a. Decoding the true latent z

Linear reg. — Nonlinear reg. (MLP) — Sinusoidal reg. — Linear cls. — Nonlinear cls. (MLP)

W MSE ($\downarrow$)

Raven's PM — Gene targeting

Rule accuracy ($\uparrow$) — Target accuracy ($\uparrow$)
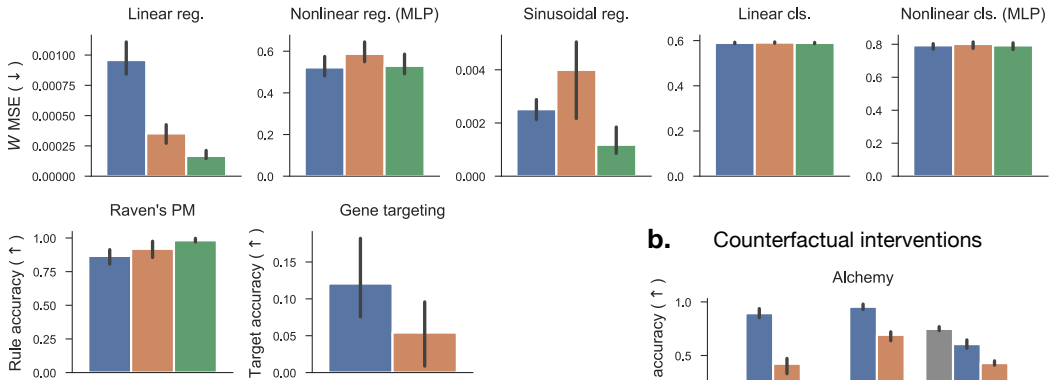
b. Counterfactual interventions

Alchemy

Relative accuracy ($\uparrow$)

Graph — Potion map — Stone map

Latent component

Implicit: $D \rightarrow$ Transformer, $x_q \rightarrow y_q$

Explicit: $D \rightarrow$ Transformer, $x_q \rightarrow z \rightarrow$ [Transformer | MLP | Known] $\rightarrow y_q$