



Foundation Models-01

Ruibo Tu, 2024 Sep. 23rd

Ruibo's blog for the talk



Foundation Models

“Gigantic pre-trained models
Trained on enormous data
Used in a few-/zero-shot manner”

Next token prediction

Language

- GPT-4
- Gemini2
- LLaMA3
- ...

Masked prediction

Vision

- SAM
- MAE
- Dino2
- ...

Bi-modal

- GPT-4V
- CLIP
- ...

Limitations of FMs

- **Not world models:** (Mechanisms, states)
 - Observational data (no interventional data, experimentation)
 - Hallucination: confidently wrong results, sounds well but wrong
- **GPT limitation:**
 - Cannot modify autoregressively generated results (OpenAI-o1)
 - Small working memory (CoT: More computation for relaxing the constraint)
- **Challenges:**
 1. **Reasoning:** No aware of structural/causal dependencies among factors
 2. **Planning:** Incremental planning✅, no hierarchical / long-term planning❌

“Predicting consequences of actions requires to capturing causal relationships correctly”

Unknown: FMs.

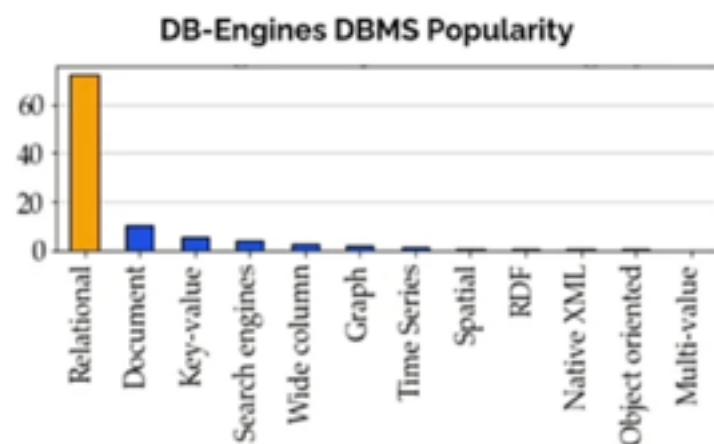
“Many tasks are based on table understanding”

Under-explored: FMs.

Data are very likely in the form of tables

Tables are **everywhere**

Spreadsheets, documents, web (pages), databases, ...



Goal: Building a foundation model for data scientists and ML engineers

Tabular data

	C1	C2	C3
R1		1.68	
R2	Healthy	1.75	Very Satisfied
R3	NULL	1.98	
R4		2.01	

	C1	C2	C3
R5	Sick	1.56	NULL
R6		1.75	Less Satisfied
R7		1.88	

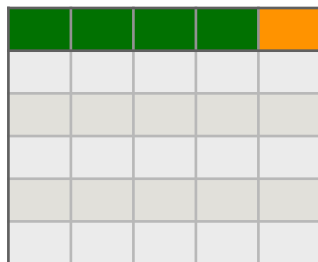
	C2	C3	C4
R1	1.68		
R2	1.75	Very Satisfied	Male
R3	1.98		
R4	2.01		

	C2	C5	C6
R1	1.68		
R2	1.75	9.9	Rare
R3	1.98		
R4	NULL	NULL	NULL
R5	NULL	NULL	NULL
R6	NULL	NULL	NULL
R7	1.88	10.1	Medium

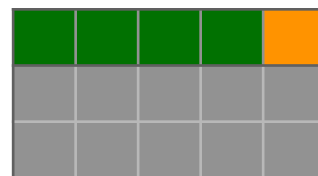
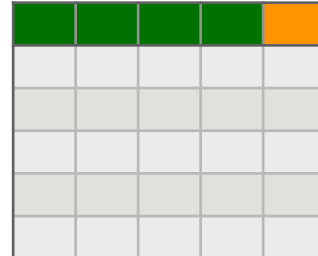
- Missing data imputation
- Feature selection
- Root cause analysis
- A/B testing
- Anomaly detection
- ML-based Predictive analysis
- Causal discovery
- Potential outcome prediction

SOTA Tabular Tasks

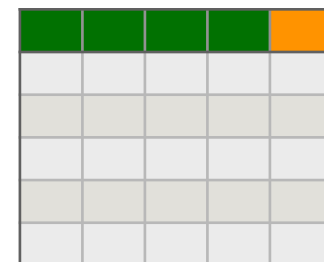
Predictive



Generative



Understanding



+ Q

A: ...

