.

Introduction

Outcome Harvesting is a powerful approach for discovering emergent changes—whether predicted or unpredicted, positive or negative—and documenting how those changes occurred. While many methods capture changes in those directly involved with a project, OH captures changes farther down the causal pathway.

However, evaluators often struggle to explore interrelationships between multiple outcomes. This case study describes how an OH practitioner (Heather Britt) collaborated with causal mapping practitioners (Steve Powell and Gabriele Caldas) to expand causal contribution analysis in OH using an AI-assisted causal mapping app. They analyzed OH data from a completed education project.

Outcome Harvesting: Analysis Limitations

While OH documents causal pathways contributing to individual outcomes well, evaluators find it difficult to make sense of interrelationships between multiple outcomes and their causal pathways. This limits their ability to answer questions about causal contribution.

Current OH practice often uses descriptive statistics to summarize data by outcome components (e.g., types of change agents or social actors) and reports findings in charts. Another approach arranges outcomes on a timeline to determine logical relationships.

Our pilot explores whether AI-assisted causal mapping can address these limitations by analyzing causal relationships between outcomes.

The Pilot

Core Question

Can AI-assisted causal mapping address the limitations of OH analysis?

Heather Britt reached out to Steve Powell and Gabriele Caldas to explore whether causal mapping with the Causal Map app could enhance OH analysis.

Causal mapping techniques, developed over 50 years ago, have been used across disciplines to identify and visually represent causal relationships in qualitative data. The Causal Map app computerizes this technique, allowing efficient coding, analysis, and visualization of information from multiple sources (interviews, reports, surveys, narratives), either manually or with AI assistance.

The AI-assisted capacities of the app were critical for revealing interrelationships between multiple outcomes.

Pilot Data Set

The pilot used data from the final evaluation of an education project (Girls Education project, 2016–2021) disrupted by political turmoil and COVID-19. The project adapted activities during lockdown, and OH was used to capture outcomes in five domains where the theory of change was no longer valid.

The evaluation team interviewed 49 change agents and drafted 103 outcome descriptions across five domains. The pilot data included both interview transcripts and outcome descriptions.

For the pilot, the domain **Increased community support for education** was selected, with 13 outcome descriptions analyzed.

Analysis Process

Step 1: Draft a Principle-Led Analysis Plan

Three guiding principles steered analysis decisions:

1. **Prioritize local leadership:** Use AI while keeping sensemaking and learning in the hands of local evaluators.

2. **Protect OH integrity:** Adapt methods as needed while staying true to OH principles, including “Less is more” (avoid collecting more data than can be analyzed).
3. **Produce accurate, actionable maps:** Human judgment is required to error-check data and interpret maps.

Step 2: Segment Data by Outcome Domain

Segmenting data by domain increases the likelihood of finding coherent causal pathways and facilitates error-checking. The pilot focused on one domain to analyze causal relationships between outcomes.

Step 3: Decide When to Apply AI-Assisted Causal Mapping

The pilot compared applying causal mapping to interview transcripts versus outcome descriptions. Outcome descriptions, crafted by local evaluators, were more accurate and required less error-checking than transcripts. Thus, mapping outcome descriptions was preferred to preserve local leadership and OH integrity.

Findings from Causal Maps

Relationships Between Outcomes

The AI identified causal links between the 13 outcome descriptions, revealing that outcomes influenced one another. For example, parents actively supporting home learning and leaders convincing parents to participate were central factors.

Factors Contributing to Domain-Level Outcome

Mapping revealed additional actors influencing the domain-level outcome “Community supports learning,” including unexpected contributors like Ministry officials.

Conclusion

AI-assisted causal mapping advanced OH analysis beyond descriptive statistics by:

- Analyzing multiple outcomes to determine causal contributions.

- Revealing interrelationships between causal pathways.
- Confirming known change agents and identifying unexpected influences.
- Showing how domain-level changes contribute to broader changes.

Causal mapping offers rich, flexible analysis that can be explored in multiple ways to answer diverse evaluation questions.

