Introducing causal inference in the energy-efficient building design process

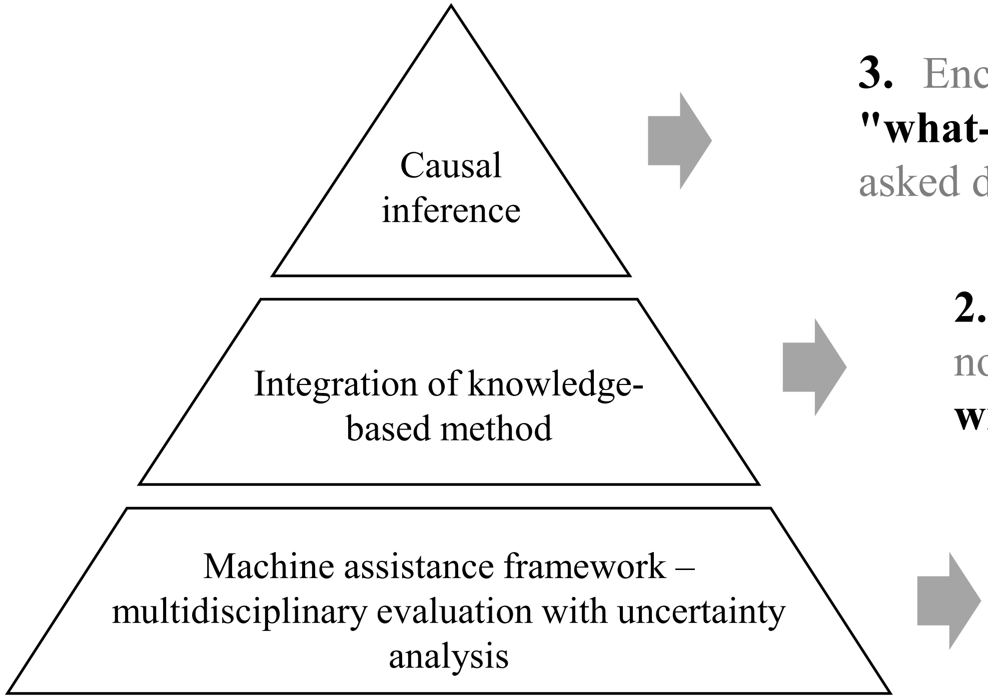
Xia Chen[a d], Jimmy Abualdenien[b], Manav Mahan Singh[c], André Borrmann[b], Philipp Geyer[a d]

[a] Technische Universität Berlin, Germany, [b] Technische Universität München, Germany, [c] Katholieke Universiteit Leuven, Belgium, [d] Leibniz University Hannover, Germany

Sustainable Building Systems Group, Prof. Dr.-Ing. Philipp Geyer

xia.chen@iek.uni-hannover.de

Our general research objective:
Machine assistance in engineering design support



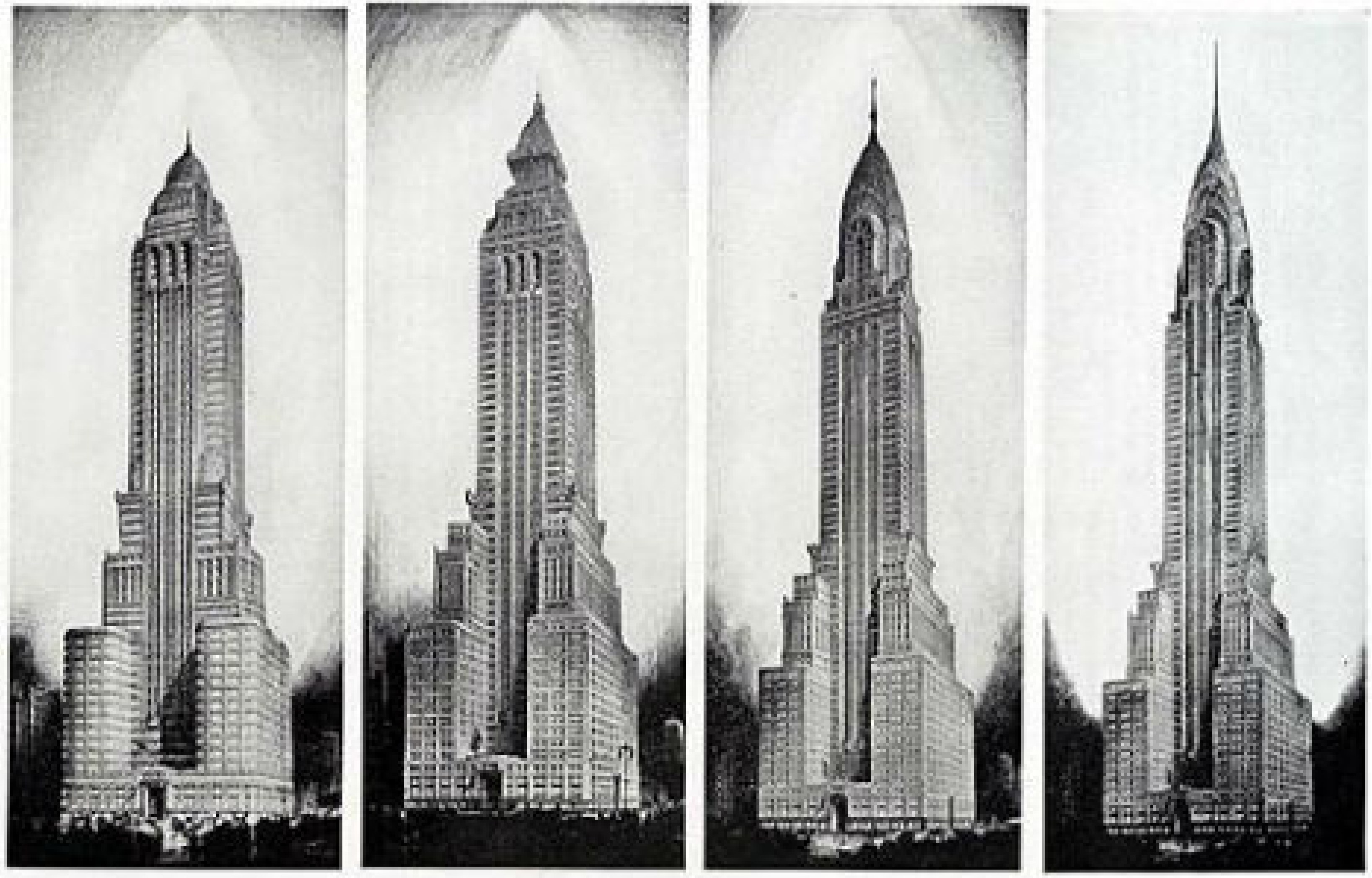
3. Encoding causal modeling process of **reasoning and answering "what-if" questions** which are intuitively generated and commonly asked during the design process

