Decoding the true latent z a. Linear reg. Nonlinear reg. (MLP) Sinusoidal reg. Linear cls. Nonlinear cls. (MLP) 0.6 0.8 0.6 0.00100 0.004 0.6 0.00075 0.4 0.4 USW 0.00050 M 0.00025 0.4 0.002 0.2 0.2 0.2 0.00000 0.000 0.0 0.0 Raven's PM Gene targeting Rule accuracy ( † ) b. Counterfactual interventions Target accuracy ( 0.15 Alchemy Relative accuracy ( 1 0.10 1.0 0.05 0.5 0.00 0.00 0.0 Implicit Explicit Graph Potion map Stone map  $D \rightarrow$  $x_q \rightarrow$ 4 → y<sub>q</sub> Latent component  $\rightarrow y_a$ Transformer Transformer MLP Transformer Known