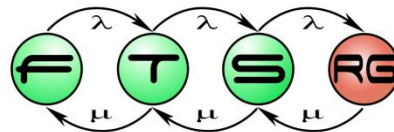



# From measurements to models

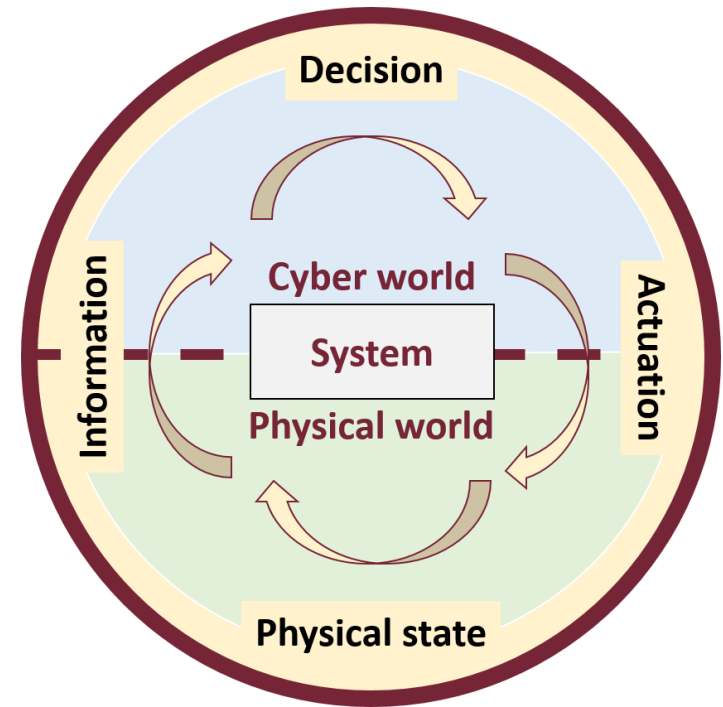
**András Pataricza-András Földvári**

**Budapest University of Technology and Economics  
Fault Tolerant Systems Research Group**

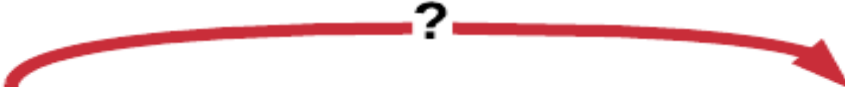


# Context

- Cyber-Physical Systems (CPS)
    - Component integration
    - Extra-functional requirements
  - Validation
    - Benchmarking
    - Operational log analysis
- 
- Big data sets
    - Hard-to-interpret
    - Many dimensional