2. Exploring methodologies to identify, analysis, and present information not only relying on empiricism, data-driven methods, but also **embedded with logicism representation of domain principles**

1. A **machine assistance framework** takes a shared set of representations to conduct **multi-disciplinary evaluation** and **uncertainties analysis** with incomplete inputs acceptance that aligned with the design process

Embedding logicism representations into **the machine assistance →**
conduct reasoning analysis and answer "what-if" questions

“What-if” question in the building design domain



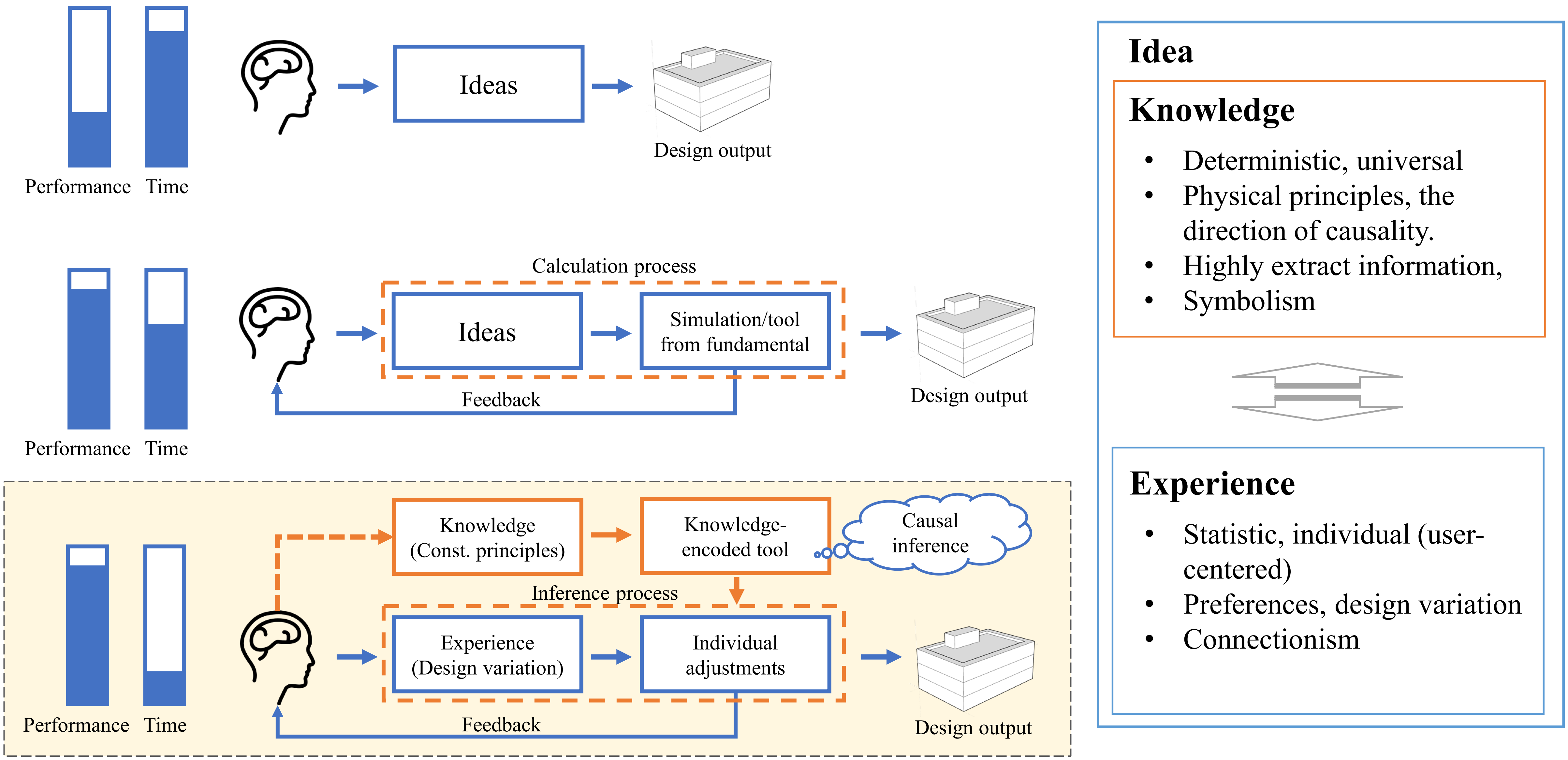
(images via: sleepny, wikimedia commons)

“What-if” questions are commonly encountered during the design process.

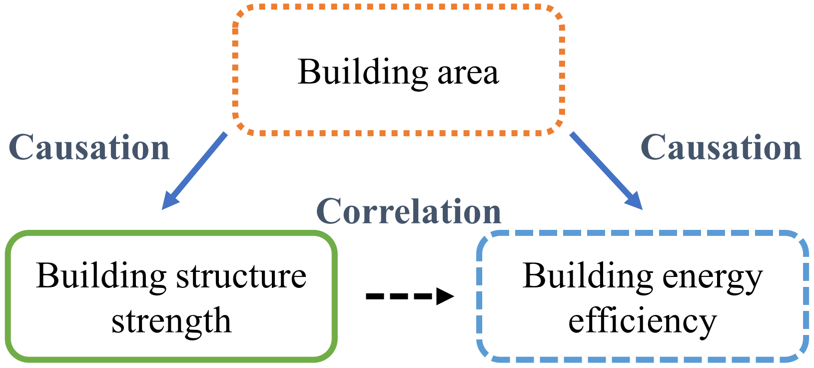
A series of actions to determine the most compatible adjustment with set objective(s)