SOTA Tabular Methods

Encoder-Decoder: Generic representation learning

- Self-supervised learning, contrastive learning
- BERT

Only-Decoder: Generative tasks as the general objective

- Predict next token
- GPT
- Diffusion models

Challenges

	C1	C2	C3
R1		1.68	
R2	Healthy	1.75	Very Satisfied
R3	NULL	1.98	
R4		2.01	

	C1	C2	C3
R1		3.01	
R2	Sick	2.22	Satisfied
R3	NULL	2.01	
R4		2.91	

	C2	C3	C4
R1	1.68		
R2	1.75	Very Satisfied	Male
R3	1.98		
R4	2.01		

1. Practical challenges

- Transformer structure
- Mixed data type
- Missing values

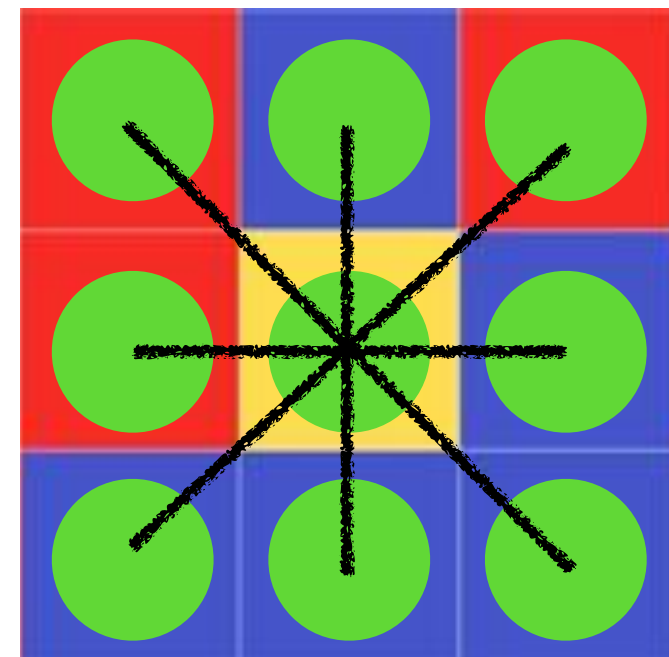
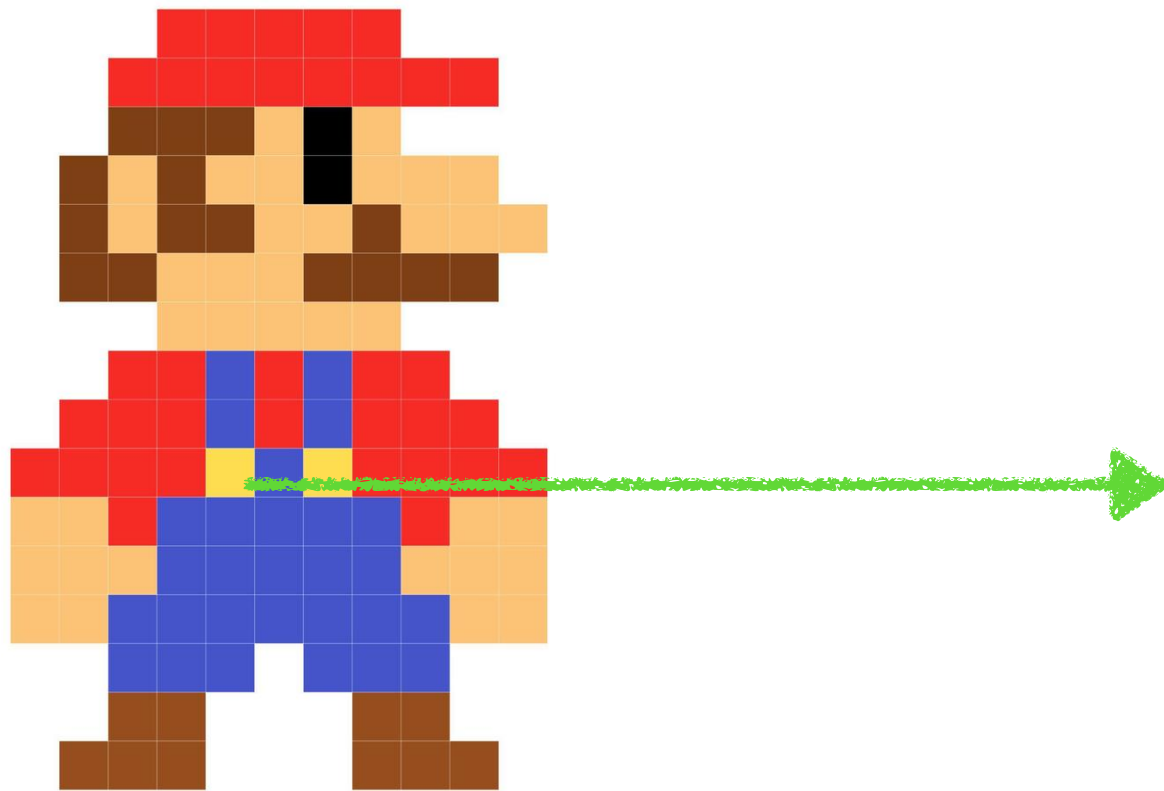
2. Structural dependencies