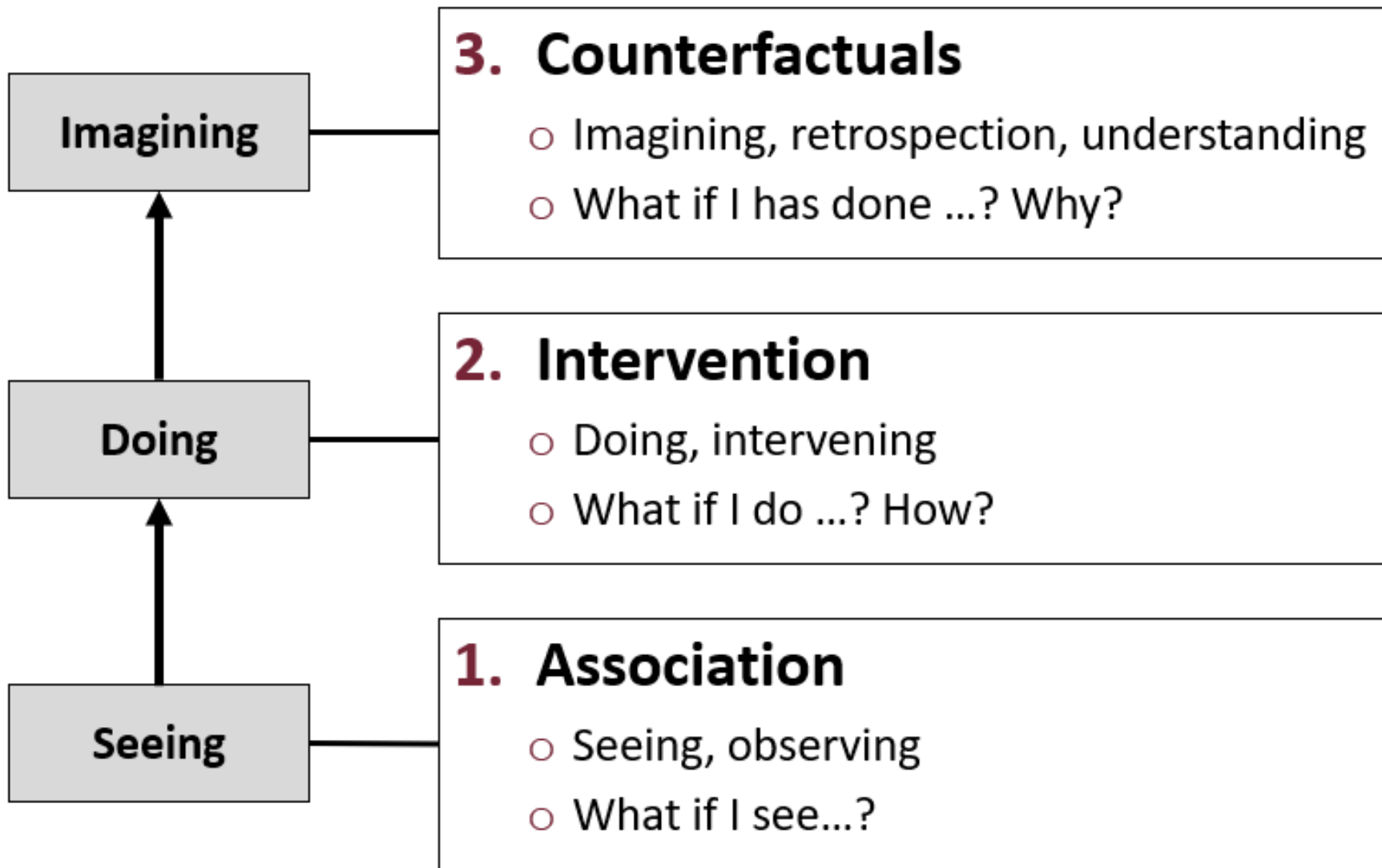


**EDA**  
**?**

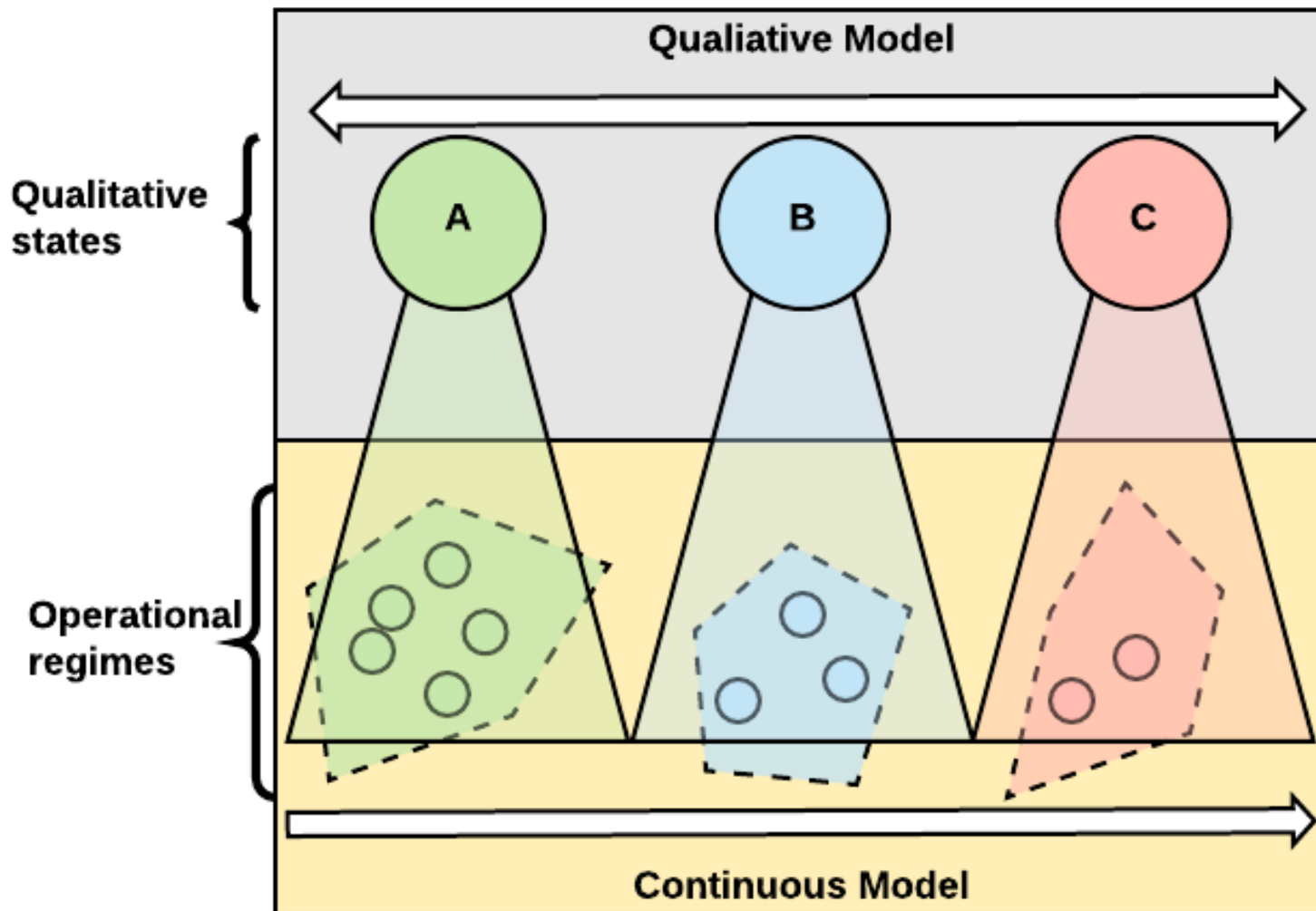


	ip	RT	RPT	RTT	location	client.type	DC	start.time	pm.pa	Country
0	10.8.146.1	2656	1031	1625	Newcastle	Microsoft	Redmond	1.36E+09	PM	UK
1	10.8.146.1	1469	719	750	Newcastle	Microsoft	Redmond	1.36E+09	PM	UK
2	10.8.146.1	1984	844	1140	Newcastle	Microsoft	Redmond	1.36E+09	PM	UK
3	10.8.146.1	2437	734	1703	Newcastle	Microsoft	Redmond	1.36E+09	PM	UK
4	10.8.146.1	1859	750	1109	Newcastle	Microsoft	Redmond	1.36E+09	PM	UK
5	10.8.146.1	2344	875	1469	Newcastle	Microsoft	Redmond	1.36E+09	PM	UK
6	10.8.146.1	1625	765	860	Newcastle	Microsoft	Redmond	1.36E+09	PM	UK
7	10.8.146.1	2281	672	1609	Newcastle	Microsoft	Redmond	1.36E+09	PM	UK
8	10.8.146.1	2640	641	1999	Newcastle	Microsoft	Redmond	1.36E+09	PM	UK

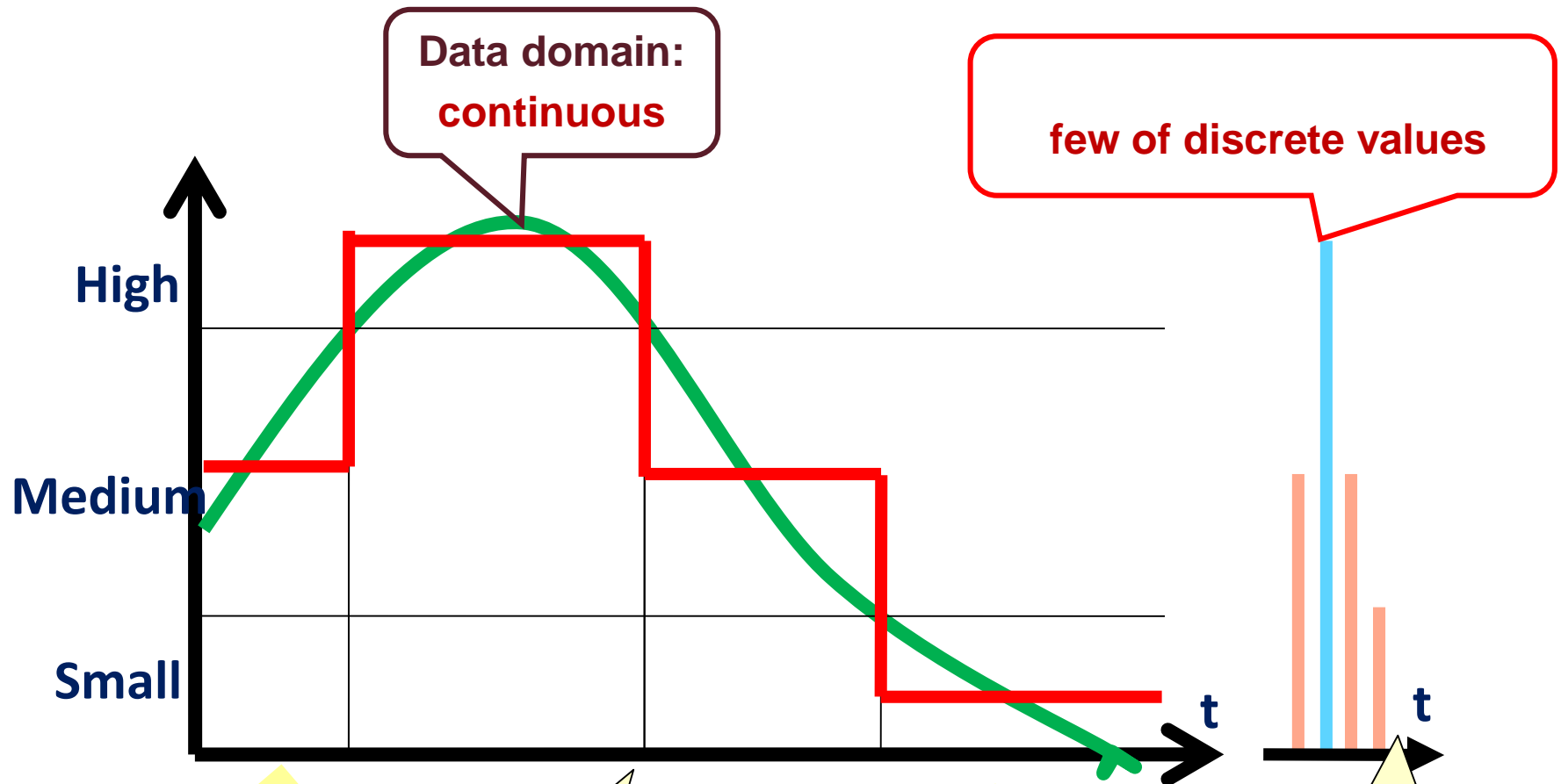
# Ladder of Causation (Judea Pearl)