-> **Core of decision-making support during the design process**

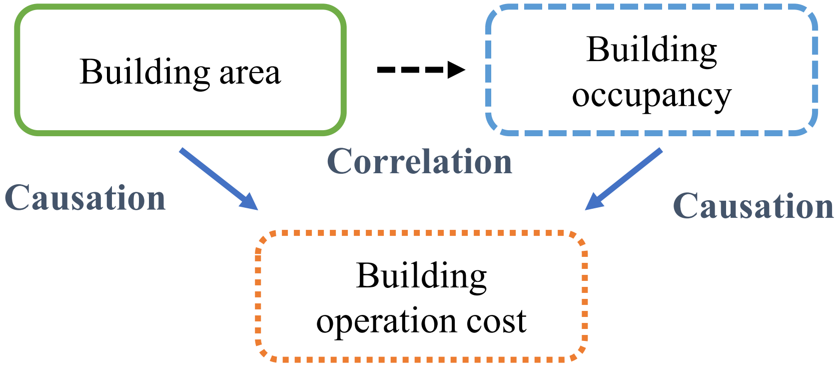
Design process paradigms: the separation of knowledge and experience



Causal inferences logic - parametric independence: Correlation does not imply causation

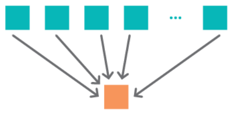


Confounder (fork) : ***Must be*** controlled for in order to accurately estimate the effect

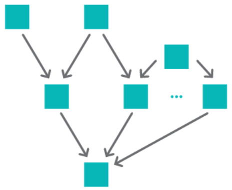


Collider: *Must not be* controlled for in order to accurately estimate the effect

Treatment   Common cause / effect   Outcome



Machine learning model



Reality

independent and identically distributed (i.i.d.): the probability distribution of each value (parameter) should have no dependence on other values;

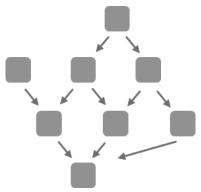
Potential confounding bias in data!

A glance of causal inference

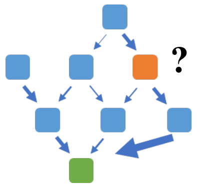
Manually hardcoding domain rules embedded causal constraints
→ Expert systems: **first-order method for causality encoding**

A mathematical rigor approach to find and encode causality directly from data, no semantic grounding required
→ Causal model: **second-order method for causality emergence**

Research objectives



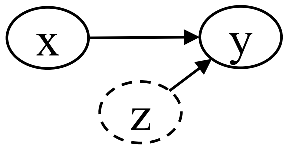
causal relations



causal effects

Asymmetry of independent changes in causality: i.e.,

- changes in P(cause) and changes in P(effect|cause) are independent
- changes in P(effect) and changes in P(cause|effect) are not independent

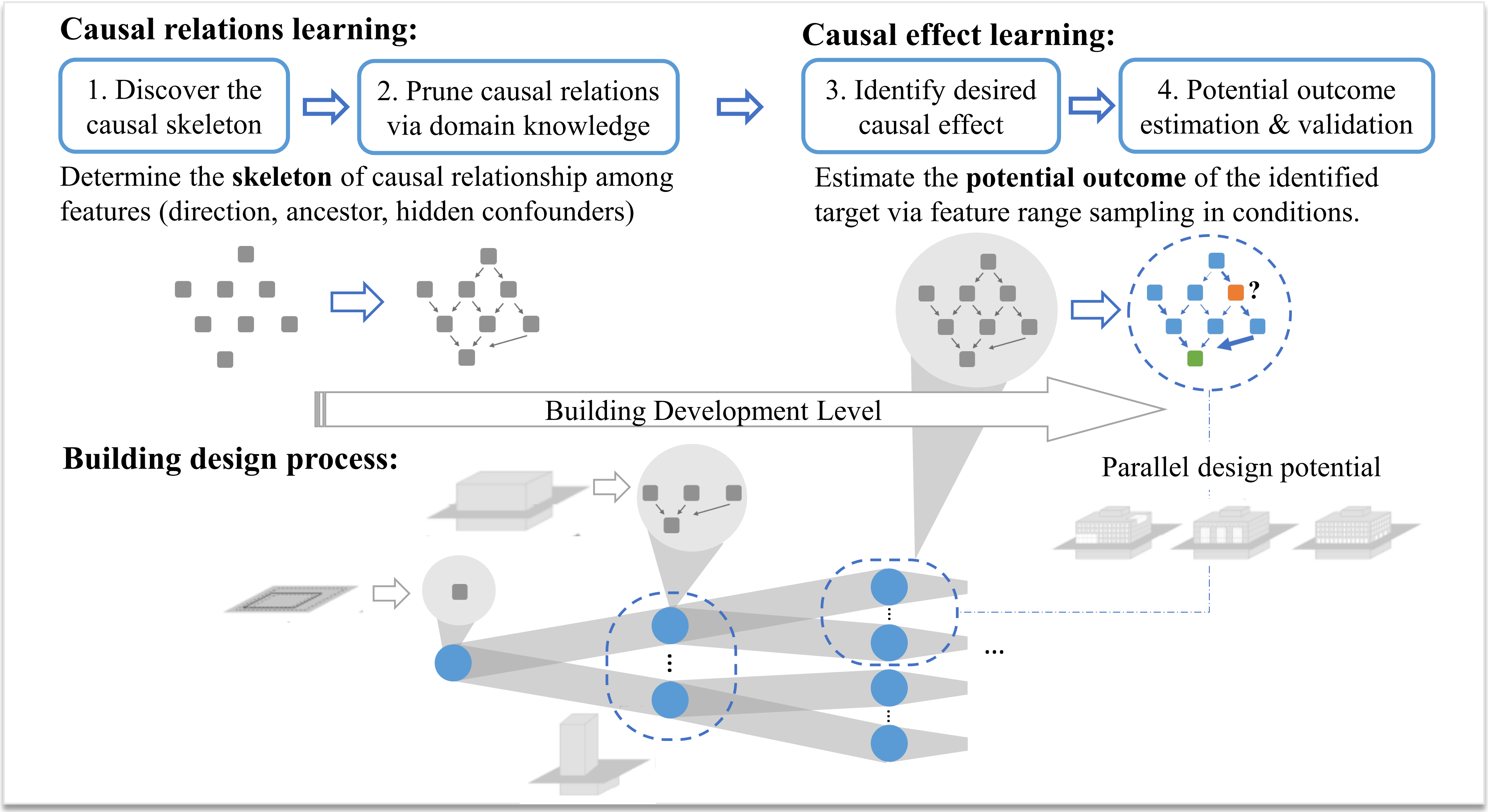


Directed Acyclic Graph (DAG)