- Prior: Causal relationships
- Domain / Distribution shift
- Cross-table reasoning

Open problems

**“How to enable FMs to
capture (causal) structural information”**

Order of an image





Order of an sentence



Order of a table

Col. order

Row. order



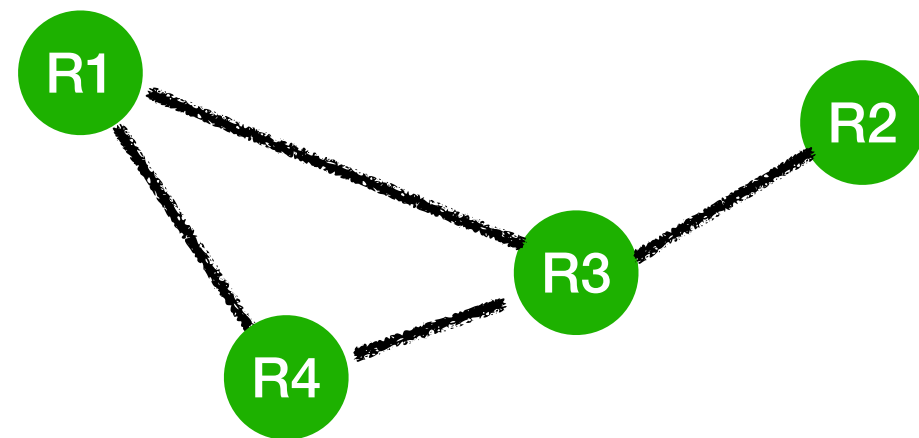
	C1	C2	C3
R1		1.68	
R2	Healthy	1.75	Very Satisfied
R3	NULL	1.98	
R4		2.01	

	C1	C3	C2
R1			1.68
R2	Healthy	Very Satisfied	1.75
R3	NULL		1.98
R4			2.01

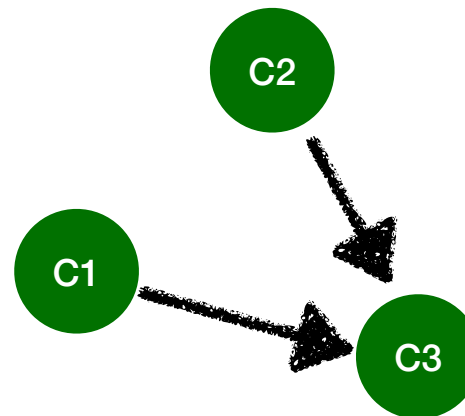
	C1	C2	C3
R1		1.68	
R2	Healthy	1.75	Very Satisfied
R4		2.01	
R3	NULL	1.98	