# Modeling paradigm: hybrid modeling



# Core idea of qualitative modeling



Data domain:  
continuous

few of discrete values

Medium

High

Medium

Small

Spatial abstraction:  
continuous → enumeration

Temporal abstraction: Stabilizing  
Peak: High

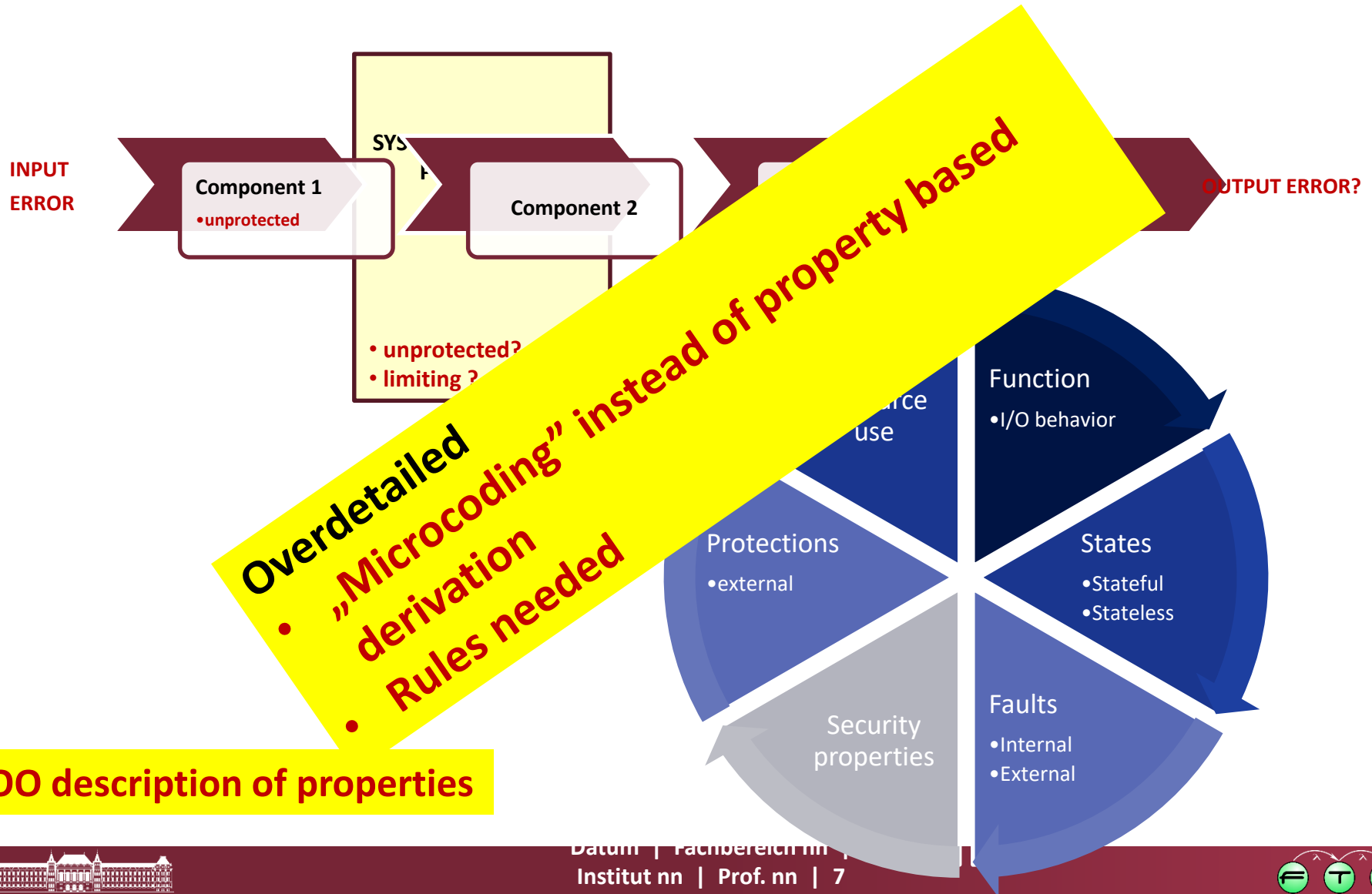
# E.g. fault model

- Input / output mapping → from risk model (?)
- Web service can be fault-free or fault-free if it uses server1 as a resource.

web service	in	fty	r		
	good	good			
	good	good			x
	good				404
	good	fty			x
	fty	fty			x
		fty	fty		x
		ins	good	fty	x
		incons	fty	fty	x
		good	good	x	x
		good	fty	x	x
	x	incons	good	x	x
	x	incons	fty	x	x

**Overdetailed  
"Microcoding" instead of property based  
derivation**

# System level impact analysis



# Model hierarchy by discretization

- **Discrete : Different operational domains → States, qualitative I/O values**
  - Clustering: domain of fundamentally similar behaviors
- Continuous
  - Intradomain variables / configuration parameters
- Domain boundaries
- Dimensioning → occurrence of some domains (e.g. saturation)
- Metamodel for all configurations :→ Single integrated model covering all cases (conditioned by the dimensioning)



