TECHNISCHE UNIVERSITÄT DARMSTADT

AIML Lab

Winter Semester 2025/26 Lecture

# Causality for AI & ML

*"Meta-Causal Models"*

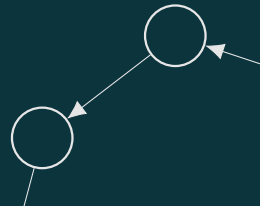

Prof. Dr. Kristian Kersting

Moritz Willig — Today's speaker

Tim Woydt

Florian Busch

Matej Zečević

# Causal AI

*“Machines’ lack of understanding of causal relations is perhaps the biggest roadblock to giving them human-level intelligence.”*

- Judea Pearl and Dana Mackenzie,
The Book of Why.

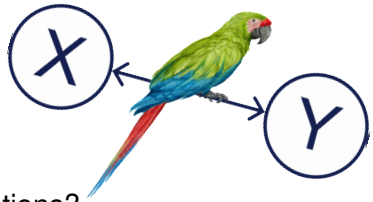
# Causal AI

*"Machines' lack of understanding of causal relations is perhaps the biggest roadblock to giving them human-level intelligence."*

- Judea Pearl and Dana Mackenzie,
The Book of Why.

This should include not only reasoning within static causal structures, but also thinking about the emergence of these causal relations in the first place.

# Recap: Are LLMs Causal Parrots?

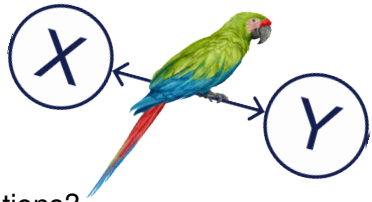



**Genuine Understanding:** LLMs might say “Smoking causes Cancer”, because they have seen that sentence thousands of times during training

. . . but do LLMs really ‘*understand*’ the real-world implications?

Zečević*, M., Willig*, M., Dhami, D.S. and Kersting, K., 2023. Causal Parrots: Large Language Models May Talk Causality But Are Not Causal. Transactions on Machine Learning Research.

# Recap: Are LLMs Causal Parrots?



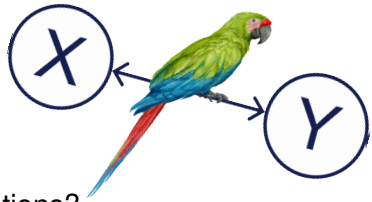

**Genuine Understanding:** LLMs might say “Smoking causes Cancer”, because they have seen that sentence thousands of times during training

. . . but do LLMs really ‘*understand*’ the real-world implications?

$\rightarrow$ Under which conditions do these relationships hold up?

Zečević*, M., Willig*, M., Dhami, D.S. and Kersting, K., 2023. Causal Parrots: Large Language Models May Talk Causality But Are Not Causal. Transactions on Machine Learning Research.

# Recap: Are LLMs Causal Parrots?




**Genuine Understanding:** LLMs might say “Smoking causes Cancer”, because they have seen that sentence thousands of times during training

. . . but do LLMs really ‘*understand*’ the real-world implications?

→ Under which conditions do these relationships hold up?

→ *“Do LLM really understand the conditions of causal emergence?”*

Zečević*, M., Willig*, M., Dhami, D.S. and Kersting, K., 2023. Causal Parrots: Large Language Models May Talk Causality But Are Not Causal. Transactions on Machine Learning Research.

# Recap: Reflection and Causal Change

**Reflection:** Causal reasoning with do-calculus commonly assumes a static SCM.

$\rightarrow$ Changes in causal structures are mostly due to 'external' interventions.

$\rightarrow$ *"Are SCM suited to model causal change?"*

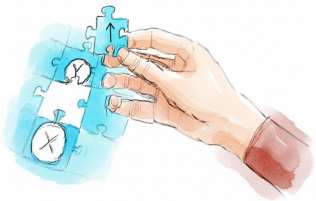
# Recap: Reflection and Causal Change

**Reflection:** Causal reasoning with do-calculus commonly assumes a static SCM.

→ Changes in causal structures are mostly due to ‘external’ interventions.

→ *“Are SCM suited to model causal change?”*

Reasoning *about* causal change might be considered a meta-task to classical causal inference.



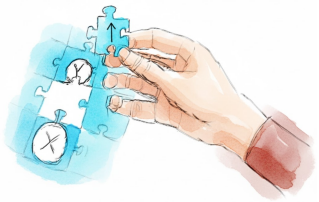
# Recap: Reflection and Causal Change

**Reflection:** Causal reasoning with do-calculus commonly assumes a static SCM.

→ Changes in causal structures are mostly due to ‘external’ interventions.

→ *“Are SCM suited to model causal change?”*

Reasoning *about* causal change might be considered a meta-task to classical causal inference.

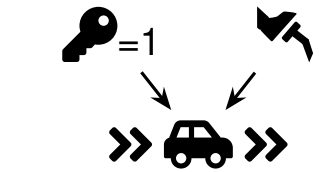


→ Objective: ‘Model how causal graphs evolve under dynamic environments.’

→ Proposed solution: *Meta-Causal Models*

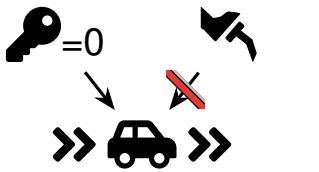
# Contextual Independence: Driving a Car

Variables can become *contextually independent* under certain states of a system.




Key → Speed ← Pedal

$Pedal \not\perp\!\!\!\perp Speed$



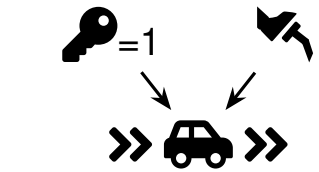

Key → Speed  Pedal

$Pedal \perp\!\!\!\perp Speed$

$$\rightarrow Pedal \perp\!\!\!\perp Speed \mid Key = 0$$

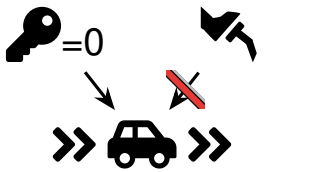
# Contextual Independence: Driving a Car

Variables can become *contextually independent* under certain states of a system.