The same table

Graph + Table



Social-Network graph

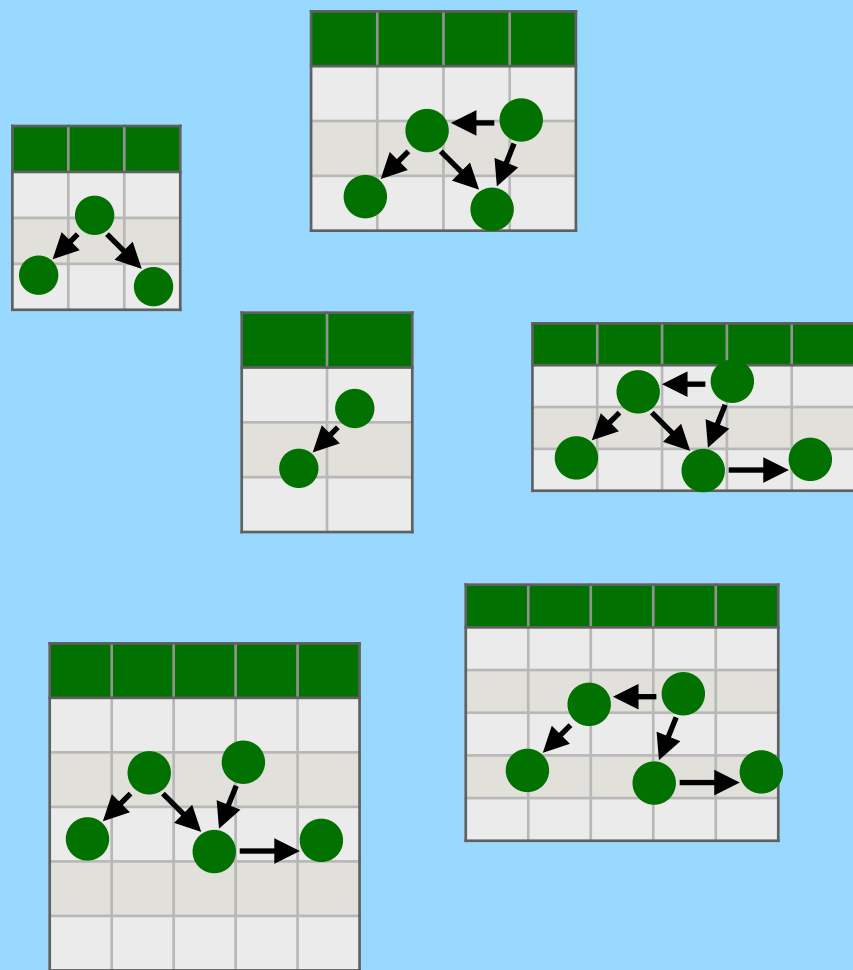


Causal graph

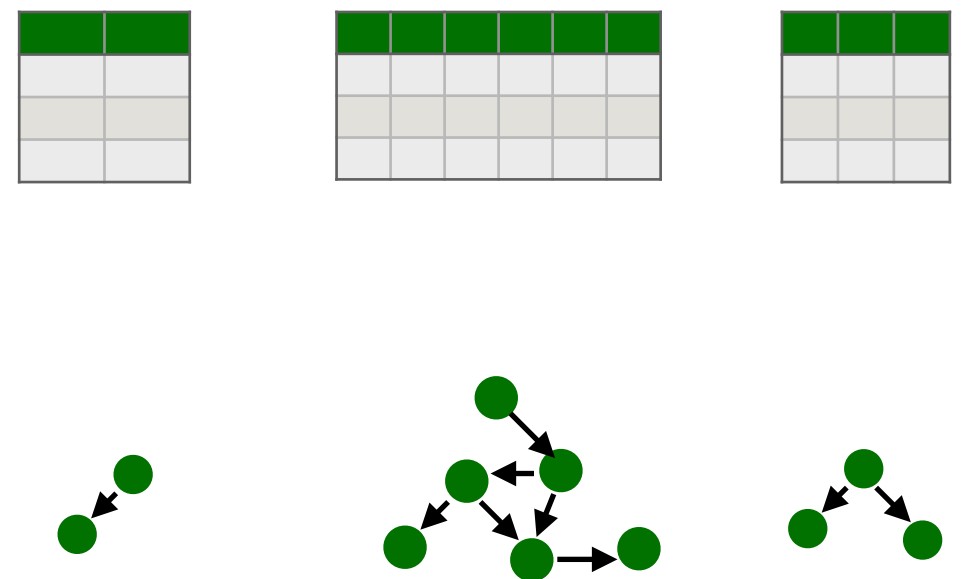
	C1	C2	C3
R1		1.68	
R2	Healthy	1.75	Very Satisfied
R3	NULL	1.98	
R4		2.01	

(Causal) Structure-Aware Foundation Models

Data



Application



**“Our benchmarks
For LLMs and other generative models”**

LLM Benchmarks

1. Common Benchmarks: well regularized
 - Conversational, arithmetic math, competition coding tests
2. Tuned for a better number
 - Skills in real-world applications?

