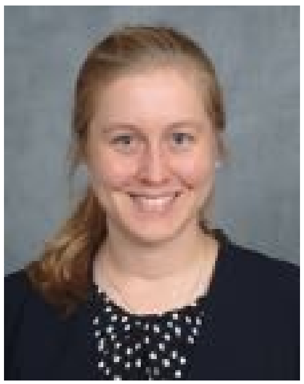
On Learning to Steer Buildings (and their occupants) Towards Greater Societal Value

Mario Bergés
Professor @ Carnegie Mellon University
Scholar @ Amazon

EPFL, 6.11.2025

https://inferlab.org

A group effort

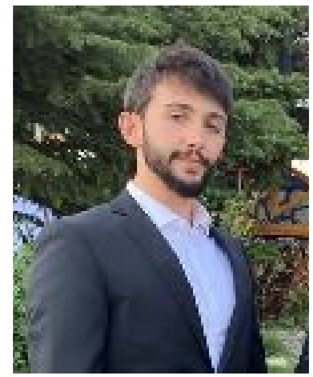


Prof. Flanigan
CMU, CEE

M. Dcotorarastoo
CMU, CEE

Sizhe Ma
CMU, CEE

Kieran Elrod
CMU, CEE



Ozan Mulayim
CMU, CEE
Soon at Google

Bingqing Chen
CMU, CEE
Now at Bosch AI

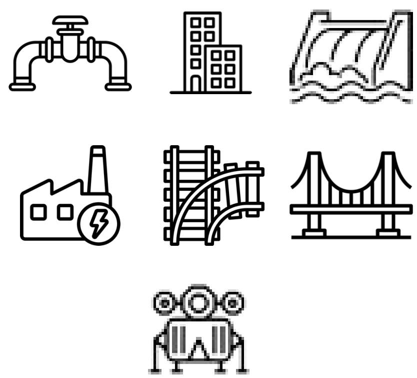
FLANIGAN SALUS LAB

INFERLab
Intelligent Infrastructure Research Laboratory

INFERLab

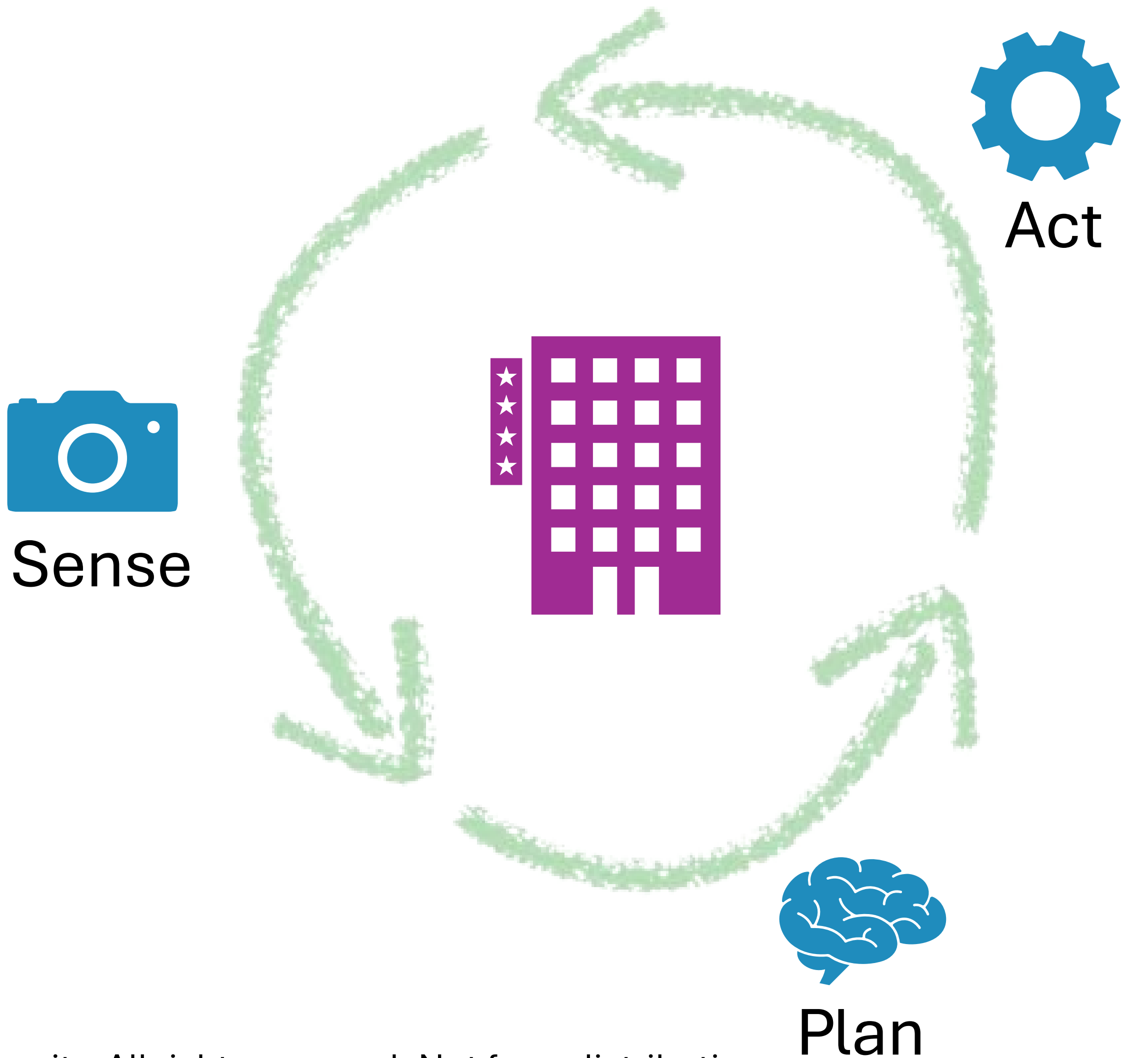
Intelligent Infrastructure Research Laboratory