Tree Aid - Empowering Communities Through Forest Management in Burkina Faso

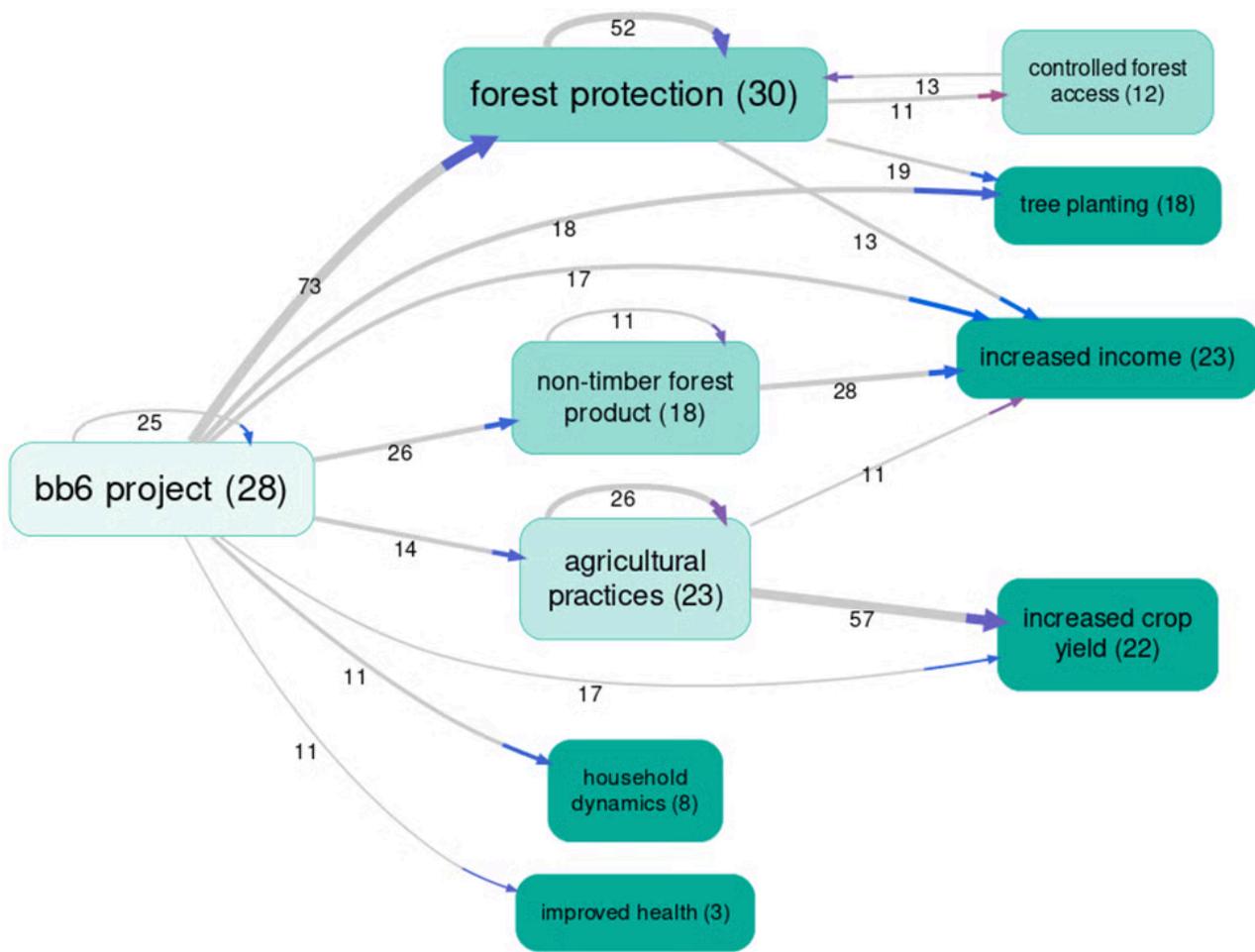
12 Dec 2025

11/12/2025

Summary

Background: Causal Map partnered with [Tree Aid](#) to evaluate their Local Governance of Forest Resources (WEOOG PAANI) project in Burkina Faso. We used [QuIP](#) to collect the data and [AI-assisted causal mapping](#) to assess the project's impact on forest governance, household and food consumption amongst project beneficiaries in two communes.

Findings: The study revealed how integrating forest management with economic empowerment and sustainable agricultural practices, has contributed to community-driven environmental conservation. And it also demonstrated how combining qualitative research with analysis tools (Causal Map app) can provide insights into community development projects.



The partner

Between February and August 2024, Causal Map has worked with Tree Aid to conduct a comprehensive evaluation of their BB6 project in Burkina Faso, focusing on 2 communes (Toécé and Gomponsom). Tree Aid is an international NGO dedicated to protecting dryland forests and supporting communities in the African drylands.