Key → Speed ← Pedal

$Pedal \not\perp\!\!\!\perp Speed$




Key → Speed  Pedal

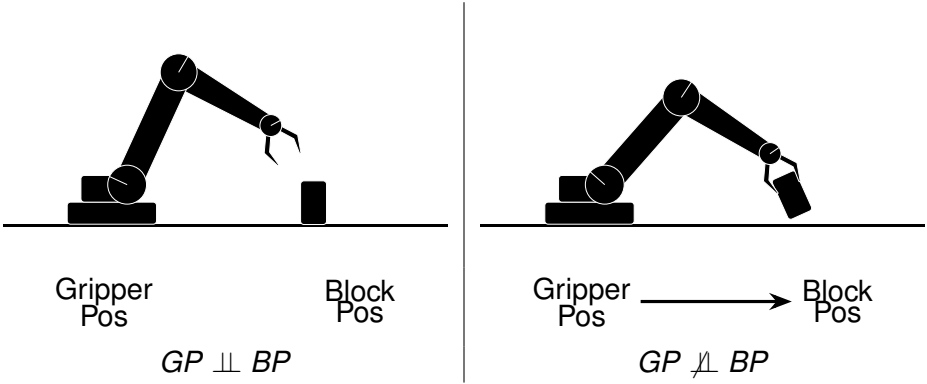
$Pedal \perp\!\!\!\perp Speed$

$$\rightarrow Pedal \perp\!\!\!\perp Speed \mid Key = 0$$

**Contextual Independency**: generally written as $A \perp\!\!\!\perp B | X = x$
(usually with some $x'$, such that $A \not\perp\!\!\!\perp B | X = x'$).

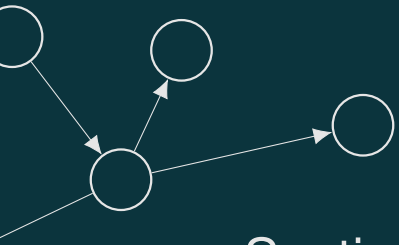
# Contextual Independence: Robot Arm

(In)dependencies can materialize under more complex conditions:




$$\rightarrow GP \perp\!\!\!\perp BP \mid (||GP, BP|| > 0)$$

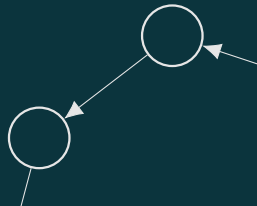
Figure inspired by: Seitzer, M., Schölkopf, B. and Martius, G., 2021. Causal influence detection for improving efficiency in reinforcement learning. Advances in Neural Information Processing Systems, 34, pp.22905-22918.



Section

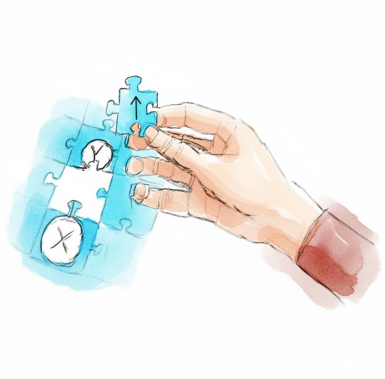
1

# Meta-Causality



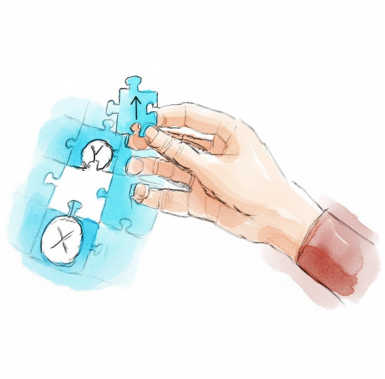
# “Causal Understanding”

We would like to have a framework that allows us to reason about and manipulate causal relations in dynamic environments.



# “Causal Understanding”

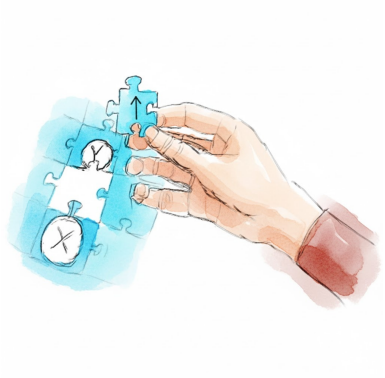
We would like to have a framework that allows us to reason about and manipulate causal relations in dynamic environments.




– Predict the stability of causal links.

# “Causal Understanding”

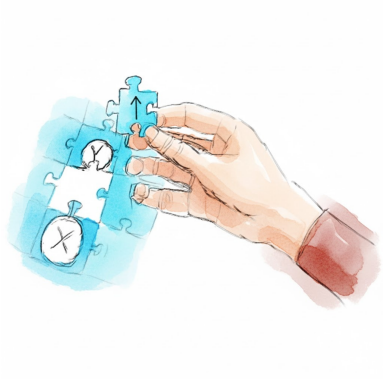
We would like to have a framework that allows us to reason about and manipulate causal relations in dynamic environments.




– Predict the stability of causal links.
– Reason over system dynamics.

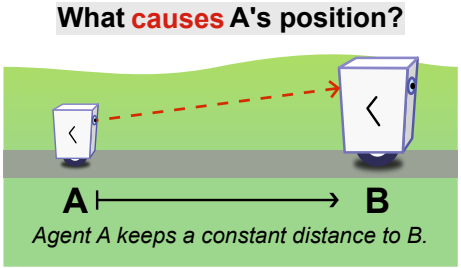
# “Causal Understanding”

We would like to have a framework that allows us to reason about and manipulate causal relations in dynamic environments.



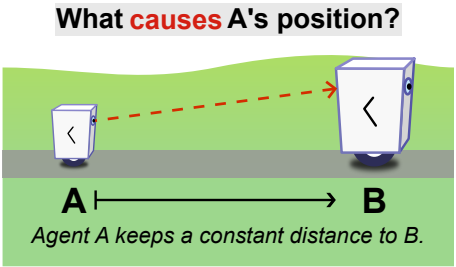

- Predict the stability of causal links.
- Reason over system dynamics.
- Attribution beyond static causal graphs.

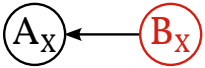
# Attributing Responsibility: A Meta-Causal Perspective




Willig, M., Tobiasch, T., Busch, F.P., Seng, J., Dhami, D.S. and Kersting, K., Systems with Switching Causal Relations: A Meta-Causal Perspective. In The Thirteenth International Conference on Learning Representations.