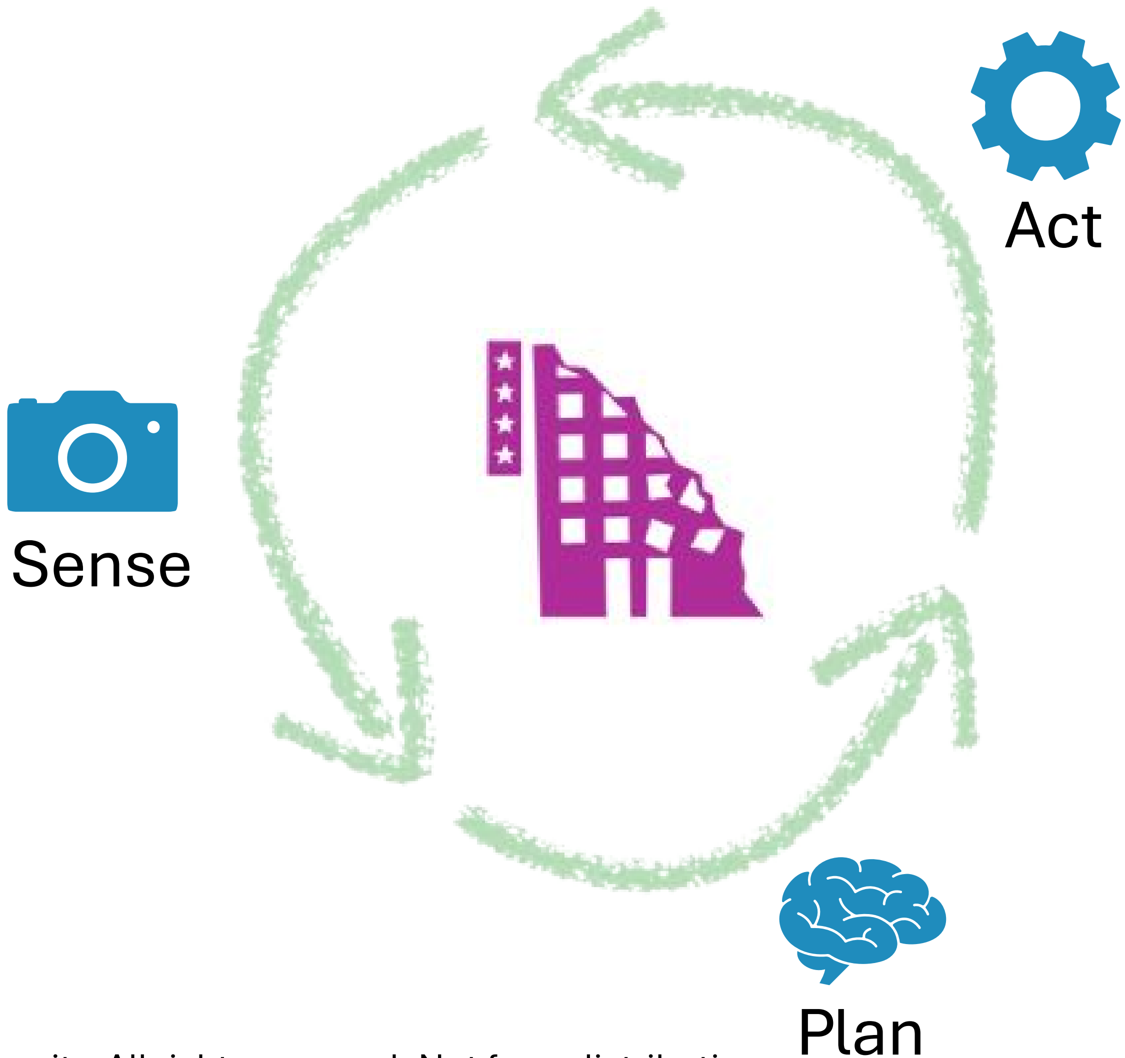
Improving the design and operation of civil infrastructures systems through data-driven solutions grounded on engineering knowledge.