The challenge

The WEOOG PAANI project (known as BB6) aimed to improve forest governance, household incomes, and food consumption across two communes, Toécé and Gomponsom. Tree Aid needed to understand **the project's complex impact pathways**, specifically, how integrating forest management with economic empowerment influenced community-driven conservation.

They faced the challenge of synthesising diverse qualitative narratives to answer specific research questions, such as:

- How do impact pathways differ between the two communes?
- What is the specific impact of the intervention on women's lives?

- Why is the project directly related to increased crop yield?
- Are there unexpected outcomes?

The Causal Map solution

Causal Map provided extensive support throughout the evaluation process.

Training

To be able to conduct quality Qualitative Impact Protocol ([QuIP](#)) interviews, training sessions were held with the evaluation team:

1. **QuIP Lead Evaluator:** the main researchers participated in the training held by [Bath SDR](#) to learn how to design and manage a [QuIP study](#).
2. **Causal Mapping training:** A 2-day immersive training in Bristol focused on understanding causal mapping, developing interviewing skills, and preparing research deliverables.
3. **Field researchers training:** Causal Map prepared the training script. Tree Aid staff who completed the first two trainings then trained the field research team in Burkina Faso, applying concepts and techniques from the previous sessions.
4. **Causal Map app training:** After data analysis, Causal Map introduced Tree Aid analysts to use the Causal Map app, enabling deeper dives into the data.

Research Design and Data collection

We helped the Tree Aid team to develop the research design, including the research questions and the interview guides.

During the data collection phase, we provided [ongoing support and feedback](#) to the Field Researchers team. This ensured:

- Consistency in interview techniques
- Adherence to QuIP methodology
- Quality and depth of collected data
- Timely addressing of any challenges encountered in the field

Analysis

The evaluation employed the QuIP methodology, which involved:

- 31 interviews, including household interviews, focus group discussions, and key informant interviews;
- Causal mapping analysis using AI-assisted coding in the Causal Map App;
- Examination of outcomes across four domains: food consumption, income, forest management, and household dynamics.

Results

Using the Causal Map app we were able to find 1288 causal claims (links) made by the respondents and we also autocoded the sentiment of each link in order to show which contributions were "positive" (blue lines) and which were "negative" (red lines).

Through the different filters in the app, many maps and tables were created to support the quantitative data collected by Tree Aid, including:

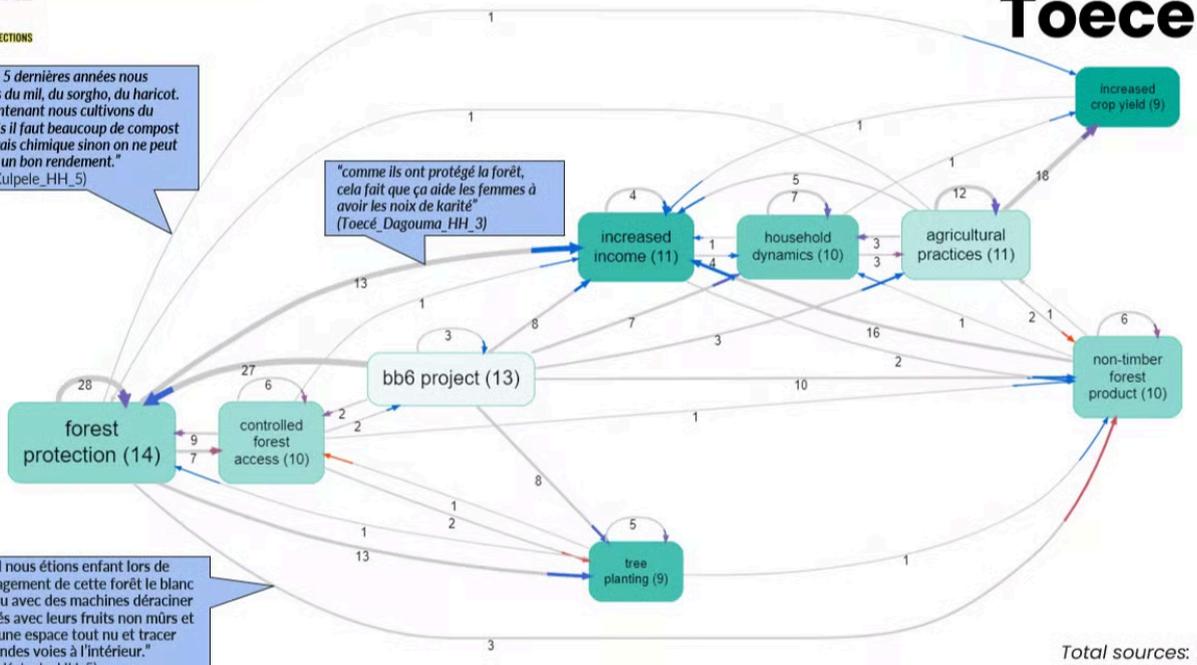
- **Comparing** maps by commune
- **Splitting** and comparing data by type of interview and domain
- **Focusing** on specific themes, such as 'Impact on women's lives'
- **Unpacking** unexpected outcomes and **answering** research questions, i.e. 'Why BB6 project is directly related to increased crop yield?'