# Attributing Responsibility: A Meta-Causal Perspective

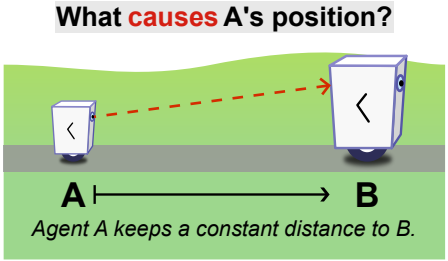


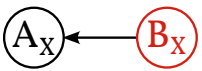




**Classical Attribution**
$A_X$ is caused by the structural equation $A_X := f(B_X)$.

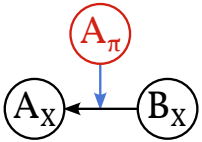
Willig, M., Tobiasch, T., Busch, F.P., Seng, J., Dhami, D.S. and Kersting, K., Systems with Switching Causal Relations: A Meta-Causal Perspective. In The Thirteenth International Conference on Learning Representations.

# Attributing Responsibility: A Meta-Causal Perspective




**Classical Attribution**
$A_X$ is caused by the structural equation $A_X := f(B_X)$.




**Meta-Causal Attribution**
But the relation $B_X \rightarrow A_X$ only *exists* due to A's policy $A_\pi$.

Meta-Causality considers factors that lead to the emergence of causal edges.

Willig, M., Tobiasch, T., Busch, F.P., Seng, J., Dhami, D.S. and Kersting, K., Systems with Switching Causal Relations: A Meta-Causal Perspective. In The Thirteenth International Conference on Learning Representations.

# Factors of Change: Meta-Causal Variables

$$\mathbf{C} := \{X_k \in \mathbf{X} \mid \exists X_i, X_j \in \mathbf{X}.\, \exists \mathbf{x}, \mathbf{x}' \in \boldsymbol{\mathcal{X}} \text{ s.t.}$$
$$(\mathbf{x}_{\bar{k}} = \mathbf{x}'_{\bar{k}}) \ \wedge \ (x_k \neq x'_k) \wedge (\mathcal{I}(\mathbf{x}, X_i, X_j) \neq \mathcal{I}(\mathbf{x}', X_i, X_j))\}$$

*"With all other variables fixed, a different value of $X_k$ can induce a different edge type"*

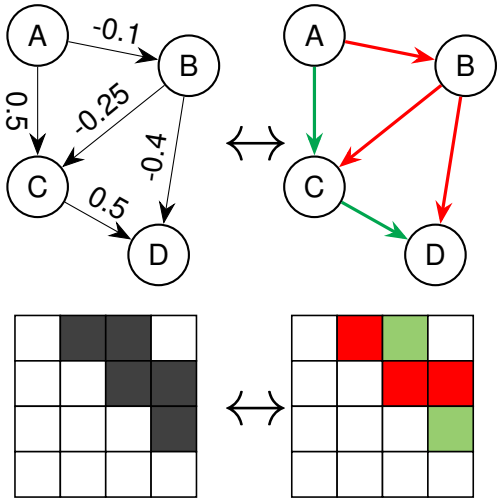
Willig, M., Tobiasch, T., Dhami, D.S. and Kersting, K., When Causal Dynamics Matter: Adapting Causal Strategies through Meta-Aware Interventions. In The Thirty-ninth Annual Conference on Neural Information Processing Systems.

# Qualitative Causal Relations

In every-day life we communicate *qualitative* properties of causal relations.

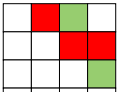
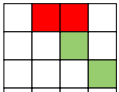
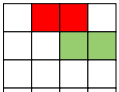
→ Abstract away from specific structural equations.

- "Temperature increases with altitude."
- "Caffeine intake increases alertness, but deprives sleep quality."
- "Time spent studying, improves exam performance, but reduces spare time."
- "Regulations reduce company gains, but improve environmental impact."
- …



# Meta-Causal States

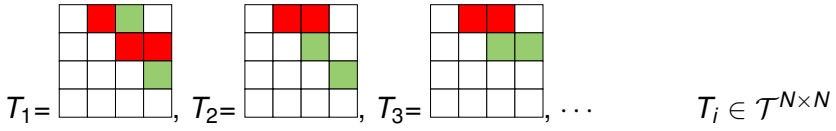
A **Meta-Causal State** (MCS) is a particular configuration of a qualitative causal graph.

$T_1 =$  , $T_2 =$  , $T_3 =$  , $\cdots$ $T_i \in \mathcal{T}^{N \times N}$

Willig, M., Tobiasch, T., Busch, F.P., Seng, J., Dhami, D.S. and Kersting, K., Systems with Switching Causal Relations: A Meta-Causal Perspective. In The Thirteenth International Conference on Learning Representations.

# Meta-Causal States

A **Meta-Causal State** (MCS) is a particular configuration of a qualitative causal graph.




The system follows an underlying mediation process $\mathcal{E} = (\mathcal{S}, \sigma)$ where

- $\mathcal{S}$ is the domain of the process.
- $\sigma : \mathcal{S} \to \mathcal{S}$ is the transition function (e.g. a Markov process).

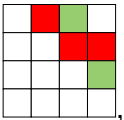
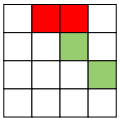
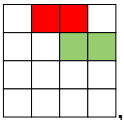
Causal variables **X** are identified via a causal abstraction: $\varphi : \mathcal{S} \to \boldsymbol{\mathcal{X}}$.

$\to$ Causal relations follow by abstracting the process $\varphi \circ \sigma$.

Willig, M., Tobiasch, T., Busch, F.P., Seng, J., Dhami, D.S. and Kersting, K., Systems with Switching Causal Relations: A Meta-Causal Perspective. In The Thirteenth International Conference on Learning Representations.

# Meta-Causal State Identification

A **Meta-Causal State** is a particular configuration of a qualitative causal graph.

$T_1=$  , $T_2=$  , $T_3=$  , $\cdots$ $\quad T_i \in \mathcal{T}^{N \times N}$

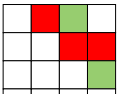
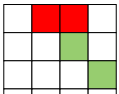
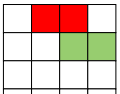
**Types** $T_{s,ij}$ are identified via the **Identification Function** $\mathcal{I} : \mathcal{S} \times \mathbf{X} \times \mathbf{X} \to \mathcal{T}$,
with $T_{s,ij} := \mathcal{I}(s, X_i, X_j) = \tau_{ij}(\varphi(s), \varphi \circ \sigma)$ and

$\rightarrow$ **Type encoders**, $\tau_{ij} : \mathcal{X} \times \boldsymbol{\mathcal{X}}^{\mathcal{S}} \to \mathcal{T}$, that identify the relation type between any two variables $X_i, X_j$ from the context $s \in \mathcal{S}$ and the causal relations $\varphi \circ \sigma$.

