Curious Causality-Seeking Agents in Open-ended Worlds

Zhiyu Zhao[1], Haoxuan Li[2], Haifeng Zhang[3,4], Jun Wang[5], Francesco Faccio[6,7], Jürgen Schmidhuber[6,7], Mengyue Yang[1*]

NeurIPS 2025

[1]University of Bristol [2]Peking University
[3]Institute of Automation, Chinese Academy of Sciences
[4]School of Artificial Intelligence, Chinese Academy of Sciences
[5]University College London
[6]King Abdullah University of Science and Technology (KAUST)
[7]The Swiss AI Lab, IDSIA-USI/SUPSI

Background: World Models

- World models are critical for reinforcement learning agents to simulate, plan, and predict outcomes

Background: World Models

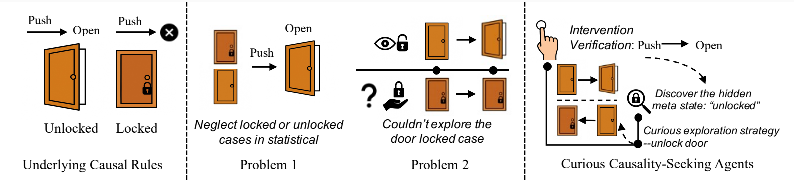
- World models are critical for reinforcement learning agents to simulate, plan, and predict outcomes
- Causal world models offer robustness by learning the underlying data-generation process

Background: World Models

- World models are critical for reinforcement learning agents to simulate, plan, and predict outcomes
- Causal world models offer robustness by learning the underlying data-generation process
- However, existing causal world models face challenges in the open-ended worlds

Motivation: The Challenge of Open-ended Worlds

- Learning a **single uniform** causal graph ✗
  - Neglecting these context-dependent changes
  - Example: **"Push → Open"** works when *unlocked* but fails when *locked*
  - Meta-Causal Graph (MCG): a unified structure that captures how causal relationships evolve across different states ✓
- Domain modeling requires a priori knowledge of state labels, limiting generalization to novel contexts ✗
  - Curious Causality-Seeking Agent: actively explore interventions to uncover MCGs ✓



Meta-Causal Graph

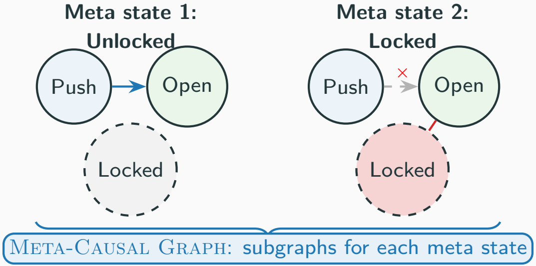
- **Definition:** An MCG is a minimal representation that contains multiple causal subgraphs, $\{\mathbf{G}_1, \mathbf{G}_2, \ldots, \mathbf{G}_k\}$.

Meta-Causal Graph

- **Definition:** An MCG is a minimal representation that contains multiple causal subgraphs, $\{\mathbf{G}_1, \mathbf{G}_2, \ldots, \mathbf{G}_k\}$.
- Each subgraph $\mathbf{G}_i$ corresponds to a unique **meta state** $\mathbf{m}_i$.

Meta-Causal Graph

- **Definition:** An MCG is a minimal representation that contains multiple causal subgraphs, $\{\mathbf{G}_1, \mathbf{G}_2, \ldots, \mathbf{G}_k\}$.
- Each subgraph $\mathbf{G}_i$ corresponds to a unique **meta state** $\mathbf{m}_i$.
- The MCG models the world as a **switching mechanism**:
  - The current meta state $\mathbf{m}$ dictates which causal subgraph $\mathbf{G}_i$ is active.
  - The active $\mathbf{G}_i$ then governs the transition dynamics: $P(\mathbf{s}_{t+1}|\mathbf{s}_t, \mathbf{a}_t)$.



Identifiability: Can we find the MCG?

Identifiability: Can we find the MCG?

Challenge: The meta states (u) and subgraphs ($\mathcal{G}_u$) are both latent.

Identifiability: Can we find the MCG?

Challenge: The meta states (u) and subgraphs ($\mathcal{G}_u$) are both latent.

We prove identifiability in two steps:

1. **Meta States ($u = C(x)$):**

Identifiability: Can we find the MCG?

Challenge: The meta states (u) and subgraphs $(\mathcal{G}_u)$ are both latent.

We prove identifiability in two steps:

1. **Meta States ($u = C(x)$):**
    - Assumption 1 (Mixed Data): Learning from **mixed data** (all samples pooled) recovers the **union** of all causal parents. $Pa_{\hat{\mathcal{G}}}(X_j) = \bigcup_{i \in S_{\mathcal{D}}} Pa_{\mathcal{G}_i}(X_j)$
    - Theorem 1 & 2: Under this assumption, meta states are identifiable (up to permutation), even if we overparameterize (use too many clusters).

Identifiability: Can we find the MCG?

Challenge: The meta states (u) and subgraphs ($\mathcal{G}_u$) are both latent.

