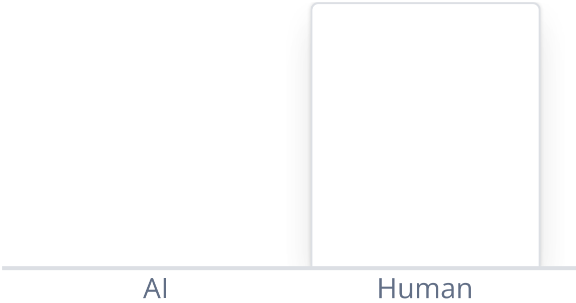
Results

0%

of text is likely AI



AI-generated 0%

AI-generated & AI-refined 0%

Human-written & AI-refined 0%

Human-written 100%

11.5 CASUAL REASONING:

Causal reasoning involves identifying cause-and-effect relationships, rather than just observing correlations, and trying to understand the mechanisms of how and why something happens. If A causes B, then when A happens B happens. Once we know the relationships, we can use them to predict, explain and intervene instead of simply observing.

How this is done in AI:

When we apply causal reasoning in AI, this can include:

Diagnosis and Trouble Shooting: Involves isolating a cause of a problem in a system. Planning and decision making: Involves predicting what could happen in the system given different actions, then taking the action that is predicted to change the state of the system.

Explanation generation: Involves providing reasons for why something happens.

Counterfactual reasoning: Involves generating "what if" questions to think about or reason about the impacts of different choices or events in the system. Scientific discovery: Involves looking for causal links in the data to understand the underlying mechanisms.

When causal reasoning is implemented in AI approaches, this typically means:

Causal models: Causal relations are represented as a graph, e.g., causal diagrams, Bayesian networks where nodes are variables and edges are causal links.

Intervention analysis: Simulate and see what happens when you intervene (i.e., do something) to the system and use this to test causal assumptions.

Counterfactual inference: Suppose some cause were changed or different and then reasoning about what would have happened.

Learning causal structure from data: Development of algorithms that can learn or infer causal relationships from observational or experimental data sets when it learning/inferring

causal relations (>often seen as a challenging area).
Example (Reasoning about the Cause of a Broken Box):

Let’s consider that the warehouse robot sees that a "fragile" box has been broken.
Observation: A "fragile" box has been broken.
Using causal reasoning, the robot would attempt to reason out the potential cause(s) of the breakage:
Hypothesis 1: Rough handling causes fragile boxes to break. The robot's internal sensors may tell it that it had handled the box roughly; therefore, it is a likely cause.
Hypothesis 2: A heavy object has fallen onto the box, causing it to break. The robot's external sensors may have detected a falling heavy object.
Hypothesis 3: The box was already damaged prior to being handled. The robot may inspect the box to verify any damage residual to the handling of the box.
The robot will use the knowledge it has regarding the world and possibly sensor data to reason about the potential causes of broke box, and determine what is the most likely cause. If for example, the robot reconfirms a model that is explicit about "Rough handling of fragile

items leads to breakage" and then finding out from its internal sensors that it really did handle the fragile box roughly, it can conclude that really the fact that it roughly handled the fragile box was indeed the cause of the broken box.

Use with LLMs (Large Language Models):

Large- language models currently have limitations in their reasoning capabilities about true causal reasoning. They are capable of determining correlations in text (e.g., "smoking is likely correlated with lung cancer"). However, LLMs are not yet able to causally reason about the statement in that they can’t independently assess the other object (in this case, smoking) against lung cancer to see if there is a plausible causality involved.

Answering "why" questions - The LLMs do have a (grounded) model of answering questions that relate to "causal" explanations. For example, if a text offers, "The king was angry, thus he banished the knight,” the LLM can answer "Why did the king banish the knight?" with "Because he was angry." again, this would be based on pattern matching (and the probability of the next word), not a structural view of causality.

Generating stories with succeeding events - LLMs generate stories, and these events logically follow, and in many cases, they imply a relationship of causality.

Understanding a limited class of causal language - LLMs often can understand certain causal terms, for example: "because", "hence", and "therefore" and "as a result."

But LLMs typically do not cope well with:

Distinguishing correlation from causation - LLMs may erroneously assume causation from a correlation in their training corpus.

Counterfactual reasoning - Asking operational "what if" questions can lead to complex causal dependencies for LLMs, which often leads to uniform or illogical reasoning.

Reasoning about interventions - Having the ability to predict the outcome of actively changing a part of a system based on a causal understanding is challenging for LLMs.

The field of incorporating true causal reasoning capabilities into LLMs is a developing research area.

Sample Real-World Application (Conceptually AI Context):

Imagine our warehouse robot is trying to understand why shelf c is always collapsing. Observation: when placing heavier objects on shelf c, it collapses. With respect to causal reasoning, the AI might perform the following:

Generate Hypotheses: - Heavy weight on shelf c causes it to collapse. - Shelf c was defective. - Nearby machinery creates vibrations that weaken shelf c under the heavy object.

Evidence Gathering: the robot could obtain data on the weight of the items placed on shelf c, look for structural damage of the shelf, monitor to see if nearby equipment creates excessive vibration.

Determining Cause: if the data indicated that heavy items were placed on shelf c with every instance of collapse, but there were no structural problems and the vibration data was within acceptable tolerances, the AI may reasonably conclude heavy items are the cause of the collapses on shelf c.

Identifying a course of action: from this causal knowledge, the AI could just enact the following rule: "No items greater than [weight limit] on shelf c", as a means to either limit, or ideally prevent, future collapses of shelf c.

Fuctionally the above works as follows:

Identifying possible causes: BRAINSTORMING BRIEFLY or retrieving possible variables that could reasonably explain the observed effect, if applicable

Validating evidence: comparing the potential possible causes against the available effects and observations, to see which possible causes appear to be supported.
Establishing causal associations: determining which of the factors were actually deemed to influence the outcome, usually with consideration of; temporal precedence (cause occurred before the effect), correlation, and ruled out potential alternative reasons.
Developing causal models: a structured representation of the cause-and-effectassociations determined.

Causal reasoning is an important process for AI to move past prediction and understanding of 'why' something happens, this develops better problem solving, planning, and decision making capabilities in complex environments.

11.6CHAIN-OF THOUGHT(COT) PROMPTING:

Chain-of-Thought (CoT) prompting is a technique used with large language models (LLMs) that encourages the model to explicitly show its reasoning process step-by step before arriving at a final answer. Instead of directly asking for the answer, the prompt is designed to elicit a sequence of intermediate thoughts that lead to the solution. This approach has been shown to significantly improve the performance of LLMs on complex reasoning tasks, such as arithmetic, common sense, and symbolic reasoning.

How it Works:

The idea behind CoT is to replicate human behavior when tackling complex problems. Humans typically approach complex problems by devising a sequence of easier problems in order to solve the more difficult problem and articulate their rationale for each step along the way.