Willig, M., Tobiasch, T., Busch, F.P., Seng, J., Dhami, D.S. and Kersting, K., Systems with Switching Causal Relations: A Meta-Causal Perspective. In The Thirteenth International Conference on Learning Representations.

# Meta-Causal State Identification

A **Meta-Causal State** is a particular configuration of a qualitative causal graph.

$T_1=$  , $T_2=$  , $T_3=$  , $\cdots$ $\quad T_i \in \mathcal{T}^{N \times N}$

**Types** $T_{s,ij}$ are identified via the **Identification Function** $\mathcal{I} : \mathcal{S} \times \mathbf{X} \times \mathbf{X} \to \mathcal{T}$,
with $T_{s,ij} := \mathcal{I}(s, X_i, X_j) = \tau_{ij}(\varphi(s), \varphi \circ \sigma)$ and

$\rightarrow$ **Type encoders**, $\tau_{ij} : \mathcal{X} \times \boldsymbol{\mathcal{X}}^{\mathcal{S}} \to \mathcal{T}$, that identify the relation type between any two variables $X_i, X_j$ from the context $s \in \mathcal{S}$ and the causal relations $\varphi \circ \sigma$.

Type encoders can be freely chosen by the user.

$\rightarrow$ A **Meta-Causal Frame** is a combination of a mediation process with a particular set of type encoders.

Willig, M., Tobiasch, T., Busch, F.P., Seng, J., Dhami, D.S. and Kersting, K., Systems with Switching Causal Relations: A Meta-Causal Perspective. In The Thirteenth International Conference on Learning Representations.

# Meta-Causal Models

Given a Meta-Causal Frame,
**Meta-Causal Models** (MCM) model the change in MCS:

$$\delta : \mathcal{T}^{N \times N} \times \mathcal{S} \to P(\mathcal{T}^{N \times N})$$

***MCM are (probabilistic) state machines***

$\to$ 'Given the current MCS and the system state, what is next MCS?'

# Meta-Causal Models

Given a Meta-Causal Frame,
**Meta-Causal Models** (MCM) model the change in MCS:

$$\delta : \mathcal{T}^{N \times N} \times \mathcal{S} \rightarrow P(\mathcal{T}^{N \times N})$$

***MCM are (probabilistic) state machines***

$\rightarrow$ 'Given the current MCS and the system state, what is next MCS?'

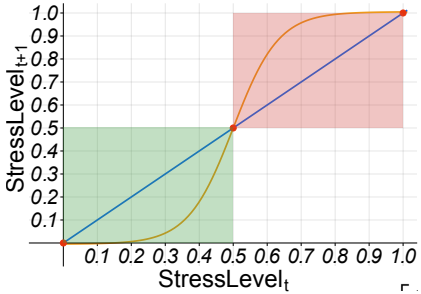
MCM model the qualitative change in cause-effect relations.
$\rightarrow$ MCM allow to reason about the emergence and vanishing of causal edges.

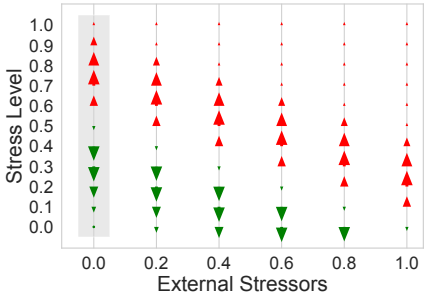
# Dynamical Systems - Self-Reinforcing Stress

Consider a person with a particular stress level $S$ and external stress factors $E$.

→ On busy days they stress themselves and stress levels rise even more.

→ After relaxed days stress-levels can reduce again.




**Variables:** 'external stress' $e$, 'decayed stress' $d$, 'resulting stress' $s$.

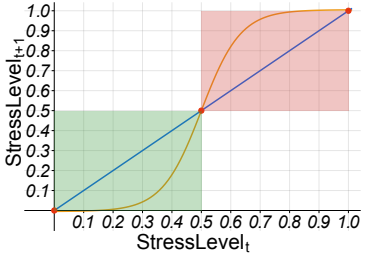
$$f_d := 0.95 \, \text{clip}_{[0,1]}(s_{t-1} + 0.5 \times E) \qquad f_s := 1.01\left(\frac{1}{1 + e^{-15d+7.5}} - 0.5\right) + 0.5$$

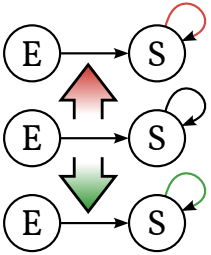
# Dynamics Systems - Self-Reinforcing Stress

‘Is the person stressing themselves at the moment?’

→ Types can change without the structural equations changing.

→ The state of the system matters!



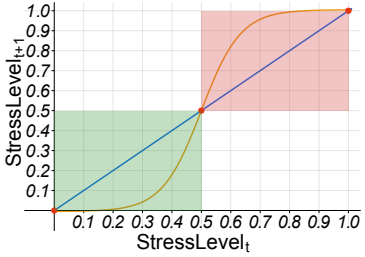



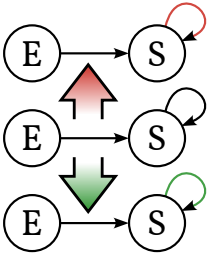
# Dynamics Systems - Self-Reinforcing Stress

‘Is the person stressing themselves at the moment?’

→ Types can change without the structural equations changing.

→ The state of the system matters!



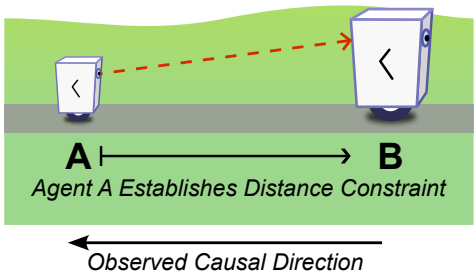




Influence of $S$ onto itself identifies suppressing or self-reinforcing dynamics.