We prove identifiability in two steps:

1. **Meta States ($u = C(x)$):**
   - Assumption 1 (Mixed Data): Learning from **mixed data** (all samples pooled) recovers the **union** of all causal parents. $Pa_{\hat{\mathcal{G}}}(X_j) = \bigcup_{i \in S_{\mathcal{D}}} Pa_{\mathcal{G}_i}(X_j)$
   - Theorem 1 & 2: Under this assumption, meta states are identifiable (up to permutation), even if we overparameterize (use too many clusters).
2. **Causal Subgraphs ($\mathcal{G}_u$):**

Identifiability: Can we find the MCG?

Challenge: The meta states (u) and subgraphs ($\mathcal{G}_u$) are both latent.

We prove identifiability in two steps:

1. **Meta States ($u = C(x)$):**
   - Assumption 1 (Mixed Data): Learning from **mixed data** (all samples pooled) recovers the **union** of all causal parents. $Pa_{\hat{\mathcal{G}}}(X_j) = \bigcup_{i \in S_{\mathcal{D}}} Pa_{\mathcal{G}_i}(X_j)$
   - Theorem 1 & 2: Under this assumption, meta states are identifiable (up to permutation), even if we overparameterize (use too many clusters).
2. **Causal Subgraphs ($\mathcal{G}_u$):**
   - Once meta states are separated, we need **interventions** to find edge direction.
   - Proposition 1: The subgraph is identifiable if we can intervene on single variables of an edge (e.g., $do(a)$ to check $a \rightarrow b$).

Framework: Curious Causality-Seeking Agent

We propose a framework to learn the MCG in a continual loop, which alternates between three core components:

Framework: Curious Causality-Seeking Agent

We propose a framework to learn the MCG in a continual loop, which alternates between three core components:

1. **Curiosity-Driven Interventional Exploration**
   - The agent uses an intrinsic reward (e.g., edge-entropy) to actively select interventions that probe uncertain causal relationships.
2. **Meta-Causal Graph Discovery**
   - From the collected experience, the agent jointly learns the mapping to latent meta states ($u = C(x)$) and their corresponding causal subgraphs ($\mathcal{G}_u$).
3. **Continual World Model Learning**
   - The agent uses the discovered $\mathrm{MCG}$to learn the environment's transition dynamics and continually updates its models and policy.

Core Component 1: Curiosity-Driven Exploration

Key Insight: Passive observation $\rightarrow$ incomplete causal graphs

Solution: Active interventions guided by curiosity

Curiosity Reward: Edge-Entropy

$$\mathcal{I}_t^{\text{edge}} = \sum_{i,j} H(\hat{M}_{C(X_t)}[i,j])$$

- Prioritizes most uncertain causal edges
- Systematically probes unknown relationships
- Reveals complete causal structure

Takeaway

Active interventions discover what passive observation cannot

Core Component 2: Meta-Causal Graph Discovery

Solution: VQ-VAE discretization

- Encode state $\rightarrow$ discrete meta state u $\rightarrow$ causal subgraph $\mathcal{G}_u$
- Automatic discovery, no labels needed

Adaptive Codebook Fusion:

- Automatically merge similar causal subgraph to maintain minimal representation

Takeaway

Hierarchical discretization discovers interpretable causal regimes

Core Component 3: Integrated Learning Objective

End-to-End Training with Multi-Task Loss:

Complete Objective Function

$$\mathcal{L} = \mathcal{L}_{\text{MLE}} + \lambda_{\text{sparse}}\mathcal{L}_{\text{sparse}} + \lambda_{\text{mask}}\mathcal{L}_{\text{mask}} + \lambda_{\text{quantization}}\mathcal{L}_{\text{quantization}}$$

$\mathcal{L}_{\textbf{MLE}}$

Predict next state

Learn dynamics

$\mathcal{L}_{\textbf{sparse}}$

Simple graphs

L1 regularization

$\mathcal{L}_{\textbf{mask}}$

Verify interventions

Refine graph

Takeaway

Multi-objective learning combines causal discovery with world model accuracy

Experimental Setup

Comprehensive Evaluation in Challenging Environments

Environments

- **Chemical:** Chain & Fork structures
- **Magnetic:** Complex physical dynamics
- Increasing noise levels (n=2,4,6)
- Out-of-distribution testing

Baselines

- **GNN/MLP:** Standard architectures
- **NCD/FCDL:** Causal discovery methods
- **Modular/NPS:** Modular approaches
- **Transformer:** Sequence modeling

Evaluation Metrics

- **Prediction Accuracy:** OOD generalization
- **Downstream Rewards:** Task performance
- **Robustness:** Varying noise levels
- **Interpretability:** Causal pattern analysis

Results: Out-of-Distribution Prediction

- **MCG** achieves highest accuracy across noise levels
- Outperforms GNN, NCD, FCDL, Transformer baselines
- Robust to distribution shifts