A benchmark for causal reasoning and decision making

Causal reasoning ability is the **core** for Reasoning and Decision-making

- Numerical and mathematical reasoning
- Applications:
 - Agent, Healthcare, Education ...

Work 1: A benchmark for causal reasoning and decision-making

Tasks

Benchmark data:

Causal Graphs

Tables

V1	V2	V3	V4
0.41	0.2	0.21	0.63
1.21	0.6	0.62	1.81

Questions about causal statements:

Given causal graphs, ask questions about

Adjacency matrix
D-separations
Causal directions

Given tables, ask questions about

Adjacency matrix
D-separations
Causal directions

Given graphs and tables, ask questions about

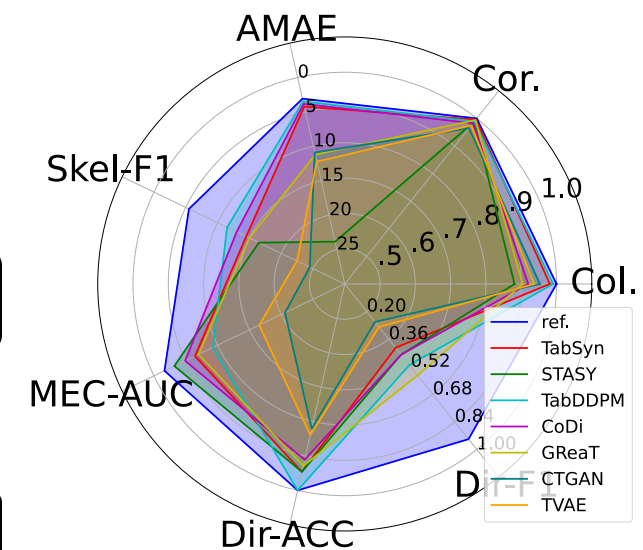
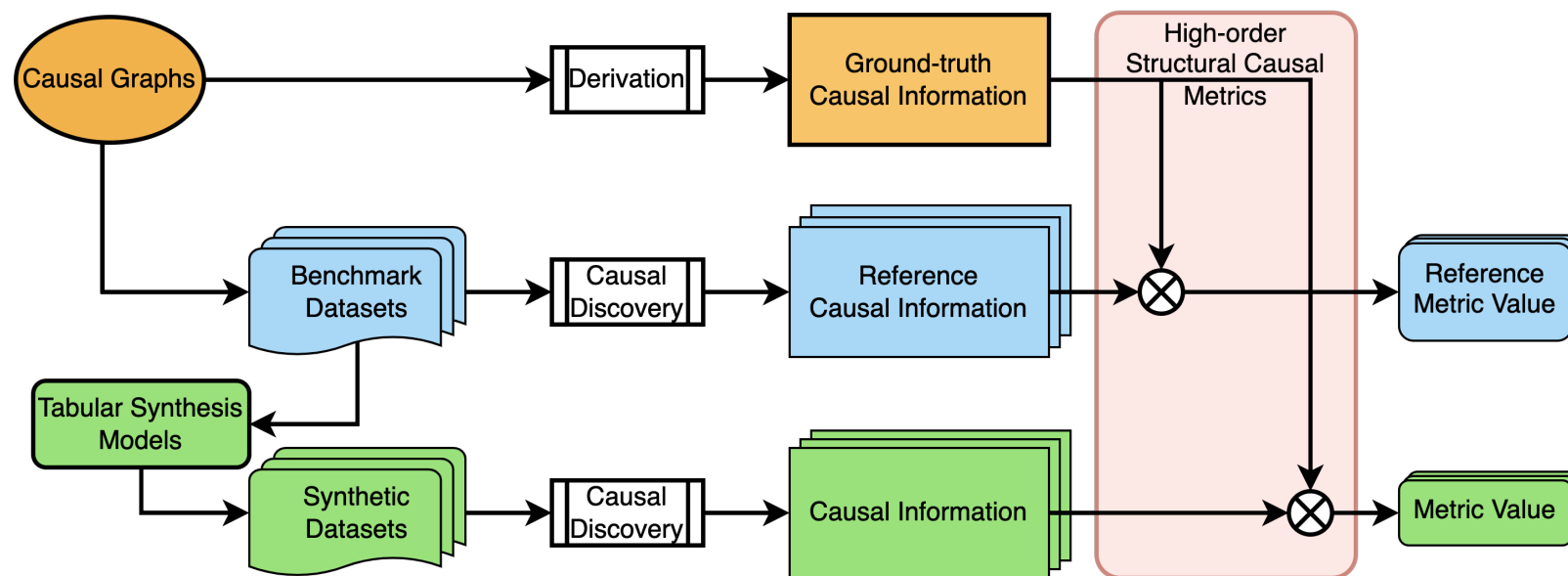
Interventional distributions
Counterfactual distributions