→ $(T_s = \text{supressing}) \Leftarrow (\ddot{f}_s < 0)$.

$$\tau\left(\begin{bmatrix} e_t \\ s_t \end{bmatrix}\right) := \begin{bmatrix} 0 & 0 \\ 1 & a \end{bmatrix} \text{ with } a := \text{sign}(\ddot{f}_s) = \text{sign}(t_s - 0.5) \in \{-1, 0, 1\}$$

# Meta-Causal State Inference




Some meta-causal states can be inferred from observations:

$$(t_{B \to A} = \text{“A chasing”}) \Leftrightarrow (\dot{A}_{pos} \cdot (B_{pos} - A_{pos}) > 0)$$

# Bridging the Gap: Meta-Causal Predictability

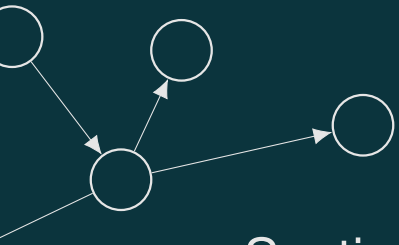
The causal abstraction might marginalize variables that are important for predicting the future trajectory of the system.

$\rightarrow$ This prevents us from accurately predicting future states from within the MCM.

**Meta-Causal Predictability**

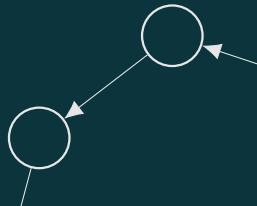
A Meta Causal Model $\mathcal{A} = (\mathcal{T}^{N \times N}, \mathcal{S}, \delta)$ is called **meta-causal predictable** if the next meta-causal state $\mathsf{T}_{\mathsf{s}_{t+1}} \in \mathcal{T}^{N \times N}$ can be predicted purely from the variable values at the current time step $\mathbf{x}_t \in \boldsymbol{\mathcal{X}} = \mathcal{S}$, so that the transition function takes the form $\delta^{\boldsymbol{\mathcal{X}}} : \boldsymbol{\mathcal{X}} \rightarrow \mathcal{T}^{N \times N}$.

Willig, M., Tobiasch, T., Dhami, D.S. and Kersting, K., When Causal Dynamics Matter: Adapting Causal Strategies through Meta-Aware Interventions. In The Thirty-ninth Annual Conference on Neural Information Processing Systems.



Section 2

# Meta-Causal Analysis



# Meta-Causal Analysis (MCA)

Similar to how causal effects quantify the influence between variables, **Meta-Causal Effects** quantify changes in the state transitions.

Questions answered by MCA:

1. What is the probability of a system to adapt a desired MCS?
2. How stable is a particular MCS?
3. Which transition pathways can be taken to obtain a particular MCS?

Willig, M., Tobiasch, T., Dhami, D.S. and Kersting, K., When Causal Dynamics Matter: Adapting Causal Strategies through Meta-Aware Interventions. In The Thirty-ninth Annual Conference on Neural Information Processing Systems.

# Linearized Meta-Causal Dynamics

**Algorithm 1** Linearized Meta-Causal Dynamics (LMCD) Algorithm

**Input:** SCM: $\mathcal{M} = (\mathbf{V}, \mathbf{U}, \mathbf{F}, P_{\mathbf{U}})$, data: $\mathbf{x}^{\mathbf{I}} = (\mathbf{x}^i)_{i=1}^N \in \mathbf{X}^N$, id. func.: $\mathcal{I} : \mathbf{X} \to \mathrm{T}$
**for each** $\mathbf{x}^i$ in $\mathbf{x}^{\mathbf{I}}$ **do**
  $\mathbf{x}^{i,t+1} \leftarrow \mathbf{F}((\mathbf{x}^i|_{\mathbf{V}}) \cup (\mathbf{u}^{t+1} \sim P_{\mathbf{U}}))$ ▷ Advance the system.
  $(\mathrm{T}^{i,t}, \mathrm{T}^{i,t+1}) \leftarrow (\mathcal{I}(\mathbf{x}^i), \mathcal{I}(\mathbf{x}^{i,t+1}))$ ▷ Identify MCS transition pair.
$U \leftarrow (\bigcup_i I(\mathrm{T}^{i,t})) \cup (\bigcup_i I(\mathrm{T}^{i,t+1}))$ ▷ Determine set of unique MCS.
**for each** $(u, v)$ in $\{1, \ldots, |U|\}^2$ **do** ▷ Approximate transition dynamics, $P \in \mathbb{R}^{|U| \times |U|}$.
  $P_{u,v} \leftarrow \sum_{i \in [1..N]} (\mathbf{1}((I(\mathrm{T}^{i,t}) = u) \wedge (I(\mathrm{T}^{i,t+1}) = v))) / \sum_{i \in [1..N]} \mathbf{1}(I(\mathrm{T}^{i,t} = v))$
$[Q \leftarrow e^{P-I}]$ ▷ Optional: Compute continuous time rate matrix. ($I$ is the identity matrix.)
**return** $P, [Q]$

1) Start with sample population.

Willig, M., Tobiasch, T., Dhami, D.S. and Kersting, K., When Causal Dynamics Matter: Adapting Causal Strategies through Meta-Aware Interventions. In The Thirty-ninth Annual Conference on Neural Information Processing Systems.

# Linearized Meta-Causal Dynamics

**Algorithm 1** Linearized Meta-Causal Dynamics (LMCD) Algorithm

**Input:** SCM: $\mathcal{M} = (\mathbf{V}, \mathbf{U}, \mathbf{F}, P_{\mathbf{U}})$, data: $\mathbf{x}^{\mathbf{I}} = (\mathbf{x}^i)_{i=1}^N \in \mathbf{X}^N$, id. func.: $\mathcal{I} : \mathbf{X} \to \mathrm{T}$
**for each** $\mathbf{x}^i$ in $\mathbf{x}^{\mathbf{I}}$ **do**
  $\mathbf{x}^{i,t+1} \leftarrow \mathbf{F}((\mathbf{x}^i|_{\mathbf{V}}) \cup (\mathbf{u}^{t+1} \sim P_{\mathbf{U}}))$ ▷ Advance the system.
  $(\mathrm{T}^{i,t}, \mathrm{T}^{i,t+1}) \leftarrow (\mathcal{I}(\mathbf{x}^i), \mathcal{I}(\mathbf{x}^{i,t+1}))$ ▷ Identify MCS transition pair.
$U \leftarrow (\bigcup_i I(\mathrm{T}^{i,t})) \cup (\bigcup_i I(\mathrm{T}^{i,t+1}))$ ▷ Determine set of unique MCS.
**for each** $(u, v)$ in $\{1, \ldots, |U|\}^2$ **do** ▷ Approximate transition dynamics, $P \in \mathbb{R}^{|U| \times |U|}$.
  $P_{u,v} \leftarrow \sum_{i \in [1..N]} (\mathbf{1}((I(\mathrm{T}^{i,t}) = u) \wedge (I(\mathrm{T}^{i,t+1}) = v))) / \sum_{i \in [1..N]} \mathbf{1}(I(\mathrm{T}^{i,t} = v))$
$[Q \leftarrow e^{P-I}]$ ▷ Optional: Compute continuous time rate matrix. ($I$ is the identity matrix.)
**return** $P$, $[Q]$

2) Identify the MCS of the data points.