# Use cases

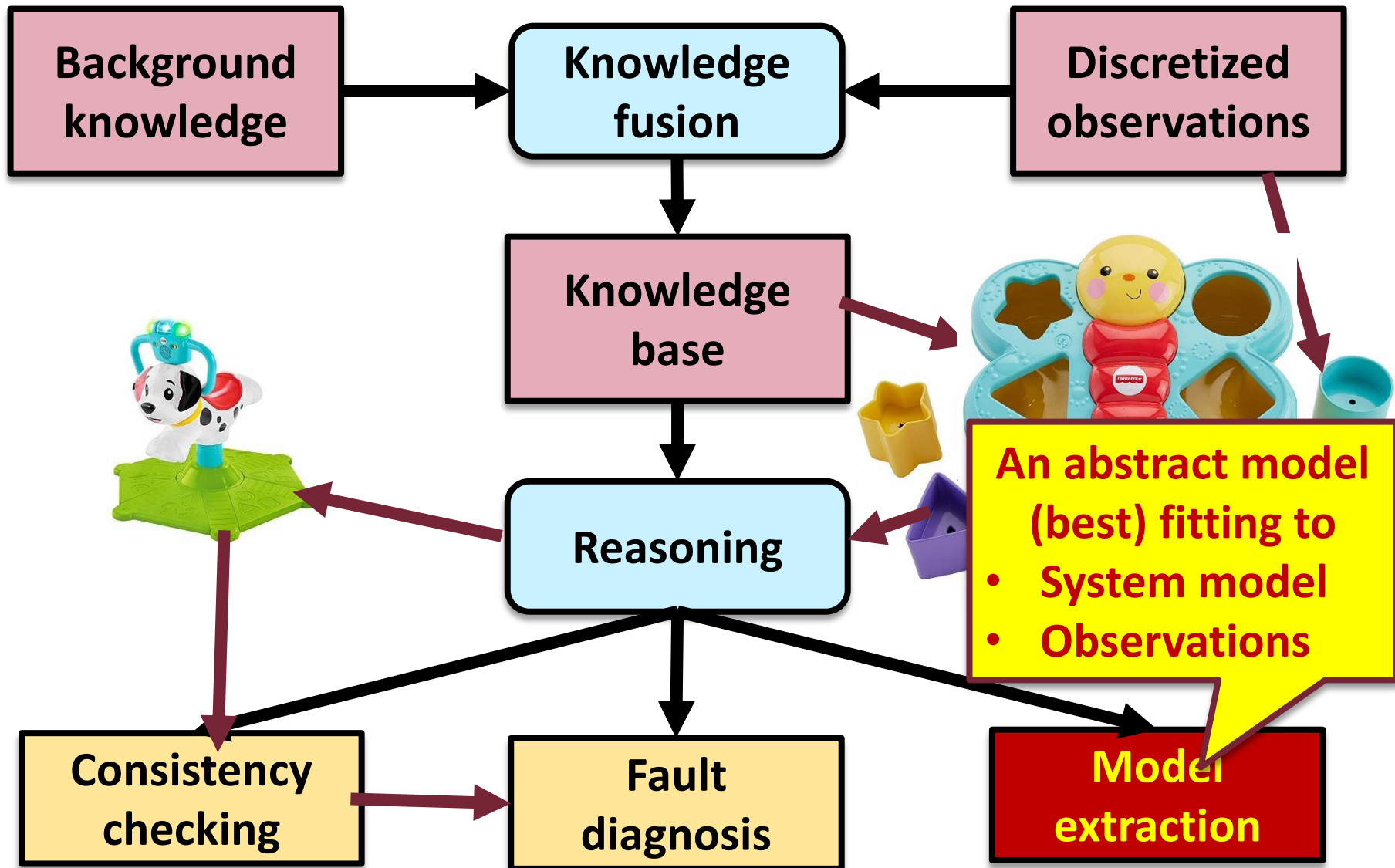
## ■ Design phase

- guideline for **dimensioning** (design pattern selection, category selection for the resources)
- V&V: **proof of correctness**/control policy  
Huge repertoire of formal methods

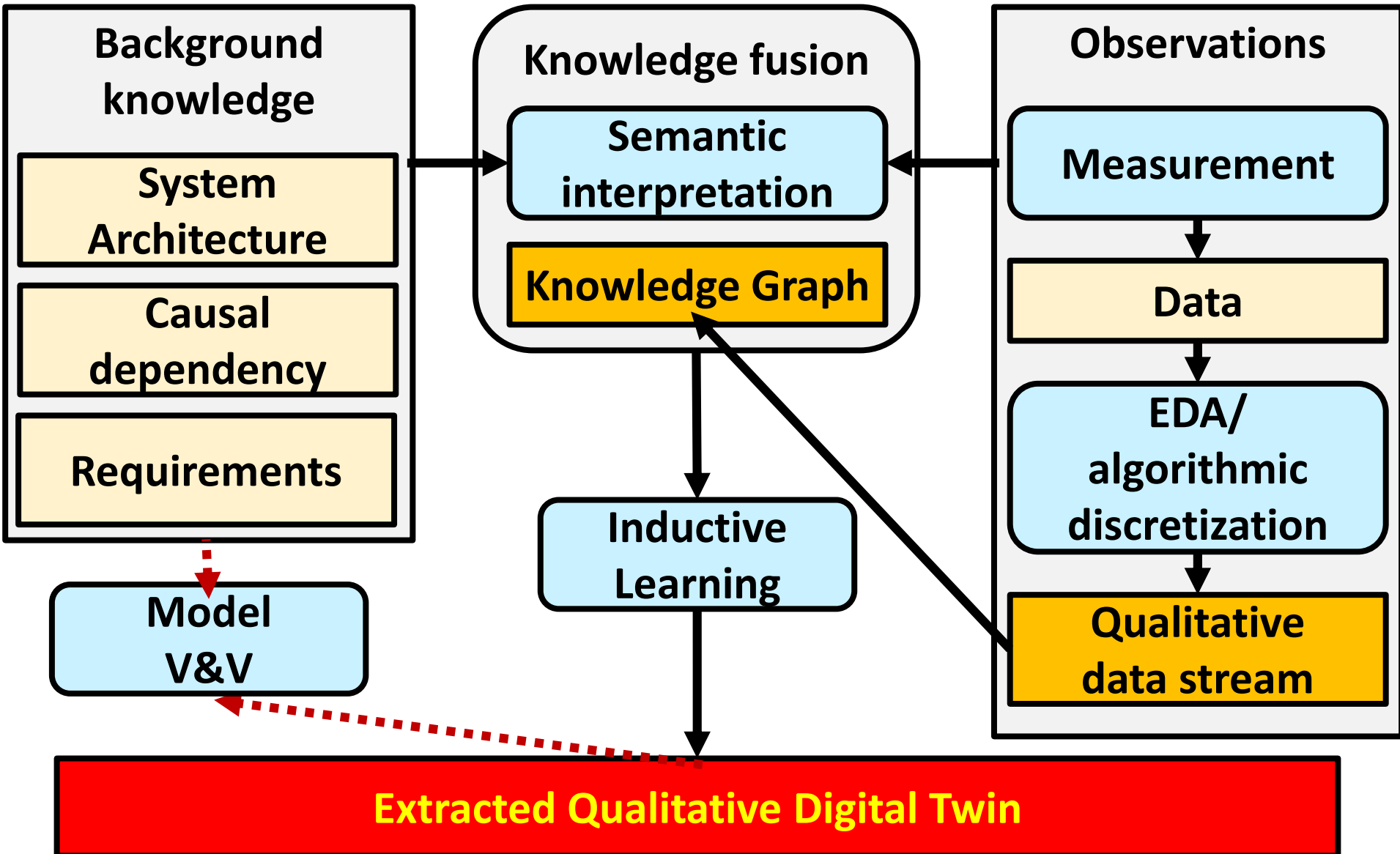
## ■ Operation phase

- **Digital twin**: monitoring and supervisory control
  - Continuous → discrete:  
classification of the **operation state**
  - Discrete: **reaction design**/impact assessment
- Discrete to continuous: **dimensioning of rules**  
(e.g. size of additional resources)