The evidence strongly suggests that Tree Aid BB6 project has demonstrated significant **positive impacts** on the communities in Toécé and Gomponsom. By integrating forest management with economic empowerment and sustainable agricultural practices, the project has created a model for community-driven environmental conservation. The strengthened local governance structures and improved household dynamics suggest that these positive changes may be sustainable in the long term.

" dans les 5 dernières années nous cultivions du mil, du sorgho, du haricot. mais maintenant nous cultivons du maïs, mais il faut beaucoup de compost ou d'engrais chimique sinon on ne peut pas avoir un bon rendement."
(Toécé_Kulpele_HH_5)

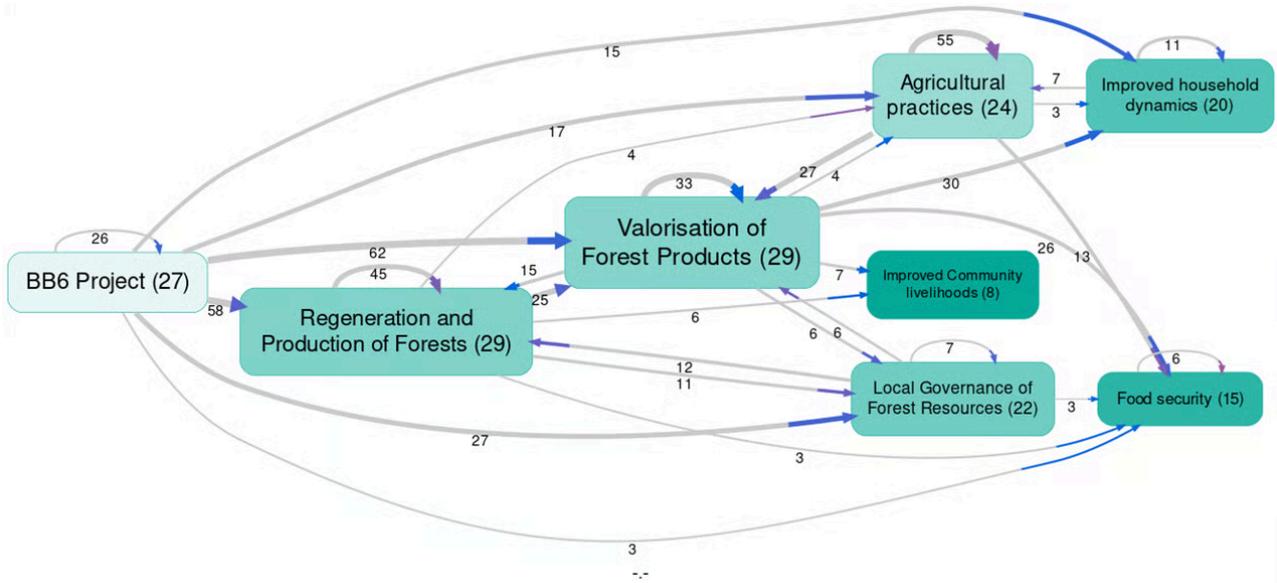
"comme ils ont protégé la forêt, cela fait que ça aide les femmes à avoir les noix de karité"
(Toécé_Dagouma_HH_3)

" quand nous étions enfant lors de l'aménagement de cette forêt le blanc est venu avec des machines déraciner les nérés avec leurs fruits non mûrs et laisser une espace tout nu et tracer des grandes voies à l'intérieur."
(Toécé_Kulpele_HH_5)



Total sources: 14

Using the “soft recoding” feature in the Causal Map app allowed us to create maps showing different perspectives of stakeholders' stories. This innovative approach enabled us to compare narratives against the project’s Theory of Change and verify the project's impacts across various domains.