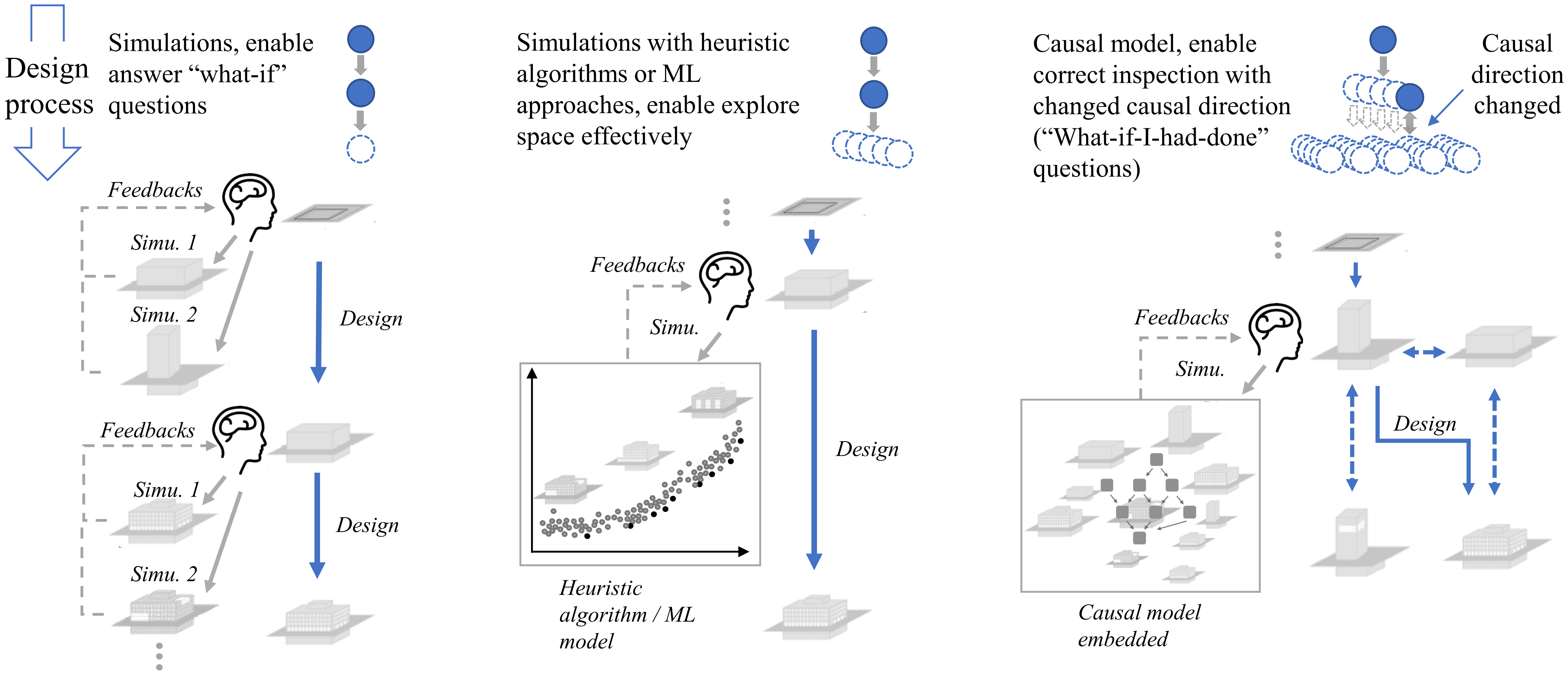
- Structural Causal Model (SCM)
- Potential Outcome Framework

Role	**Example**	**Methods (i.e.)**
Causal skeleton → **Knowledge**	Cause-effect relationships	• D-separation • Back-door criteria
Causal effect → **Experience**	Variable manipulation/intervene	• Average Treatment Effect (ATE) • Conditional ATE

Causal inference in the building design process: a four-step framework



Design assistance dimension: process-oriented informative support



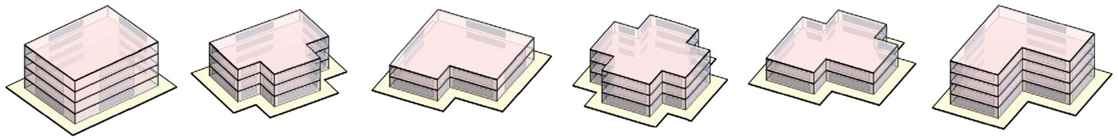
(a) One-dimensional point:
Potential design scheme

(b) Two-dimensional line:
Consequential design combinations

(c) Three-dimensional space:
Potential design spaces (alternative realities)

Case study: building early design phase scenario

Random building shapes



Sampled ranges

Parameter	Unit	Min	Max
Ground Floor Area [1]	m^2	250	800
Height	m	3	4
Number of Floors	-	2	5
External Wall Area	m^2	200	1800
Windows Area	m^2	30	850
u-Value (Wall)	W/m^2K	0.15	0.25
u-Value (Internal Wall)		0.4	0.6
u-Value (Ground Floor)		0.15	0.25
u-Value (Roof)		0.15	0.25
u-Value (Internal Floor)		0.4	0.6
u-Value (Windows)		0.7	1.0
g-Value (Windows)	-	0.3	0.6
Permeability	m^3/m^2h	6	9
WWR [2]	-	0.1	0.5
Equipment Heat Gain	W/m^2	10	14
Building Occupancy	Person/m^2	16	24

[1] Ground Floor Area for random shapes buildings

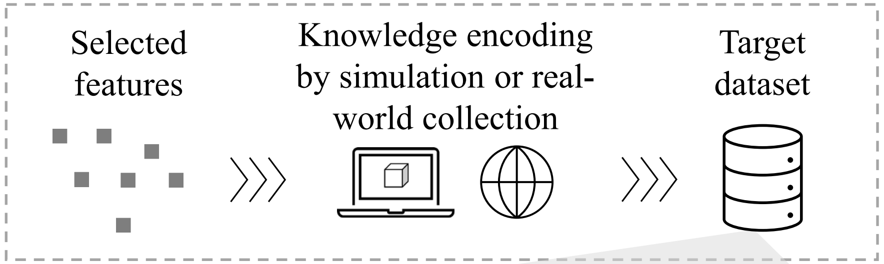
[2] Window-to-wall ratio (WWR) varies independently in each direction

