

How can domain knowledge help in identifying causal relationships in data

Domain knowledge helps identify causal relationships in data by providing expert insights about which variables are relevant and how they are causally connected, guiding the construction of causal models and reducing false discoveries. Specifically:

- Experts use their understanding of the system to specify important variables and plausible cause-effect links, which can be encoded as causal diagrams or Bayesian networks to structure the causal inference process^{[1] [2]}.
- Prior domain knowledge helps distinguish genuine causal effects from spurious correlations and confounding, making causal discovery more accurate and efficient than unguided data-driven methods^{[1] [3]}.
- It can fill gaps when some variables are latent or unobserved by informing reconstruction or estimation of missing data, improving causal inference robustness^[4].
- Domain knowledge also enables validation and refinement of causal assumptions, ensuring models reflect real-world mechanisms rather than purely statistical associations^{[5] [3]}.
- Tools and processes like expert elicitation and Delphi methods systematically capture and combine expert knowledge to build reliable causal graphs used in causal AI models^[1].

In summary, domain knowledge acts as a critical foundation for identifying valid causal relationships by informing model structure, variable selection, and assumptions, thereby enhancing the accuracy, interpretability, and applicability of causal AI.

✱

1. <https://www.causalwizard.app/inference/article/domain-knowledge>
2. https://www.pywhy.org/dowhy/v0.11/user_guide/modeling_causal_relations/index.html
3. <https://graduateschool.charlotte.edu/leveraging-domain-knowledge-enhanced-causal-structure-learning-and-out-distribution-generalization>
4. <https://pmc.ncbi.nlm.nih.gov/articles/PMC7050445/>
5. <https://www.sciencedirect.com/science/article/pii/S0370157323003411>