# Automated model extraction

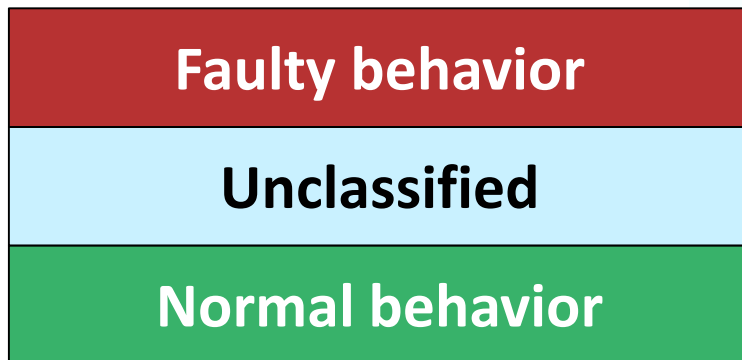
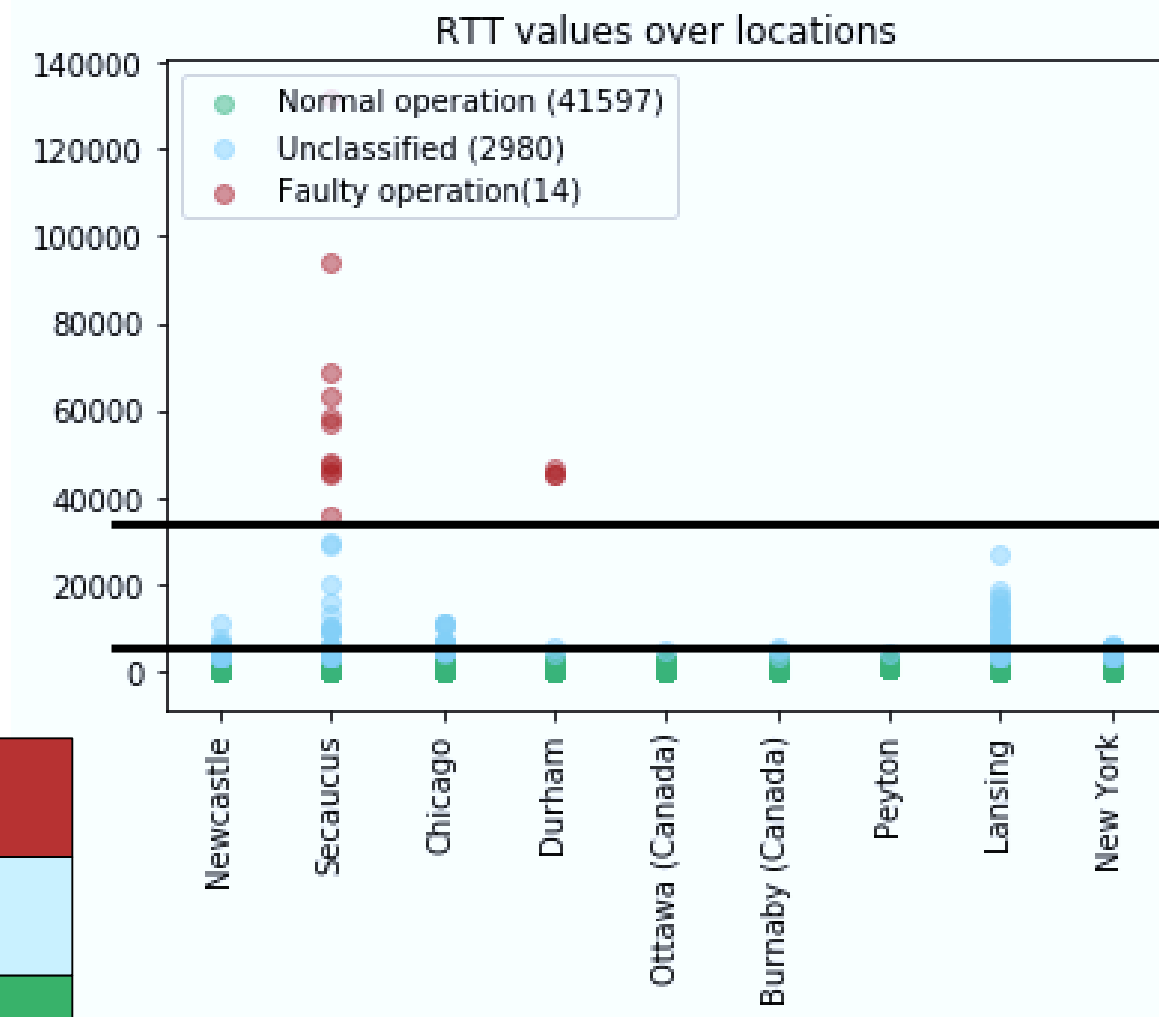


# Method: Steps



# Discretization: operational modes

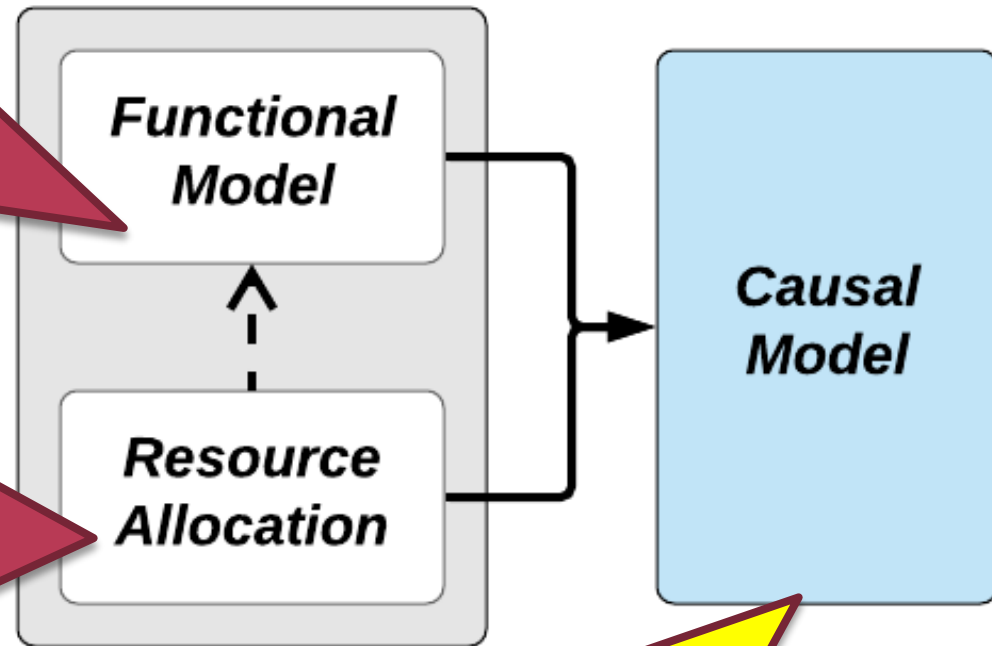
- Homogenous domains
  - Similar behaviors
- Visual EDA
  - Thresholds
- Clustering techniques
  - K-mean
- 3 operational domains



# Causal model building

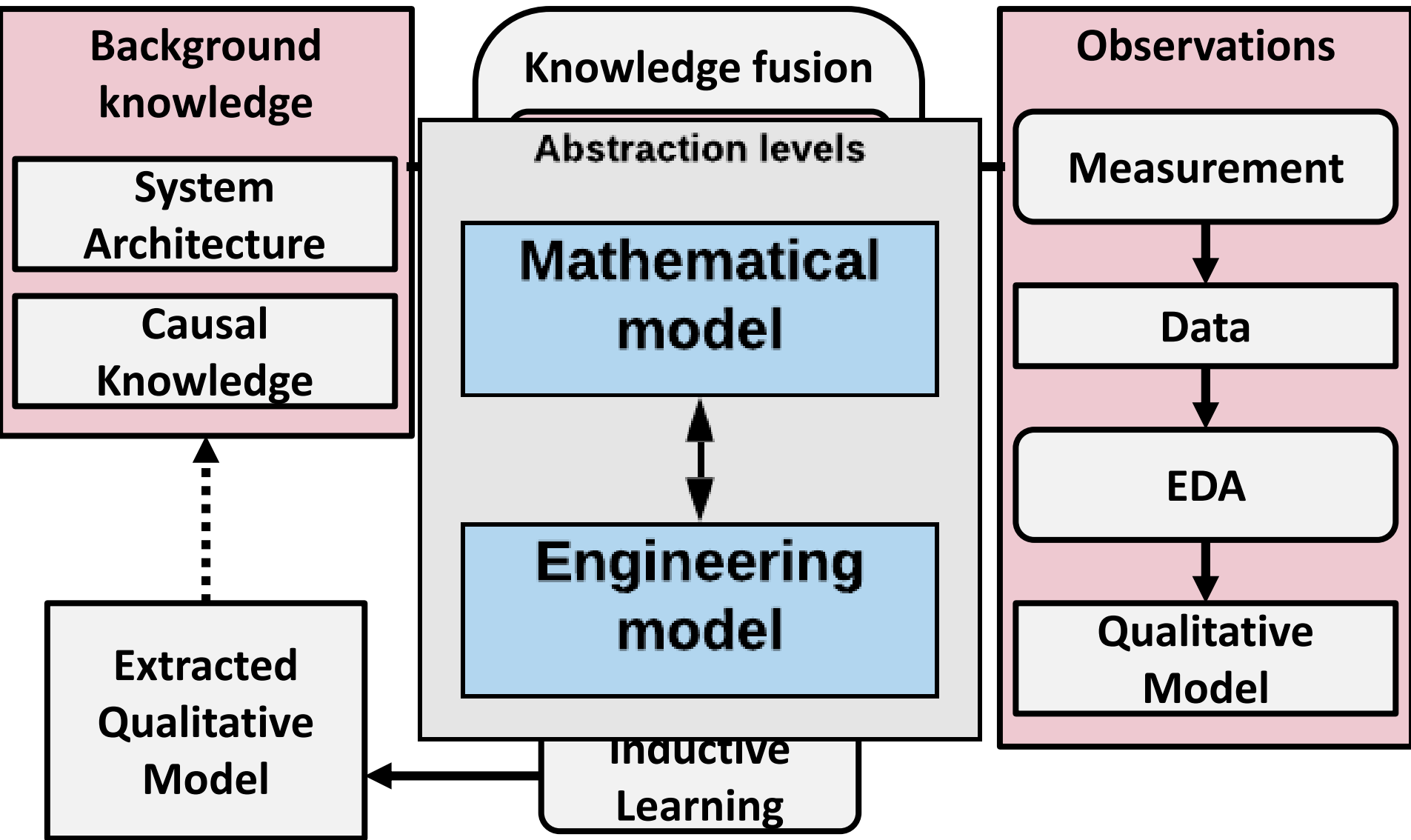
- Continuous processes
- “Skeleton” of the causal model