Generated data

Height	Volume	Number of Floors	External Wall	Ground Floor	Window Area	u-Value (Wall)	u-Value (Ground	u-Value (Roof)
3.74219	8039.57	4	1051.36	537.09	357.575	0.23828	0.21797	0.20234
3.24219	5150.12	3	610.043	529.49	305.469	0.18828	0.16797	0.15234
3.82813	11041.8	4	1050.95	721.1	597.062	0.22031	0.15156	0.24531
3.46875	2524.66	3	467.647	242.61	195.751	0.23438	0.15313	0.21563
3.65625	7635.9	5	1018.12	417.69	476.369	0.20313	0.23438	0.19688
3	864	2	259.2	144	28.8	0.15	0.15	0.15
3.64063	6369.49	4	1039.34	437.39	200.297	0.22656	0.24531	0.15156
3.14063	2683.73	2	341.977	427.26	192.714	0.17656	0.19531	0.20156
3.96875	9691.53	4	1119.71	610.49	463.823	0.18438	0.20313	0.16563
3.15625	8205.33	3	871.051	866.57	243.894	0.15313	0.18438	0.24688
3.75	7315.31	3	803.25	650.25	344.25	0.225	0.175	0.175
3.80469	7637.68	4	938.138	501.86	425.843	0.20703	0.18672	0.19609
3.30469	1186.71	2	243.495	179.55	110.933	0.15703	0.23672	0.24609
3.89063	4455.16	3	691.053	381.7	302.516	0.20156	0.17031	0.22656
3.39063	6138.12	4	790.975	452.58	363.533	0.15156	0.22031	0.17656
3.04688	3689.52	3	503.973	403.64	253.556	0.17969	0.16719	0.19844
3.90625	2976.56	3	583.649	254	163.422	0.22813	0.15938	0.22188
3.5625	2896.1	3	533.64	270.98	171.735	0.15625	0.15625	0.15625

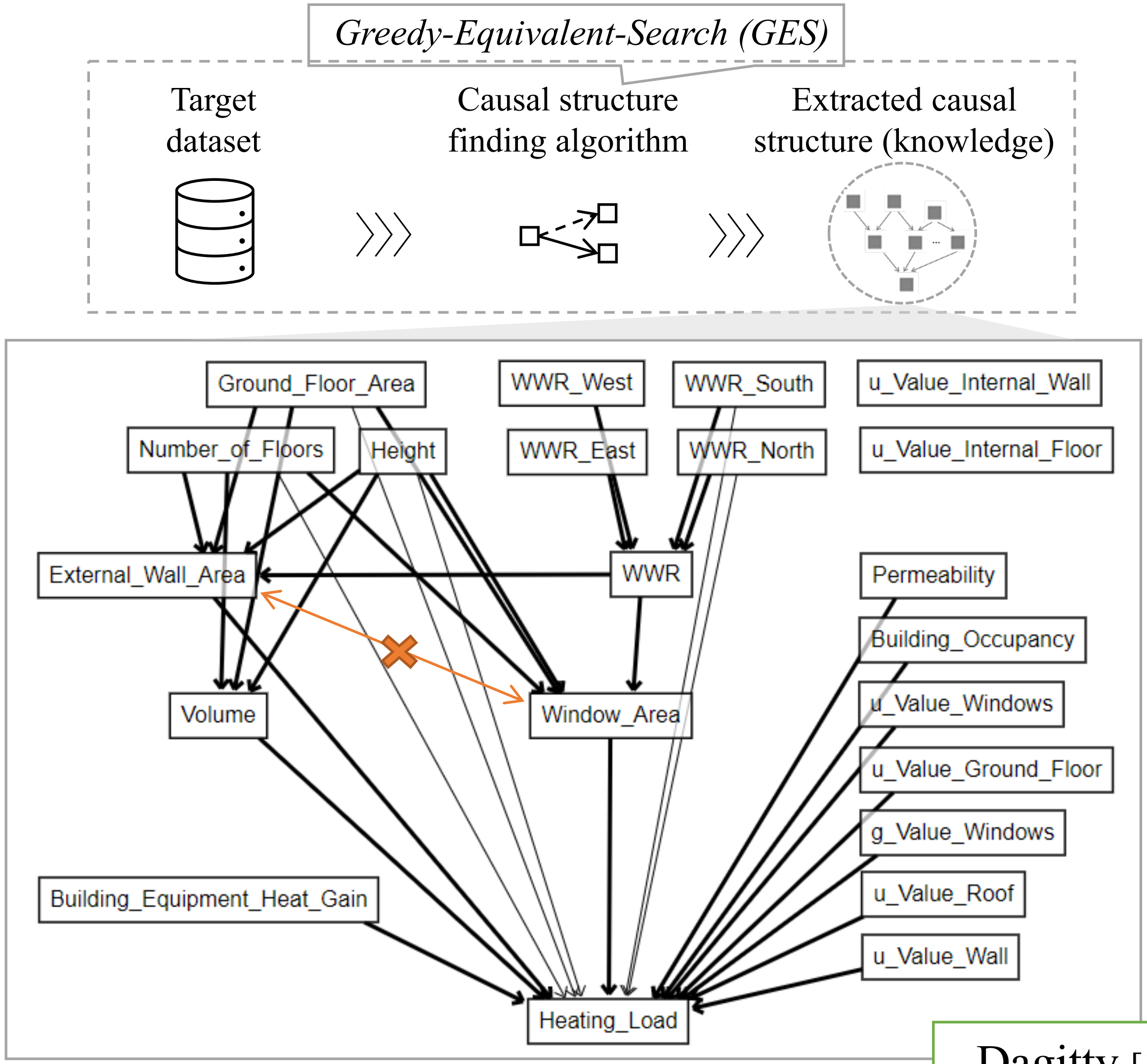
- *Singh MM, Singaravel S, Klein R, Geyer P. Quick energy prediction and comparison of options at the early design stage. Advanced Engineering Informatics 2020;46:101185.*
- *Geyer P, Singh MM, Chen X. Explainable AI for engineering design: A unified approach of systems engineering and component-based deep learning; 2021.*