Willig, M., Tobiasch, T., Dhami, D.S. and Kersting, K., When Causal Dynamics Matter: Adapting Causal Strategies through Meta-Aware Interventions. In The Thirty-ninth Annual Conference on Neural Information Processing Systems.

# Linearized Meta-Causal Dynamics

**Algorithm 1** Linearized Meta-Causal Dynamics (LMCD) Algorithm

**Input:** SCM: $\mathcal{M} = (\mathbf{V}, \mathbf{U}, \mathbf{F}, P_{\mathbf{U}})$, data: $\mathbf{x}^{\mathbf{I}} = (\mathbf{x}^i)_{i=1}^N \in \mathbf{X}^N$, id. func.: $\mathcal{I} : \mathbf{X} \to \mathrm{T}$
**for each** $\mathbf{x}^i$ in $\mathbf{x}^{\mathbf{I}}$ **do**
  $\mathbf{x}^{i,t+1} \leftarrow \mathbf{F}((\mathbf{x}^i|_{\mathbf{V}}) \cup (\mathbf{u}^{t+1} \sim P_{\mathbf{U}}))$ ▷ Advance the system.
  $(\mathrm{T}^{i,t}, \mathrm{T}^{i,t+1}) \leftarrow (\mathcal{I}(\mathbf{x}^i), \mathcal{I}(\mathbf{x}^{i,t+1}))$ ▷ Identify MCS transition pair.
$U \leftarrow (\bigcup_i I(\mathrm{T}^{i,t})) \cup (\bigcup_i I(\mathrm{T}^{i,t+1}))$ ▷ Determine set of unique MCS.
**for each** $(u, v)$ in $\{1, \ldots, |U|\}^2$ **do** ▷ Approximate transition dynamics, $P \in \mathbb{R}^{|U| \times |U|}$.
  $P_{u,v} \leftarrow \sum_{i \in [1..N]} (\mathbf{1}((I(\mathrm{T}^{i,t}) = u) \wedge (I(\mathrm{T}^{i,t+1}) = v))) / \sum_{i \in [1..N]} \mathbf{1}(I(\mathrm{T}^{i,t} = v))$
$[Q \leftarrow e^{P-I}]$ ▷ Optional: Compute continuous time rate matrix. ($I$ is the identity matrix.)
**return** $P, [Q]$

3) Advance the system and again identify MCS.

Willig, M., Tobiasch, T., Dhami, D.S. and Kersting, K., When Causal Dynamics Matter: Adapting Causal Strategies through Meta-Aware Interventions. In The Thirty-ninth Annual Conference on Neural Information Processing Systems.

# Linearized Meta-Causal Dynamics

**Algorithm 2** Linearized Meta-Causal Dynamics (LMCD) Algorithm

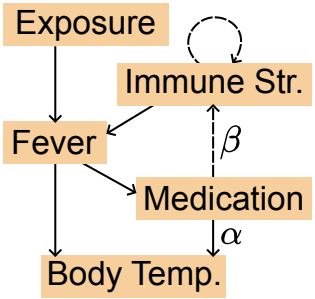
**Input:** SCM: $\mathcal{M} = (\mathbf{V}, \mathbf{U}, \mathbf{F}, P_{\mathbf{U}})$, data: $\mathbf{x}^{\mathbf{I}} = (\mathbf{x}^i)_{i=1}^N \in \mathbf{X}^N$, id. func.: $\mathcal{I} : \mathbf{X} \to \mathrm{T}$
**for each** $\mathbf{x}^i$ in $\mathbf{x}^{\mathbf{I}}$ **do**
  $\mathbf{x}^{i,t+1} \leftarrow \mathbf{F}((\mathbf{x}^i|_{\mathbf{V}}) \cup (\mathbf{u}^{t+1} \sim P_{\mathbf{U}}))$ ▷ Advance the system.
  $(\mathrm{T}^{i,t}, \mathrm{T}^{i,t+1}) \leftarrow (\mathcal{I}(\mathbf{x}^i), \mathcal{I}(\mathbf{x}^{i,t+1}))$ ▷ Identify MCS transition pair.
$U \leftarrow (\bigcup_i I(\mathrm{T}^{i,t})) \cup (\bigcup_i I(\mathrm{T}^{i,t+1}))$ ▷ Determine set of unique MCS.
**for each** $(u, v)$ in $\{1, \ldots, |U|\}^2$ **do** ▷ Approximate transition dynamics, $P \in \mathbb{R}^{|U| \times |U|}$.
  $P_{u,v} \leftarrow \sum_{i \in [1..N]} (\mathbf{1}((I(\mathrm{T}^{i,t}) = u) \wedge (I(\mathrm{T}^{i,t+1}) = v))) / \sum_{i \in [1..N]} \mathbf{1}(I(\mathrm{T}^{i,t} = v))$
$[Q \leftarrow e^{P-I}]$ ▷ Optional: Compute continuous time rate matrix. ($I$ is the identity matrix.)
**return** $P, [Q]$

3) Compute transition statistics.

Willig, M., Tobiasch, T., Dhami, D.S. and Kersting, K., When Causal Dynamics Matter: Adapting Causal Strategies through Meta-Aware Interventions. In The Thirty-ninth Annual Conference on Neural Information Processing Systems.