Answers:

- The adjacency matrix is ...;
- V2 is d-separated from V3 given V1;
- V2 --> V4;
- $p(V_j | do(V_i)=x) = \dots$;
- $p(V_j | do(V_i)=x, V1=a, V2=b, \dots)$

Causal graph reasoning	Causal graphs -> Adjacency matrix	Causal graphs -> D-separations	Causal graphs -> Causal directions
Knowledge discovery	<p>Given a graph among V1, V2, V3, and V4. Graph has the following directed edges: V1 -> V2, V1 -> V3, V2 --> V4, V3 --> V4.</p> <p>Question: What is the adjacency matrix ? or What are the neighbor nodes of V2?</p>	<p>Given a graph among V1, V2, V3, and V4. Graph has the following directed edges: V1 -> V2, V1 -> V3, V2 --> V4, V3 --> V4.</p> <p>Question: Is V2 d-separated from V3 given V1? Are V2 and V3 the parents of V4.</p>	<p>Given a graph among V1, V2, V3, and V4. Graph has the following directed edges: V1 -> V2, V1 -> V3, V2 --> V4, V3 --> V4.</p> <p>Question: Does V2 cause V1?</p>
	<p style="text-align: center;">Tables -> Adjacency matrix</p> <p>The corresponding tabular data is V1 V2 V3 V4 \n ----- ---- ----- \n 0.41 0.2 0.21 0.63 \n 1.21 0.6 0.62 1.81 .</p> <p>Question: What is the adjacency matrix ?</p>	<p style="text-align: center;">Tables -> D-separations</p> <p>The corresponding tabular data is V1 V2 V3 V4 \n ----- ---- ----- \n 0.41 0.2 0.21 0.63 \n 1.21 0.6 0.62 1.81 .</p> <p>Question: Is V2 d-separated from V3 given V1?</p>	<p style="text-align: center;">Tables -> Causal directions</p> <p>The corresponding tabular data is V1 V2 V3 V4 \n ----- ---- ----- \n 0.41 0.2 0.21 0.63 \n 1.21 0.6 0.62 1.81 .</p> <p>Question: Does V2 cause V1?</p>
	<p>Causal graphs + Tables -> Counterfactual distributions</p> <p>Given a graph among V1, V2, V3, and V4. Graph has the following directed edges: V1 -> V2, V1 -> V3, V2 --> V4, V3 --> V4.</p> <p>The corresponding tabular data is V1 V2 V3 V4 \n ----- ---- ----- \n 0.41 0.2 0.21 0.63 \n 1.21 0.6 0.62 1.81 .</p> <p>Question: Given a new observation, V1= 0.1, V2 = 0.3, V3=4, V4=0.1, V5= 0.2, what is the counterfactual distribution of all variables when the intervention value of V2 is 0.3.</p>		

Work 2: A benchmark for evaluating tabular synthesis model capability for preserving causal information



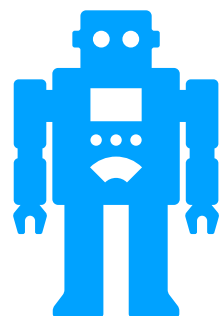
- Diffusion-based
- LLM-based

Part 2

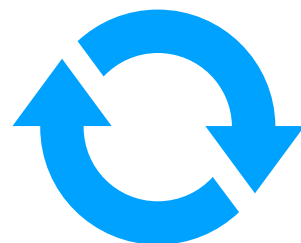
“Decision-making of FMs (agents)”