Causal structure finding



Height	Volume	Number of Floors	External Wall	Ground Floor	Window Area	u-Value (Wall)	u-Value (Ground	u-Value (Roof)
3.74219	8039.57	4	1051.36	537.09	357.575	0.23828	0.21797	0.20234
3.24219	5150.12	3	610.043	529.49	305.469	0.18828	0.16797	0.15234
3.82813	11041.8	4	1050.95	721.1	597.062	0.22031	0.15156	0.24531
3.46875	2524.66	3	467.647	242.61	195.751	0.23438	0.15313	0.21563
3.65625	7635.9	5	1018.12	417.69	476.369	0.20313	0.23438	0.19688
3	864	2	259.2	144	28.8	0.15	0.15	0.15
3.64063	6369.49	4	1039.34	437.39	200.297	0.22656	0.24531	0.15156
3.14063	2683.73	2	341.977	427.26	192.714	0.17656	0.19531	0.20156
3.96875	9691.53	4	1119.71	610.49	463.823	0.18438	0.20313	0.16563
3.15625	8205.33	3	871.051	866.57	243.894	0.15313	0.18438	0.24688
3.75	7315.31	3	803.25	650.25	344.25	0.225	0.175	0.175
3.80469	7637.68	4	938.138	501.86	425.843	0.20703	0.18672	0.19609
3.30469	1186.71	2	243.495	179.55	110.933	0.15703	0.23672	0.24609
3.89063	4455.16	3	691.053	381.7	302.516	0.20156	0.17031	0.22656
3.39063	6138.12	4	790.975	452.58	363.533	0.15156	0.22031	0.17656
3.04688	3689.52	3	503.973	403.64	253.556	0.17969	0.16719	0.19844
3.90625	2976.56	3	583.649	254	163.422	0.22813	0.15938	0.22188
3.5625	2896.1	3	533.64	270.98	171.735	0.15625	0.15625	0.15625

(a) Dataset

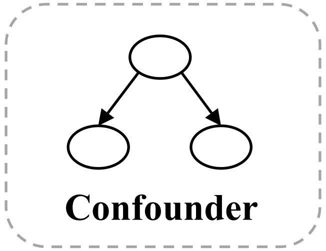


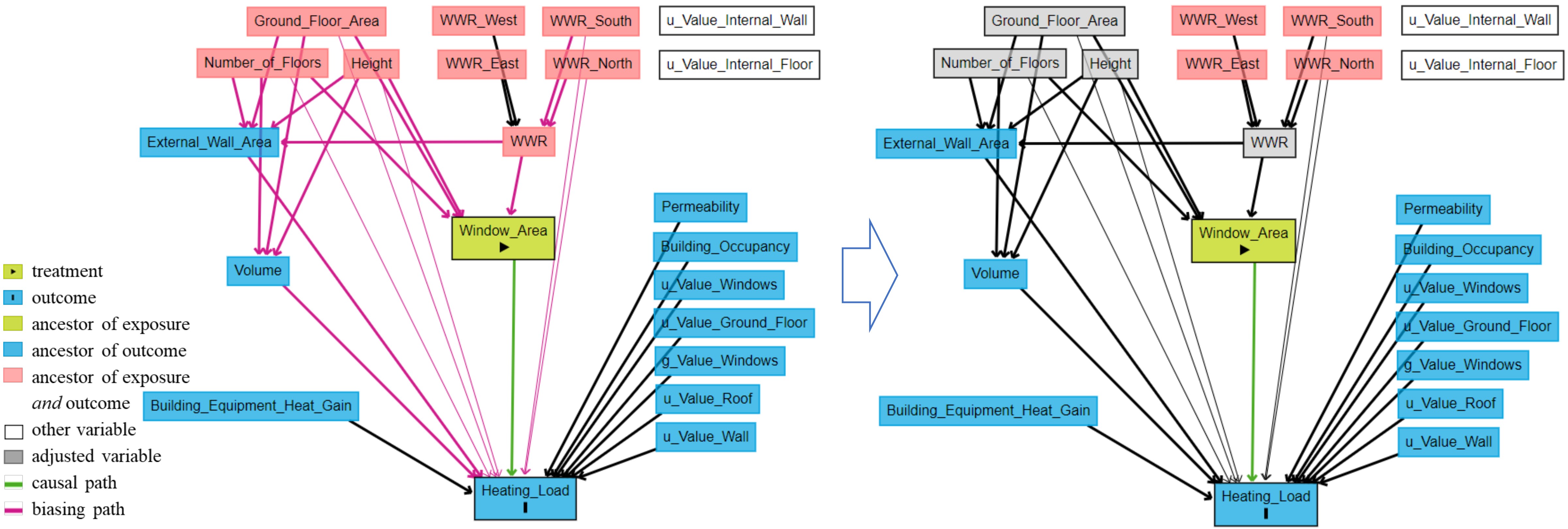
Dagitty [1]

(b) Causal DAG

- *[1] Johannes Textor, Benito van der Zander, Mark K. Gilthorpe, Maciej Liskiewicz, George T.H. Ellison. Robust causal inference using directed acyclic graphs: the R package 'dagitty'.International Journal of Epidemiology 45(6):1887-1894, 2016.*

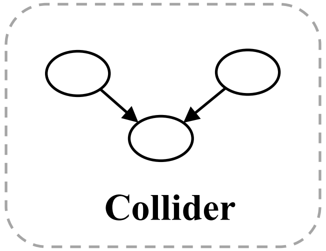
“What-if” scenario i :
Direct causal effect from *window area* to *heating load*?