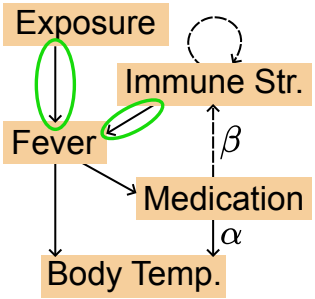
# Medicating Flu

When do causal edges become active? $\rightarrow$



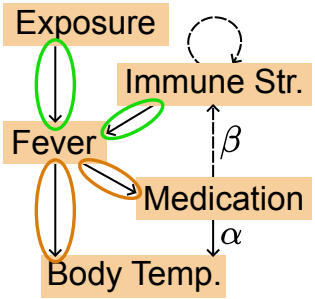
# Medicating Flu

When do causal edges become active? $\rightarrow$



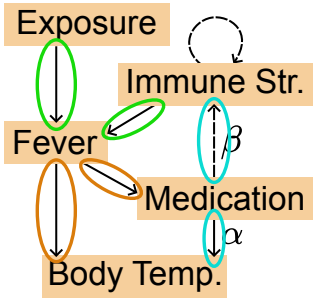
# Medicating Flu

When do causal edges become active? $\rightarrow$



# Medicating Flu

When do causal edges become active? $\rightarrow$

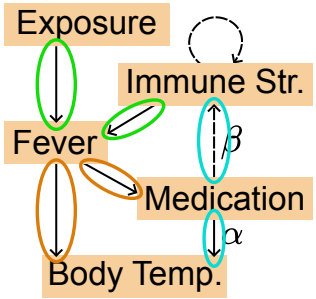


# Medicating Flu

When do causal edges become active? $\rightarrow$

Compare 2 Medications: 3 Unique MCS:

1. **Healthy:** No Fever, No Medication.
2. **Mild Fever:** Fever, No Medication.
3. **Strong Fever:** Fever, Medication.
4. ~~**???:** No Fever, Medication.~~

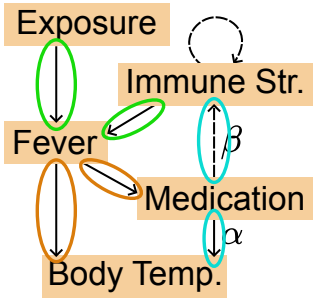


# Medicating Flu

When do causal edges become active? →

Compare 2 Medications: 3 Unique MCS:

1. **Healthy:** No Fever, No Medication.
2. **Mild Fever:** Fever, No Medication.
3. **Strong Fever:** Fever, Medication.
4. ~~**???:** No Fever, Medication.~~

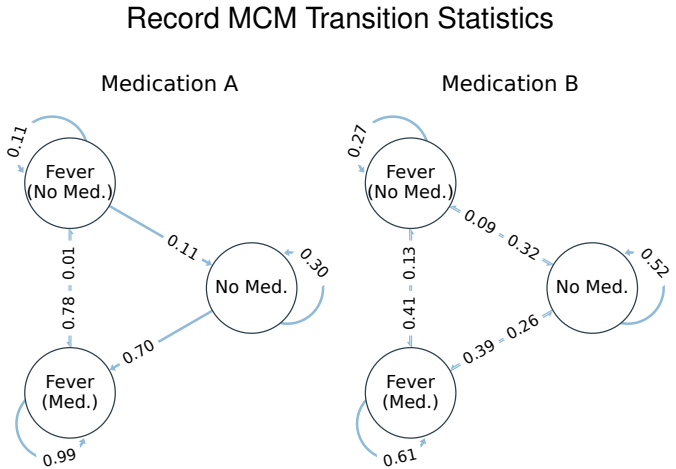



Compare two medications:

**A** Higher effect on fever symptoms and higher immune suppression.

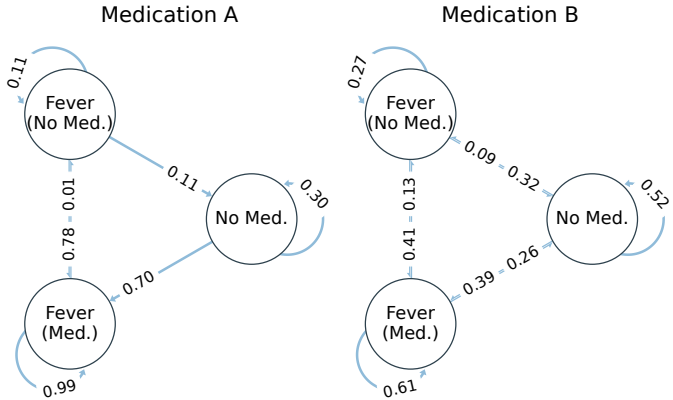
**B** Lower effect on fever symptoms and milder immune suppression.

# Medicating Flu: MCA



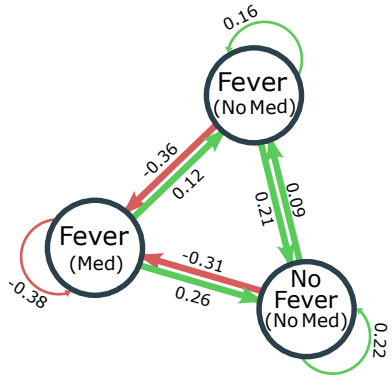
# Medicating Flu: MCA

Record MCM Transition Statistics



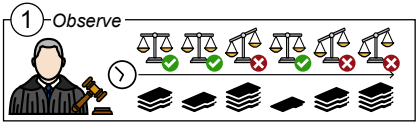

**Meta-Causal ATE**

$\text{sMCATE}(P_A, P_B) := P_B - P_A$



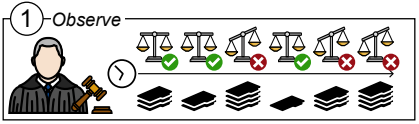
# Judicial Decision Making

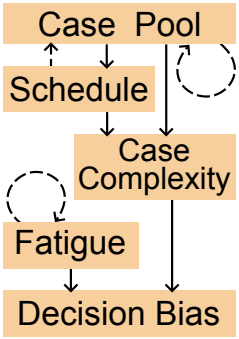
*“A judge notices that their decisions are becoming inaccurate over time.”*



# Judicial Decision Making

*"A judge notices that their decisions are becoming inaccurate over time."*




**Structural Equations:**

$$\text{CasePool}_t := \text{CasePool}_{t-1} \setminus \{\text{CasePool}_{t-1}[\text{Schedule}_{t-1}]\}$$

$$\text{Schedule}_t^{\text{initial}} := 0$$

$$\text{CaseComplexity}_t := \text{CasePool}_t[\text{Schedule}_t]$$

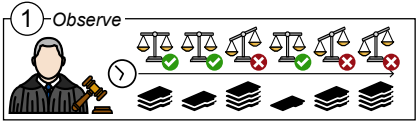
$$\text{Fatigue}_t := \text{Fatigue}_{t-1} + 0.5$$

$$\text{DecisionBias}_t := \max(\text{Fatigue}_t + \text{CaseComplexity}_t - 5, 0)$$

$$\textit{(initial)}\ \text{CasePool}_0 := [3, 2, 4, 1, 3, 4]$$

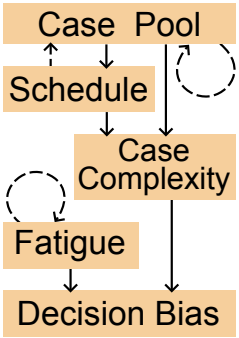
# Judicial Decision Making

*“A judge notices that their decisions are becoming inaccurate over time.”*



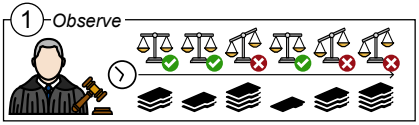

‘When do decisions become biased?’