OOD Prediction Accuracy

Algorithm	Fork			Chain		
	n=2	n=4	n=6	n=2	n=4	n=6
GNN	36.29±3.45	25.80±3.48	21.58±3.44	29.22±3.39	23.28±4.98	20.53±6.96
MLP	31.11±1.69	30.44±2.28	32.39±1.76	28.66±3.65	26.52±4.26	24.15±4.17
NCD	41.60±5.08	37.47±2.13	42.27±1.82	40.04±6.21	37.47±2.98	41.19±1.66
FCDL	57.82±9.90	49.29±8.90	47.70±6.68	50.66±10.10	48.81±8.91	48.05±5.86
Modular	26.53±3.45	24.73±5.61	26.73±8.31	25.24±4.68	24.94±4.81	25.09±5.91
NPS	40.56±4.61	26.81±4.37	23.02±4.27	38.73±2.63	27.69±4.28	24.45±3.84
CDL	35.59±1.85	35.82±1.40	42.22±1.39	34.90±1.59	36.52±1.72	42.06±1.29
GRADER	37.93±1.06	38.94±1.63	45.74±2.25	36.82±3.12	37.41±2.84	43.48±4.14
Oracle	33.87±1.34	36.48±1.80	42.47±0.75	34.63±1.78	38.31±2.48	42.87±2.08
Sandy-Mixure	31.93±0.16	32.47±0.0161	33.72±2.11	30.08±3.12	29.31±5.18	27.43±2.20
Transformer	25.13±0.63	24.37±2.58	21.90±2.35	29.62±0.65	30.37±2.81	29.78±1.08
MCG (ours)	**63.18±13.94**	**50.47±9.87**	**50.04±8.56**	**51.99±6.58**	49.78±4.11	**49.69±5.14**

Results: Downstream Task Performance

- **MCG** yields top downstream rewards
- Strong generalization to unseen contexts
- Effective in both Chemical and Magnetic environments

Downstream Rewards

Algorithm	Chain			Fork			Magnetic
	n=2	n=4	n=6	n=2	n=4	n=6	
GNN	6.89±0.28	6.38±0.28	6.56±0.53	6.61±0.92	6.15±0.74	6.95±0.78	2.23 ± 0.90
MLP	7.39±0.65	6.63±0.58	6.78±0.93	6.49±0.48	5.93±0.71	6.84±1.17	2.10 ± 0.22
NCD	9.60±1.52	8.86±0.23	10.32±0.37	10.95±1.63	9.11±0.63	9.11±0.63	2.85 ± 0.47
FCDL	11.16±3.5	10.39±2.84	10.62±2.52	13.98±2.01	13.36±2.14	12.91±2.40	2.77 ± 0.45
Modular	6.61±0.63	7.01±0.55	7.04±1.07	6.05±0.70	5.65±0.50	6.43±1.00	0.88 ± 0.52
NPS	6.92±1.03	6.88±0.79	6.80±0.39	5.82±0.83	5.75±0.57	5.54±0.80	0.91 ± 0.69
CDL	8.71±0.55	8.65±0.38	10.23±0.50	9.37±1.33	8.23±0.40	9.50±1.18	1.10 ± 0.67
Oracle	8.47±0.69	8.85±0.78	10.29±0.37	7.83±0.87	8.04±0.62	9.66±0.21	0.95 ± 0.55
Sandy-Mixture	6.81±0.17	6.73±0.20	7.07±0.26	6.95±0.21	6.71±0.20	7.03±0.40	1.63 ± 0.02
Transformer	6.45±0.28	6.73±0.18	7.31±0.33	6.54±0.18	6.68±0.27	7.00±0.39	2.13 ± 0.01
MCG(ours)	**13.82±3.84**	**12.49±2.39**	**12.45±1.37**	**14.65±2.75**	**14.06±2.64**	**13.28±2.04**	**3.19 ± 0.14**

Ablation Studies

- **No intervention verification** → weaker causal alignment
- **No curiosity-driven exploration** → lower exploration efficiency & accuracy
- **Both components** critical for robust performance

Ablation Results

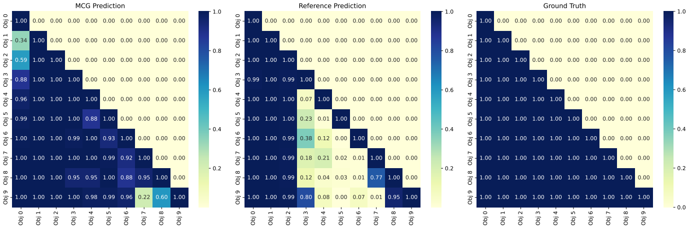
Algorithm	Fork			Chain		
	n=2	n=4	n=6	n=2	n=4	n=6
MCG (wo intervention verification)	58.18±14.51	51.70±8.58	46.04±8.56	47.75±7.52	46.19±5.87	47.94±6.60
MCG (wo curiosity-driven exploration)	48.28±8.15	43.48±5.05	46.54±3.72	50.36±8.05	**50.55±6.83**	48.72±4.42
MCG (ours)	**63.18±13.94**	**50.47±9.87**	**50.04±8.56**	**51.99±6.58**	49.78±4.11	**49.69±5.14**

Analysis & Visualization

Learned Subgraphs

- Closely match ground truth

Learned vs. Ground Truth Graphs



Conclusion

Thank you for your listening!

Paper Link