- Detailed causal relations
- E.g., shared resources



- Data propagation
- Hidden relationships (e.g., shared resources)
- Filtering false associations

# Semantic interpretation



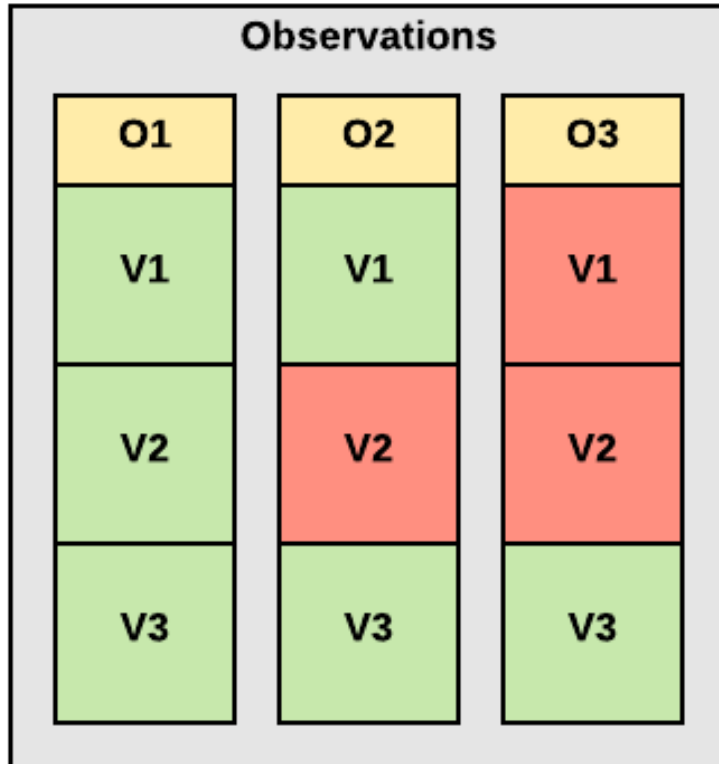
# Knowledge graph

- Information representation
- Knowledge fusion
- Reasoning

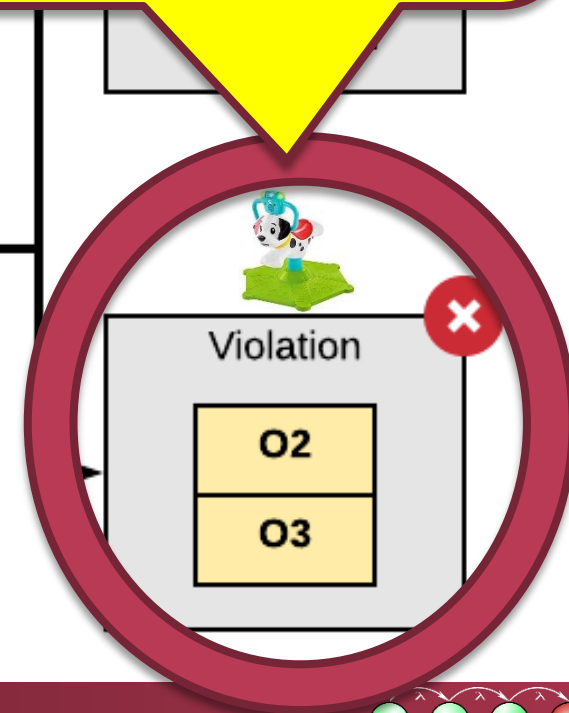
**Mismatch:**

**Digital twin  $\leftrightarrow$  observation?**

- Faulty operation
- Noisy inputs
- Modeling faults



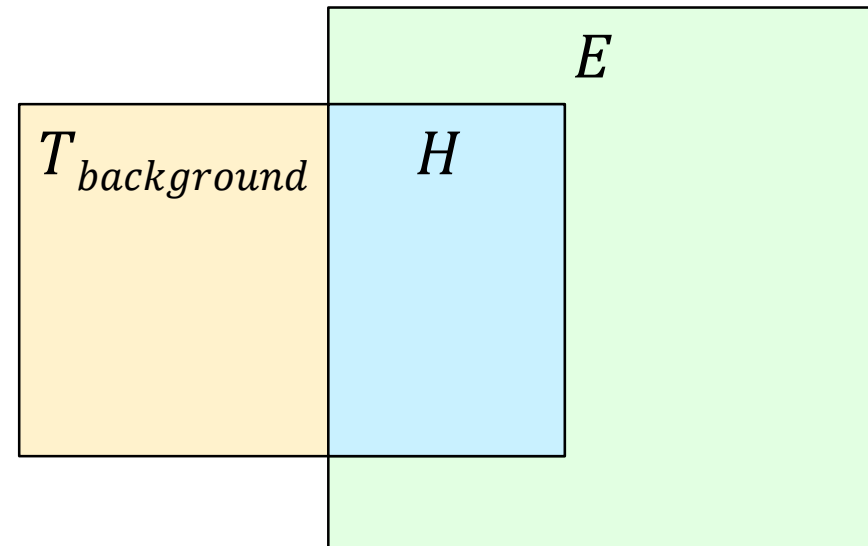
Constraints  
(schema, rules)



# Inference goal: a hypothesis (H)

- Hypothesis=abstract model
- Explains all the selected outcome variables (E)
- by reducing the included condition variables (background knowledge) ( $T_{background}$ ).
- Excluding contradictory data