Agents (RL)

Agents



Interactions



Environments



Dialogue Agent

QA, conversation

Human

Game Agent

APIs

Text, Video Games

Goal: Robust agents learn causal world models

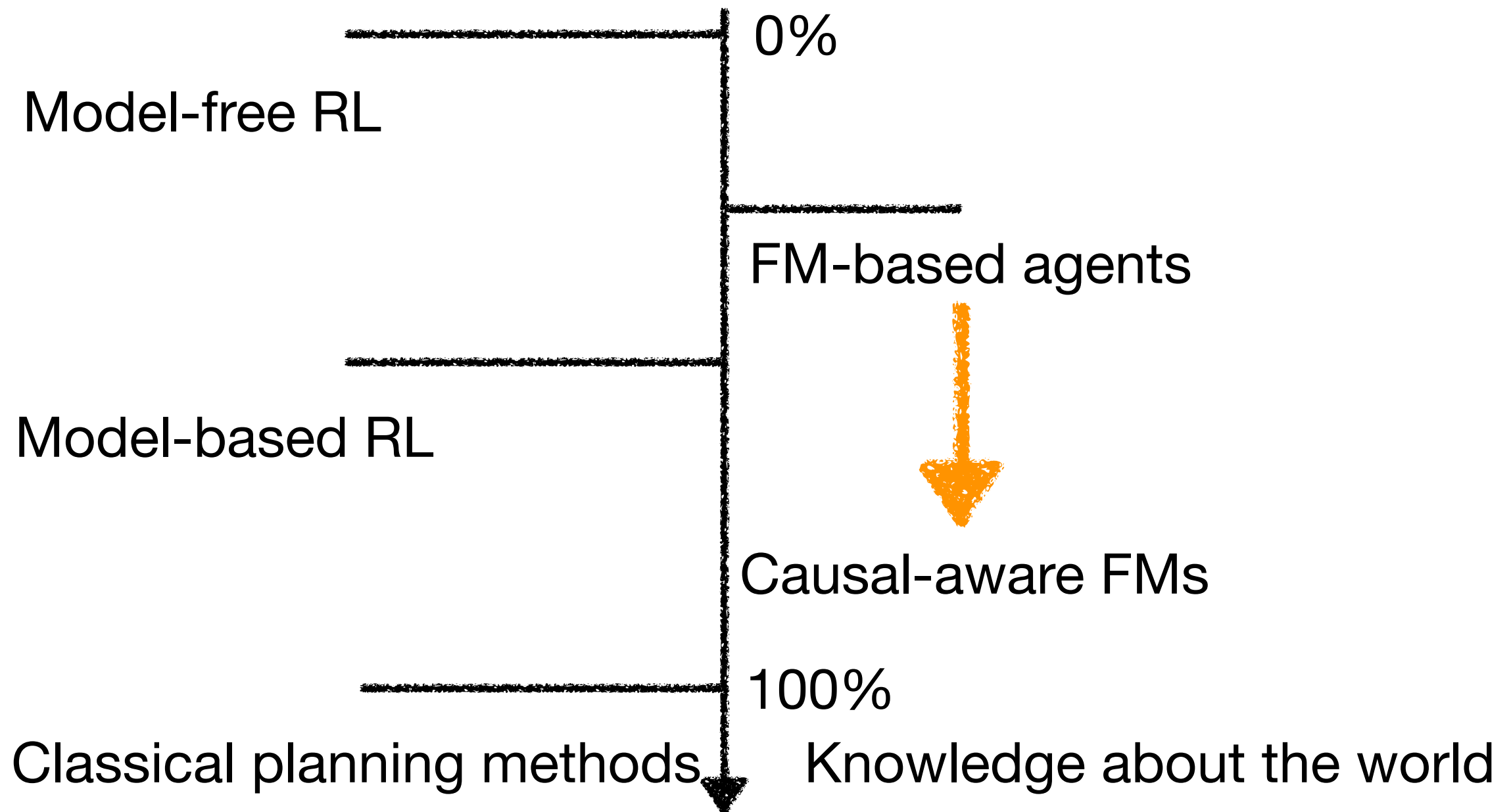
“Predict consequences of actions”

Minimum requirement for generalization

- Efficiency: Sample quality and efficiency
- Effectiveness: generalization

“Only interacting with the world and causal discovery can get causal information.”

