perplexity

What specific data sources are most effective for training causal AI models in a business context

The most effective data sources for training causal AI models in a business context are those that are high-quality, domain-specific, and representative of the causal factors influencing the business problem. Key data types include:

- **Operational and transactional data:** Detailed records of business processes, customer interactions, sales, claims, and financial transactions that capture the variables influencing outcomes[1].
- **Event and observability data:** Logs from applications, infrastructure, and user experience layers that reveal causal relationships across systems and business metrics[2] [3].
- **Real-time and time-series data:** Enables causal AI models to adapt and learn from evolving conditions, maintaining accuracy over time[4].
- **Geospatial and environmental data:** Useful especially in insurance for catastrophe modeling and risk exposure analysis[5].
- **Marketing and customer behavior data:** To identify causal drivers of engagement, churn, and conversion[5].
- **Domain knowledge and expert annotations:** To guide causal feature selection and model design, improving interpretability and reducing confounding[1] [5].

Data preparation steps such as cleaning, normalization, and identifying potential confounders are critical to ensure causal validity[1]. Leveraging diverse, context-rich datasets that capture cause-effect mechanisms enables causal AI to provide actionable, explainable insights beyond correlation-based models.

1. https://kanerika.com/blogs/causal-ai/
2. https://www.dynatrace.com/news/blog/causal-ai-use-cases-for-modern-observability/
3. https://www.dynatrace.com/news/blog/what-is-causal-ai-deterministic-ai/
4. https://www.a3logics.com/blog/causal-ai-use-cases/
5. https://www.leewayhertz.com/causal-ai/