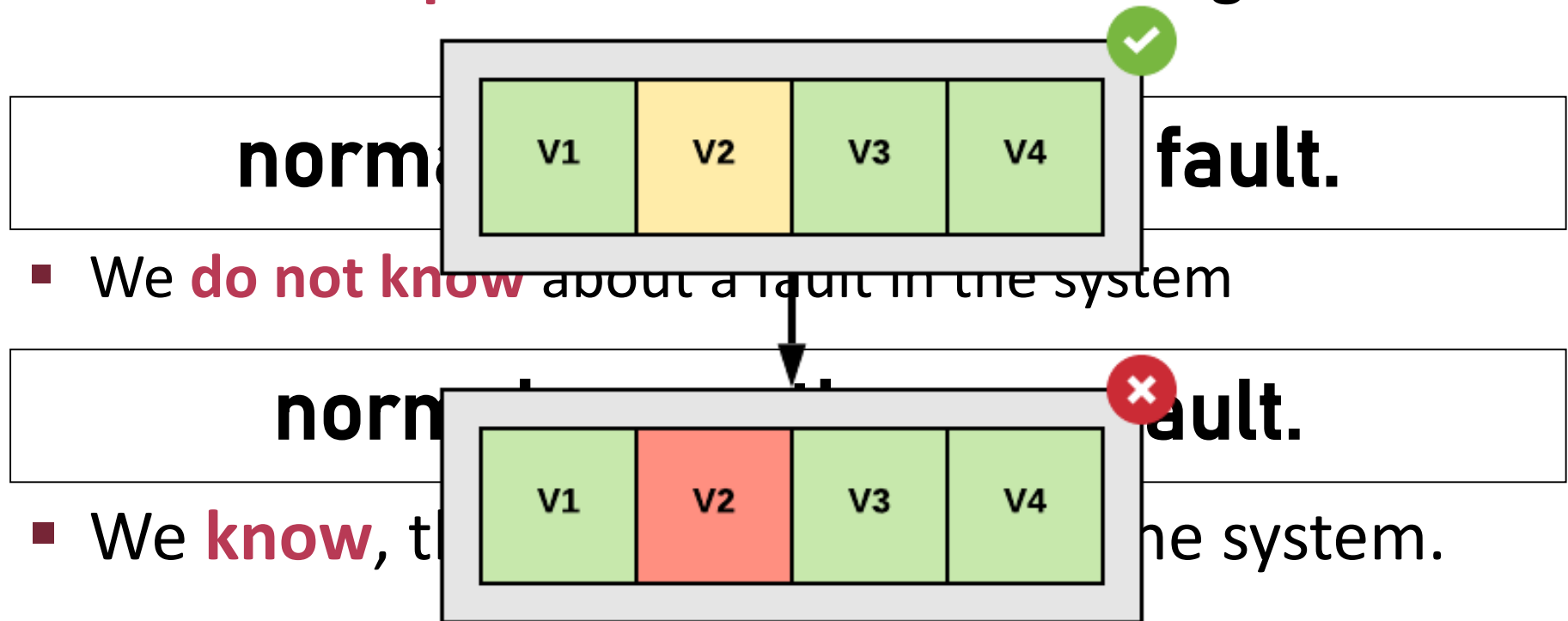
	Explain	Include for reduction	Exclude
Outcome 0	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>
Outcome 1	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>
Contradictions	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>





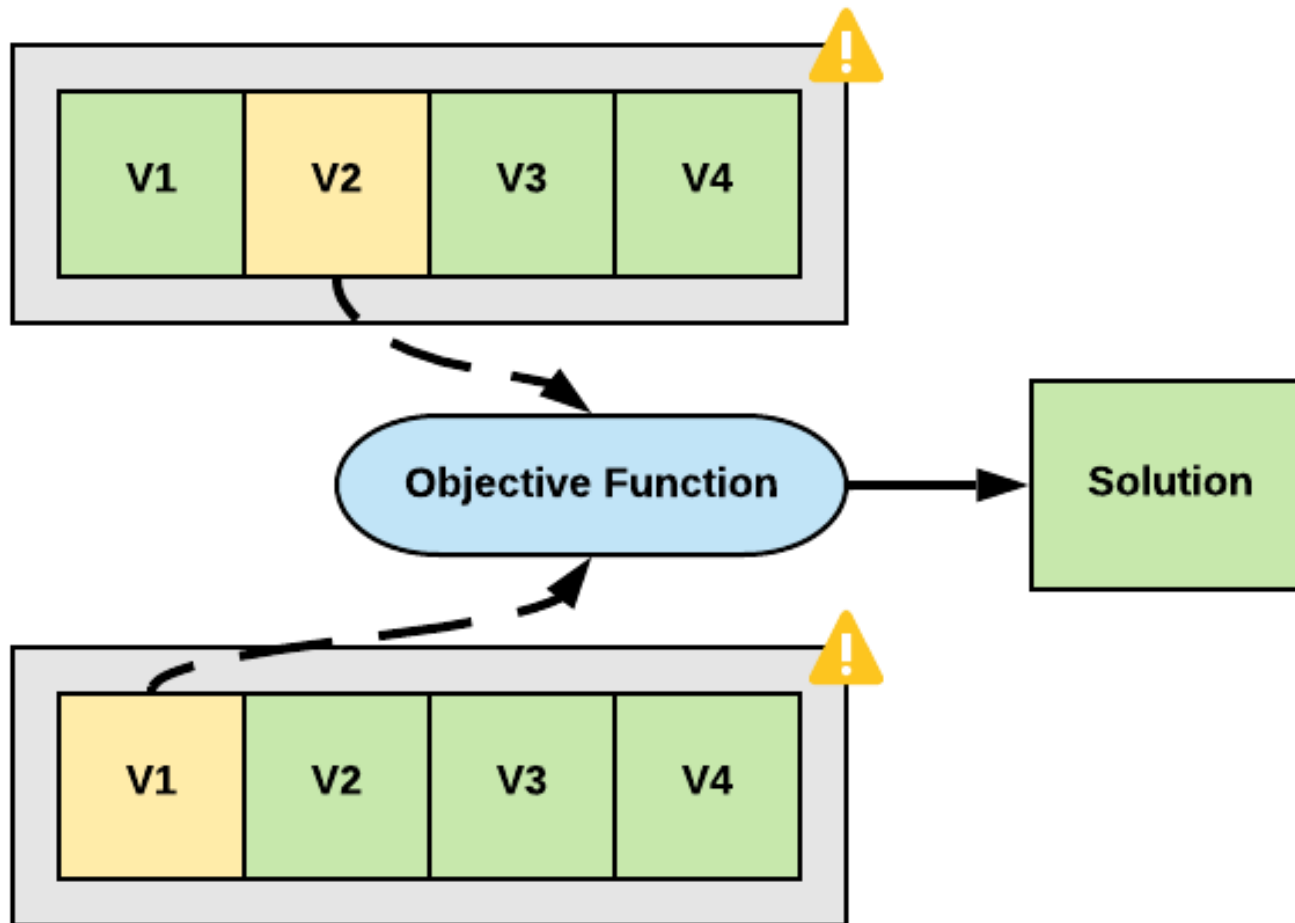
# Incomplete knowledge

- Answer Set Programming (ASP)
  - Stable Model Semantics
    - Logic program + **negation as failure**
    - **Incomplete** and **default** knowledge