→ Combination of *Fatigue* or *Case Complexity*.



# Judicial Decision Making

*"A judge notices that their decisions are becoming inaccurate over time."*

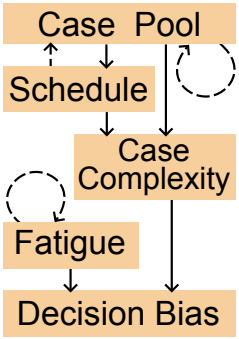



'When do decisions become biased?'

$\rightarrow$ Combination of *Fatigue* or *Case Complexity*.

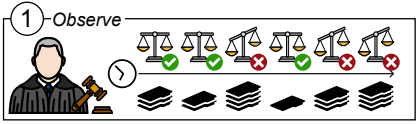
**Idea:** Schedule the hard cases first:

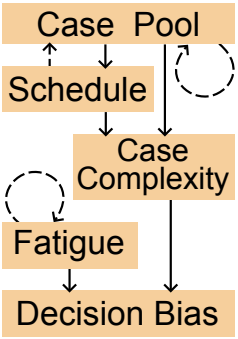
$$\text{Schedule}_t^{\text{adapted}} := \text{argmax}_i(\text{CasePool}_t)$$



# Judicial Decision Making

*"A judge notices that their decisions are becoming inaccurate over time."*



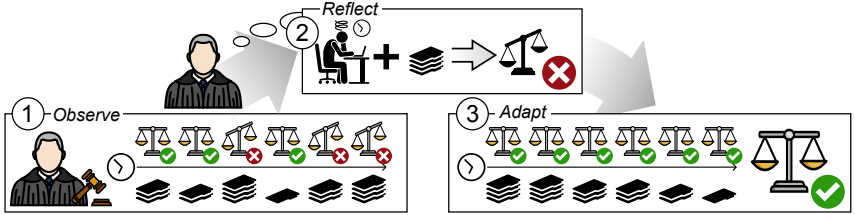



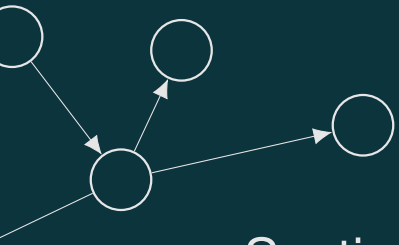

'When do decisions become biased?'

$\rightarrow$ Combination of *Fatigue* or *Case Complexity*.

**Idea:** Schedule the hard cases first:

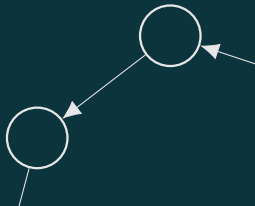
$$\text{Schedule}_t^{\text{adapted}} := \text{argmax}_i(\text{CasePool}_t)$$





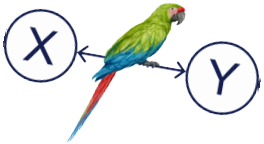
Section 3

# ‘Genuine’ Causal Understanding



# From Parroting to Understanding: A Meta-Causal Path

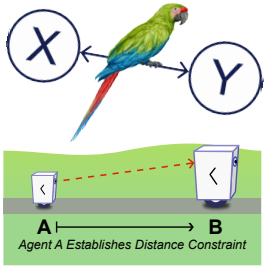
**Reflection & Adaptation:** Genuine understanding isn't just about knowing that 'A causes B', but understanding the conditions under which that relationship holds, and to adapt when it changes.



# From Parroting to Understanding: A Meta-Causal Path

**Reflection & Adaptation:** Genuine understanding isn't just about knowing that 'A causes B', but understanding the conditions under which that relationship holds, and to adapt when it changes.

**Meta-Causal Models** allow to explicitly model and reason about *how* and *why* causal relationships change.

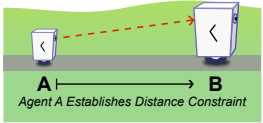


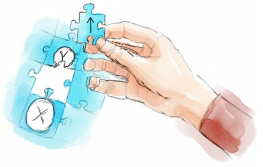
# From Parroting to Understanding: A Meta-Causal Path

**Reflection & Adaptation:** Genuine understanding isn't just about knowing that 'A causes B', but understanding the conditions under which that relationship holds, and to adapt when it changes.

**Meta-Causal Models** allow to explicitly model and reason about *how* and *why* causal relationships change.

**Future AI Systems** should not just produce due to their intrinsic weights, but deliberately think about the underlying mechanisms at play.

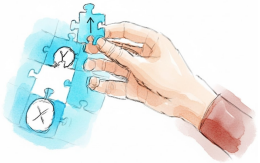



# From Parroting to Understanding: A Meta-Causal Path

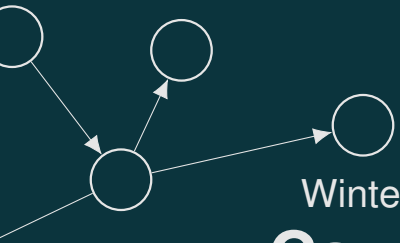
**Reflection & Adaptation:** Genuine understanding isn't just about knowing that 'A causes B', but understanding the conditions under which that relationship holds, and to adapt when it changes.

**Meta-Causal Models** allow to explicitly model and reason about *how* and *why* causal relationships change.

**Future AI Systems** should not just produce due to their intrinsic weights, but deliberately think about the underlying mechanisms at play.

*Meta-Causality may be the dividing line between systems that merely describe the world from those that truly understand it.*



TECHNISCHE UNIVERSITÄT DARMSTADT

AIML Lab

Winter Semester 2025/26 Lecture

# Causality for AI & ML

Feel free to reach out:

**Moritz Willig** https://moritz-willig.de/

Computer Science Department
Technical University of Darmstadt
moritz.willig@cs.tu-darmstadt.de