Knowledge about the world



Agents with a causal world model

Model-based RL

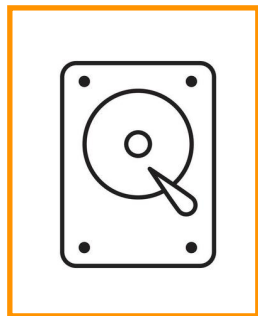


Causal (SCM)-based

Causal-aware FM agents

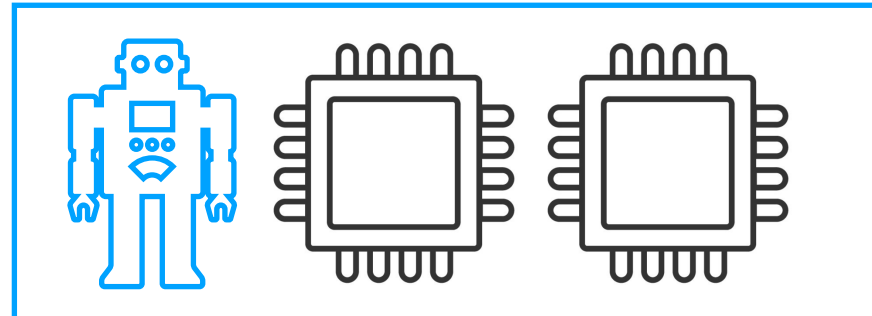
World Model

- graphs



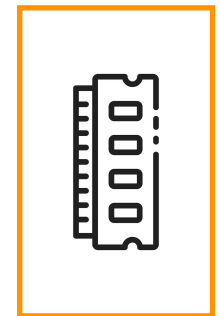
Algorithms,

- Discovery, Inference



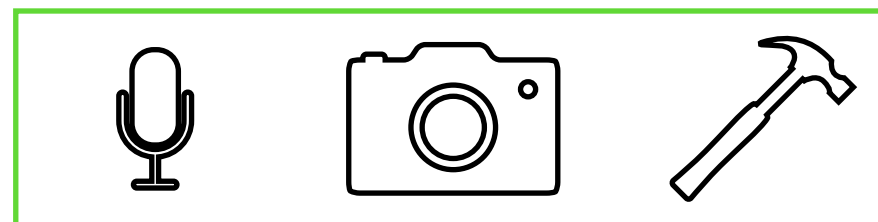
Task-related information

- Parameters



Communication

Interaction



Non-causal FM agents

1. Augmenting memories and hard drives
2. Iterative prompts with in/external feedbacks
3. Combine with planning methods
4. On/offline fine-tuning

Wang, G., Xie, Y., Jiang, Y., Mandlekar, A., Xiao, C., Zhu, Y., ... & Anandkumar, A. (2023). Voyager: An open-ended embodied agent with large language models. arXiv preprint arXiv:2305.16291.

Li, Shuang, Xavier Puig, Chris Paxton, Yilun Du, Clinton Wang, Linxi Fan, Tao Chen et al. "Pre-trained language models for interactive decision-making." Advances in Neural Information Processing Systems 35 (2022): 31199-31212.

Yoneda, T., Fang, J., Li, P., Zhang, H., Jiang, T., Lin, S., Picker, B., Yunis, D., Mei, H. and Walter, M.R., 2023. Statler: State-maintaining language models for embodied reasoning. arXiv preprint arXiv:2306.17840.

Hao, S., Gu, Y., Ma, H., Hong, J.J., Wang, Z., Wang, D.Z. and Hu, Z., 2023. Reasoning with language model is planning with world model. arXiv preprint arXiv:2305.14992.

Pouplin, T., Sun, H., Holt, S. and Van der Schaar, M., 2024. Retrieval-Augmented Thought Process as Sequential Decision Making. arXiv preprint arXiv:2402.07812.

**“Causal-aware FMs for
Reasoning and decision-making”**

**Thank you for
listening!**