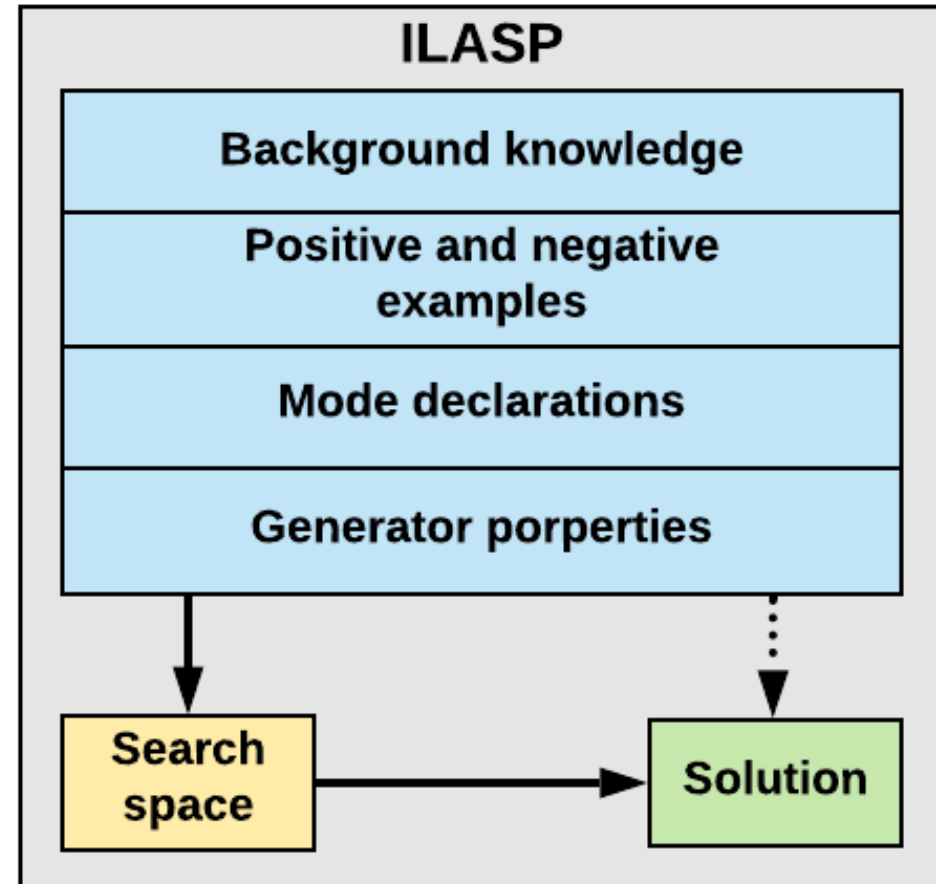
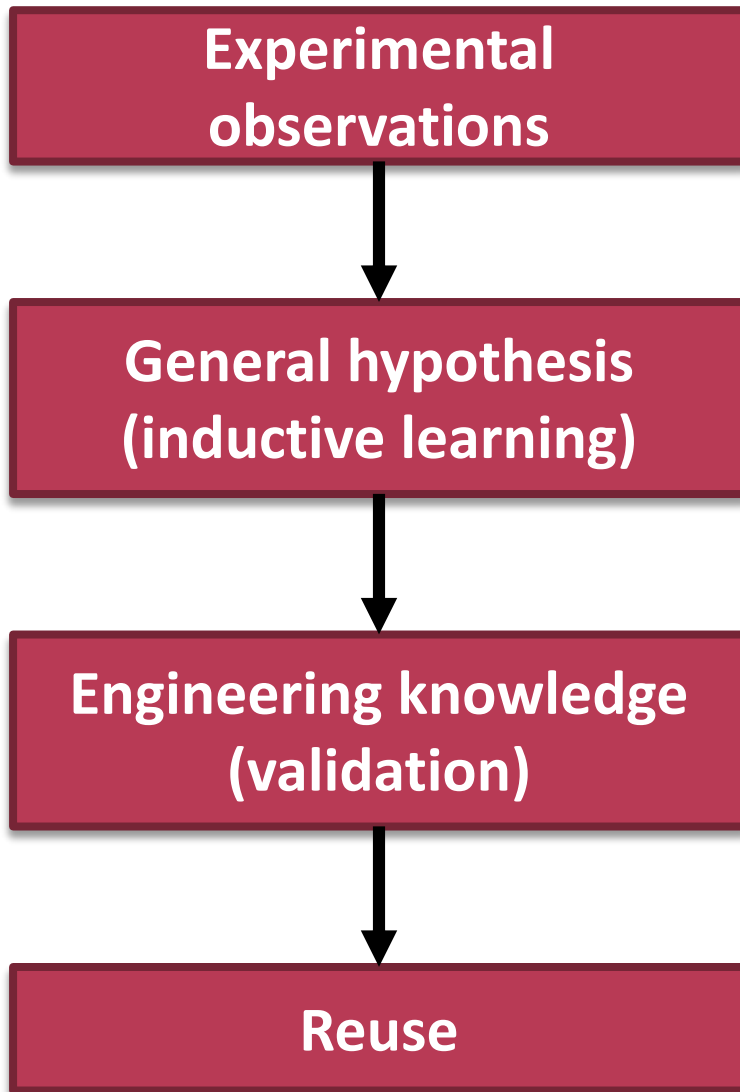


# Weak constraints

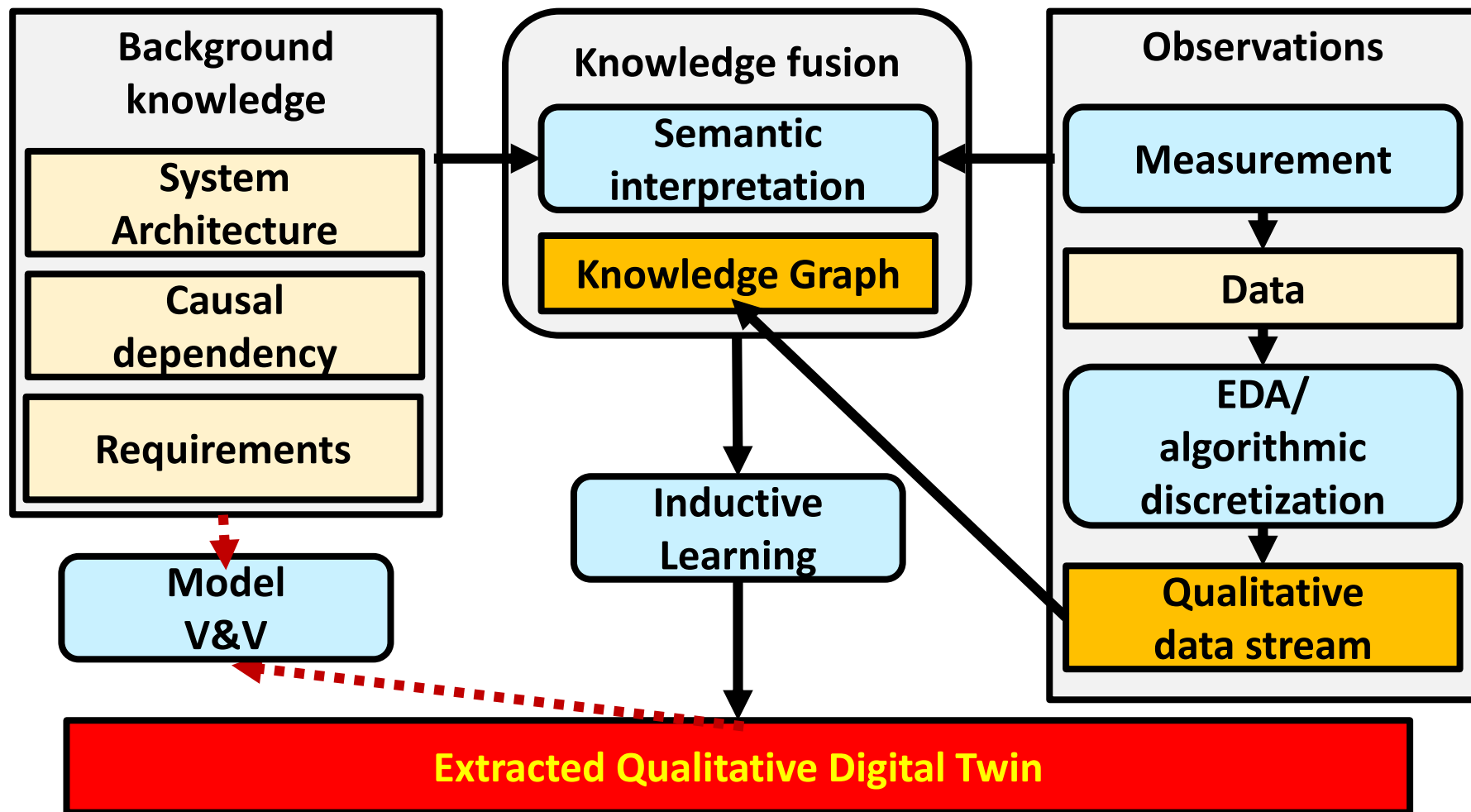
- Optimization problems



# Inductive learning



# Summary



The results presented in this presentation were established in the framework of the professional community of Balatonfüred Student Research Group of BME-VIK to promote the economic development of the region. During the development of the achievements, we took into consideration the goals set by the Balatonfüred System Science Innovation Cluster and the plans of the "BME Balatonfüred Knowledge Center", supported by EFOP 4.2.1-16-2017-00021.)