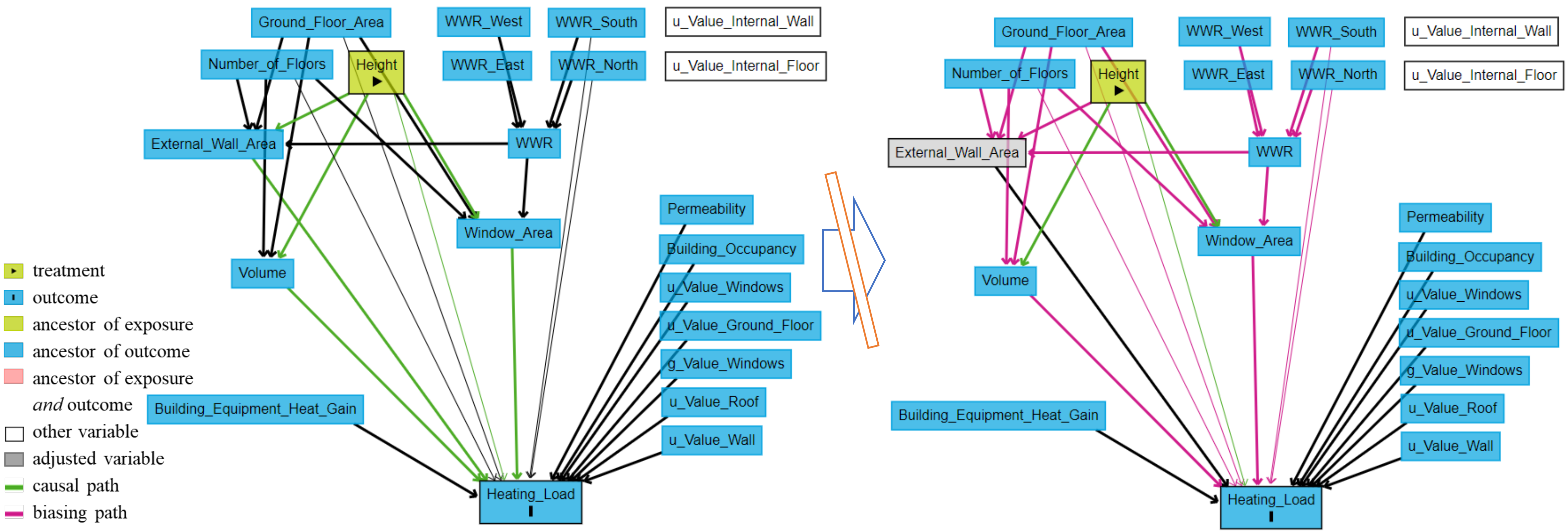




Suggestion to Scenario i: To correctly estimate the total causal effect from ***[Window Area]*** to ***[Heating Load]***, *WWR*, *Ground Floor Area*, *Number of Floors* and *Height* **should be controlled** (fixed) to eliminate biasing paths (red arrows).

"What-if" scenario ii:
Direct causal effect from *building floor height* to *heating load?*



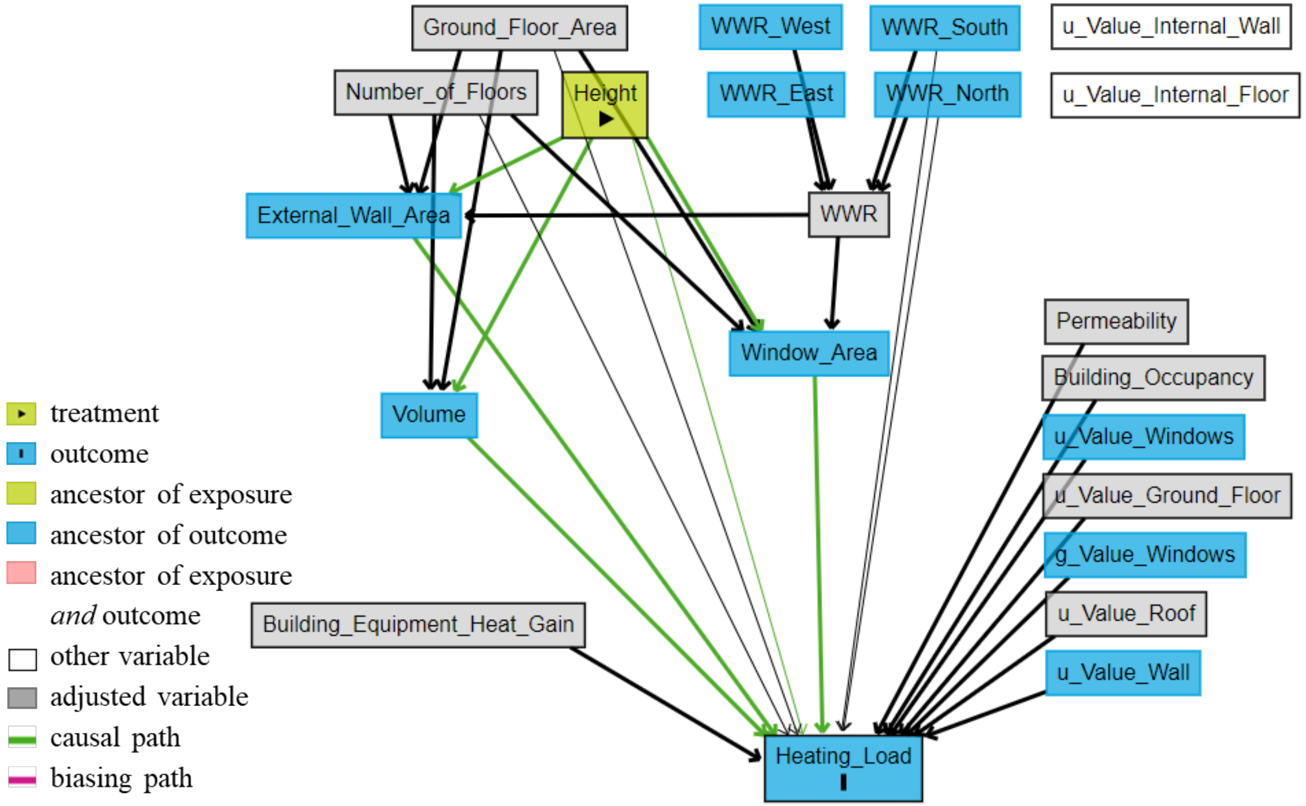


Suggestion to Scenario ii: To correctly estimate the total causal effect from ***[Height]*** to ***[Heating Load]***, *Window Area*, *External Wall Area*, and *Volume* **should not be adjusted** (controlled) to avoid biasing paths.

Causal effects quantification:
what if I had changed floor height from 3 meters to 3.2 meters?

Generated data with scenario condition

Parameter	*Unit*	*Value*
Ground Floor Area	m^2	300
Height	m	3 → 3.2
Number of Floors	-	3
External Wall Area	m^2	Unknown
u-Value (Wall)		Unknown
u-Value (Internal Wall)		Unknown
u-Value (Ground Floor)	W/m^2K	0.2
u-Value (Roof)		0.2
u-Value (Internal Floor)		Unknown
u-Value (Windows)		Unknown
g-Value (Windows)	-	Unknown
Permeability	m^3/m^2h	7.5
WWRs	-	0.3
Equipment Heat Gain	W/m^2	Unknown
Building Occupancy	Person/m^2	Unknown

