# 🌻 AI-assisted causal mapping – Summary (validation)

(Powell & Cabral, 2025)

- **Goal / research question**
- Test whether an **untrained LLM** can **identify and label causal claims** in qualitative interview “stories” well enough to be useful, compared with **human expert coding** (a criterion study).
- Focus is on **validity/usefulness of causal-claim extraction**, not causal inference.

- **Core framing: causal mapping vs systems modelling**

- In systems mapping, an edge (X \rightarrow Y) is often read as “(X) causally influences (Y)”.
- In causal mapping (as used here), an edge means **there is evidence that (X) influences (Y) / a stakeholder claims (X) influences (Y)**.
- Output is therefore a **repository of evidence with provenance**, not a predictive system model.

- **“Naive” (minimalist) causal coding definition**

- Deliberately avoids philosophical detail; codes **undifferentiated causal influence** only.
- Does **not** encode effect size/strength; does **not** do causal inference; does **not** encode polarity as a separate field (left implicit in labels like “employment” vs “unemployment”).
- Coding decision reduced to: **where is a causal claim, and what influences what?**

- **Data and criterion reference**

- Corpus from a **QuIP** evaluation (2019) of an “Agriculture and Nutrition Programme”.
- Dataset previously hand-coded by expert analysts (used as a **criterion study**).
- Validation subset: **3 sources**, **163 statements**, **~15 A4 pages**.

- **Extraction procedure (AI as low-level assistant)**

- Implemented via the **Causal Map** web app using **GPT-4.o**.
- Temperature set to **0** for reproducibility.

 

- AI instructed to produce an **exhaustive, transparent** list of claims with **verbatim quotes**; synthesis is done later by causal mapping algorithms.
- Exclusions: **ignore hypotheticals/wishes**.
- Output per claim: statement ID + quote + influence factor + consequence factor.

- **Two validation variants**

- **Variant 1 — open coding ("radical zero-shot")**
  - No codebook; includes an "orientation" so the AI understands the research context.
  - Uses a multi-pass prompting process (initial extraction + revision passes).
- **Variant 2 — codebook-assisted ("closed-ish")**

  - Adds a partial codebook (most-used top-level labels from the human coding).
  - Uses hierarchical labels `general concept; specific concept`.
- **Validation metrics and headline results**

- **Precision** (human-rated, four criteria): correct endpoints; correct causal claim; not hypothetical; correct direction.
  - Variant 1: 180 links; perfect composite score (8/8) for **84%** of links.
  - Variant 2: 172 links; perfect composite score (8/8) for **87%** of links.
- **Recall (proxy)**: compared link counts vs the human-coded set (acknowledging no true ground truth because granularity is underdetermined).

- **Utility check (overview-map similarity)**

- Detailed maps differ (expected in qualitative coding).
- When zoomed out to top-level labels and filtered to the most frequent nodes/links, AI and human overview maps are **broadly similar**.

- **Scope limits / risks**

- Small sample; single (relatively "easy") dataset; informal rating process.
- Label choice/consistency remains a major source of variation; batching can introduce inconsistency across prompts.
- Suitable for mapping "how people think" and building auditable evidence sets; not suitable for high-stakes adjudication of specific links without checking.

## References

Powell, & Cabral (2025). *AI-assisted Causal Mapping: A Validation Study*. Routledge. https://www.tandfonline.com/doi/abs/10.1080/13645579.2025.2591157.

 