Using AI to facilitate feedback on the learning experiences of doctoral students

12 Dec 2025

- Gabriele Cabral, Causal Map Ltd
- James Copestake, University of Bath
- Steve Powell, Causal Map Ltd

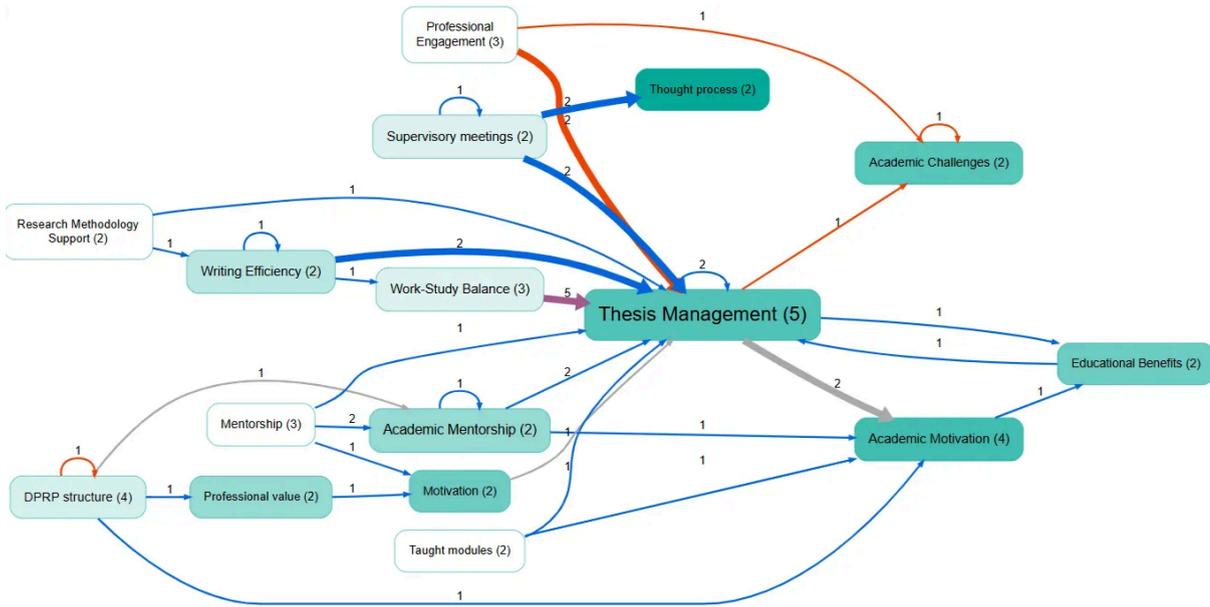
Summary

The Causal Map team has conducted a trial of an innovative approach to securing feedback from students using online open-ended interviews conducted by the app [QualiaInterviews](#), which uses generative AI (gen-AI), followed by a second use of gen-AI within the app **Causal Map** to semi-automate causal coding of the narrative transcripts thereby generated. The trial was conducted with students registered on the doctorate in policy research and practice (DPRP) at the **University of Bath**.

The pilot

We report on the trial of an innovative approach to securing feedback from students using online open-ended interviews conducted by the app Qualia, which uses generative AI (gen-AI), followed by a second use of gen-AI within the app Causal Map to semi-automate causal coding of the narrative transcripts thereby generated. The trial was conducted with students registered on the doctorate in policy research and practice (DPRP) at the University of Bath, a part-time programme for mid-career policy professionals. This generated credible evidence of diverse positive and negative drivers of learning from eleven students. The trial suggests that incorporation of gen-AI into causal mapping of narrative data about students' study experiences enhances the potential to use the method cost-effectively on a larger scale, whether alongside or instead of more traditional approaches to eliciting student feedback on teaching and learning.

Keywords: *AI; Causal mapping; Doctoral studies; Generative AI; Qualitative data analysis, Student evaluation*



Link: 685. Filename: james-dprp-research. Citation coverage 59%: 44 of 74 total citations and 6 of 6 total coded sources are shown here. Numbers on factors show source count. Factor sizes show citation count. Darker factor colours show greater outcomeness. Numbers on links show citation count. Zooming in to level 1 of the hierarchy. Auto clustering factors using label set 2. Top 20 factors by source count.

[See our findings in this paper](#)

[See a summarised report in this presentation](#)