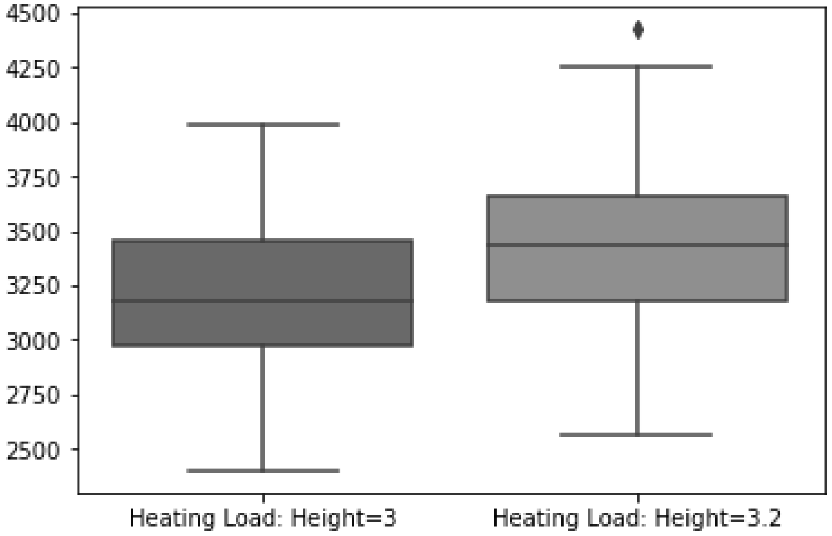


$$\tau = \mathbb{E}[\text{Heating Load}|\text{Height} = 3.2\text{m}, \boldsymbol{X}] - \mathbb{E}[\text{Heating Load}|\text{Height} = 3\text{m}, \boldsymbol{X}]$$

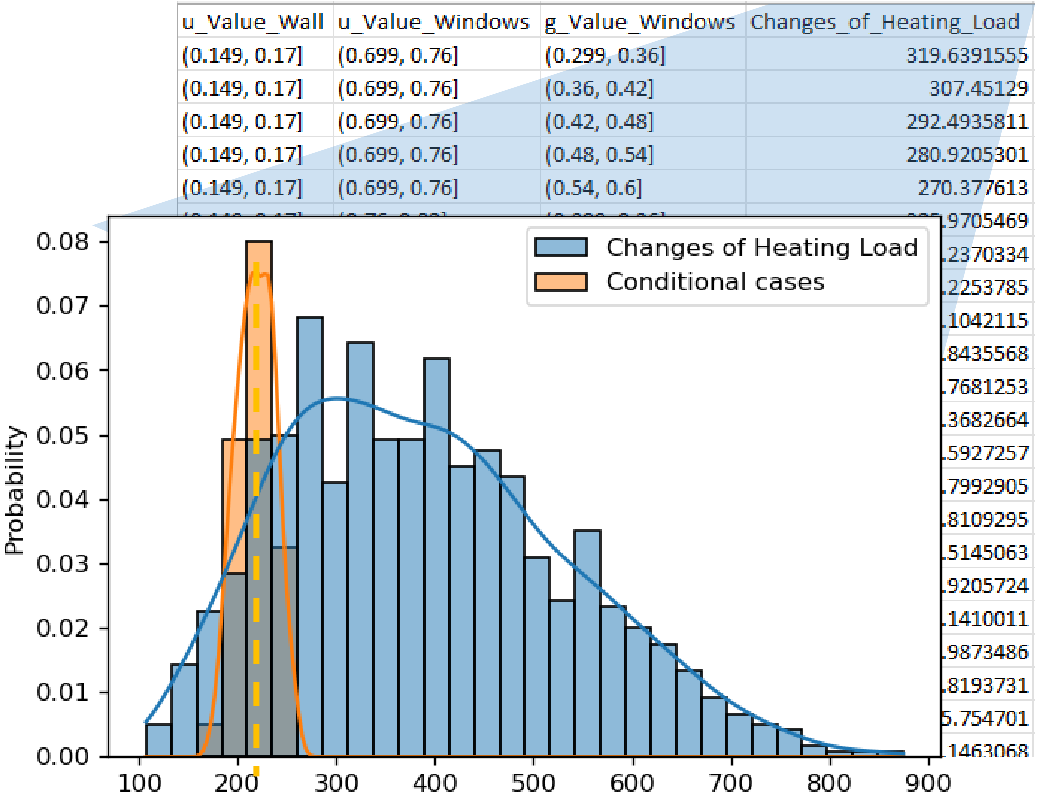
- ***Window Area, External Wall Area, and Volume*** *should not be adjusted*
- If we calculate the CATE, the $\boldsymbol{X}$ becomes the set of [*Building Equipment Heat Gain, Building Occupancy, Ground Floor Area=300, Number of Floors=3, WWRs=0.3, u Value Roof=0.2, u Value Ground Floor =0.2, Permeability=7.5*]

Causal effects quantification:
what if I had changed floor height from 3 meters to 3.2 meters?

Exhaustive simulations result (200 runs)



Causal model result



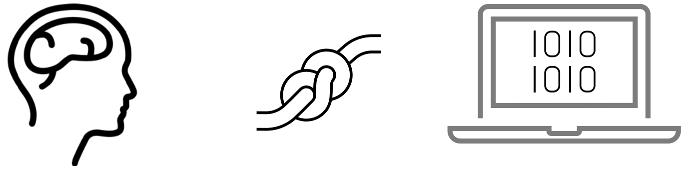
Accuracy performance of three typical data-driven methods for predicting heating load

	MAPE	R^2
LightGBM	6.972 %	**0.924**
RF	11.016 %	0.81
ANN	13.152 %	0.746

Changes of Heating Load, CATE	**kWh/year**
CATE based on simulations	**218.52**
CATE based on causal model	**207.28**
CATE based on pure ML model (LightGBM)	**47.24**

Output: If the treatment variable **[*Height*]** increases from **3** to **3.2 m** based on Table 3 condition, causes an increase of **207.28 kWh/year** in the direct expected value of the outcome **[*Heating_Load*]**

Key takeaways



1. Parametric dependency check is important.

2. An analogy between personal **experience** and physical **knowledge** provides a channel for integrating data-driven and knowledge-based methods through causal DAGs. This separation would achieve a fast cross-sectional examination and avoid conducting erroneous conclusions.

3. Causal model provides a **data-driven knowledge extraction method** for design process analysis with reduced computational difficulty; The causal model allows users to quickly check potential design alternatives in a higher dimension.

4. We clarify the **boundary of design assistance** based on DAGs. The growth of DAG with reduced uncertainties aligns with the nature of the design process.

5. A **four-step framework** is proposed to implement causal inference into the design domain with **causal structure finding** and **causal relationship quantification**.

Leibniz
Universität
Hannover

Thank you!
Questions?

xia.chen@iek.uni-hannover.de
chenxiachan.github.io

Contact & More research insights