

Level 3: Counterfactuals

What would Y be if X had been different?

Level 2: Interventions

What happens to Y under $\text{do}(X=x)$?

Level 1: Associations

What is $P(Y|X)$?

LLMs: strong at L1; brittle at L2/L3 without explicit causal grounding.