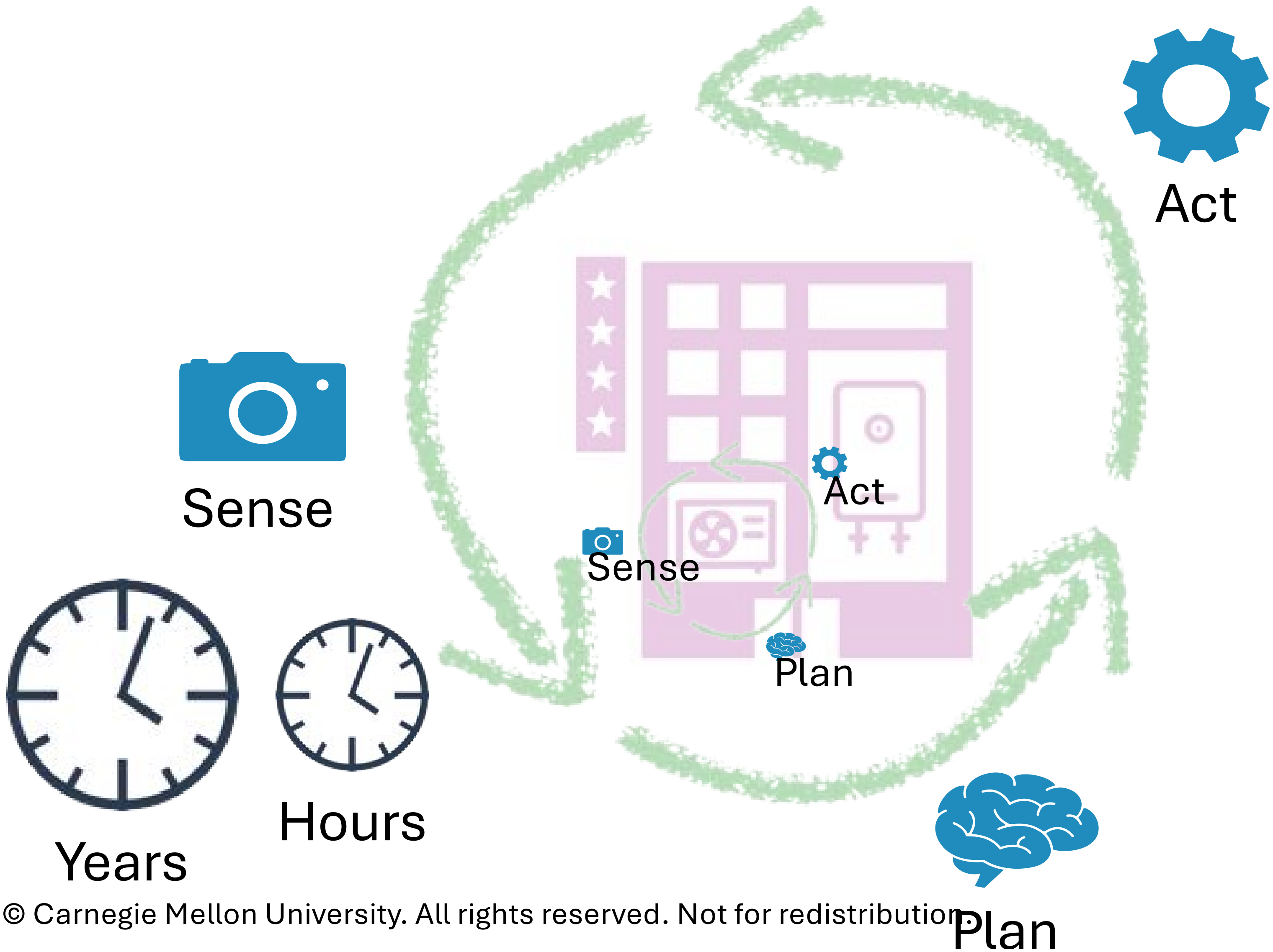
We spend over 90% of our time in *designed* environments

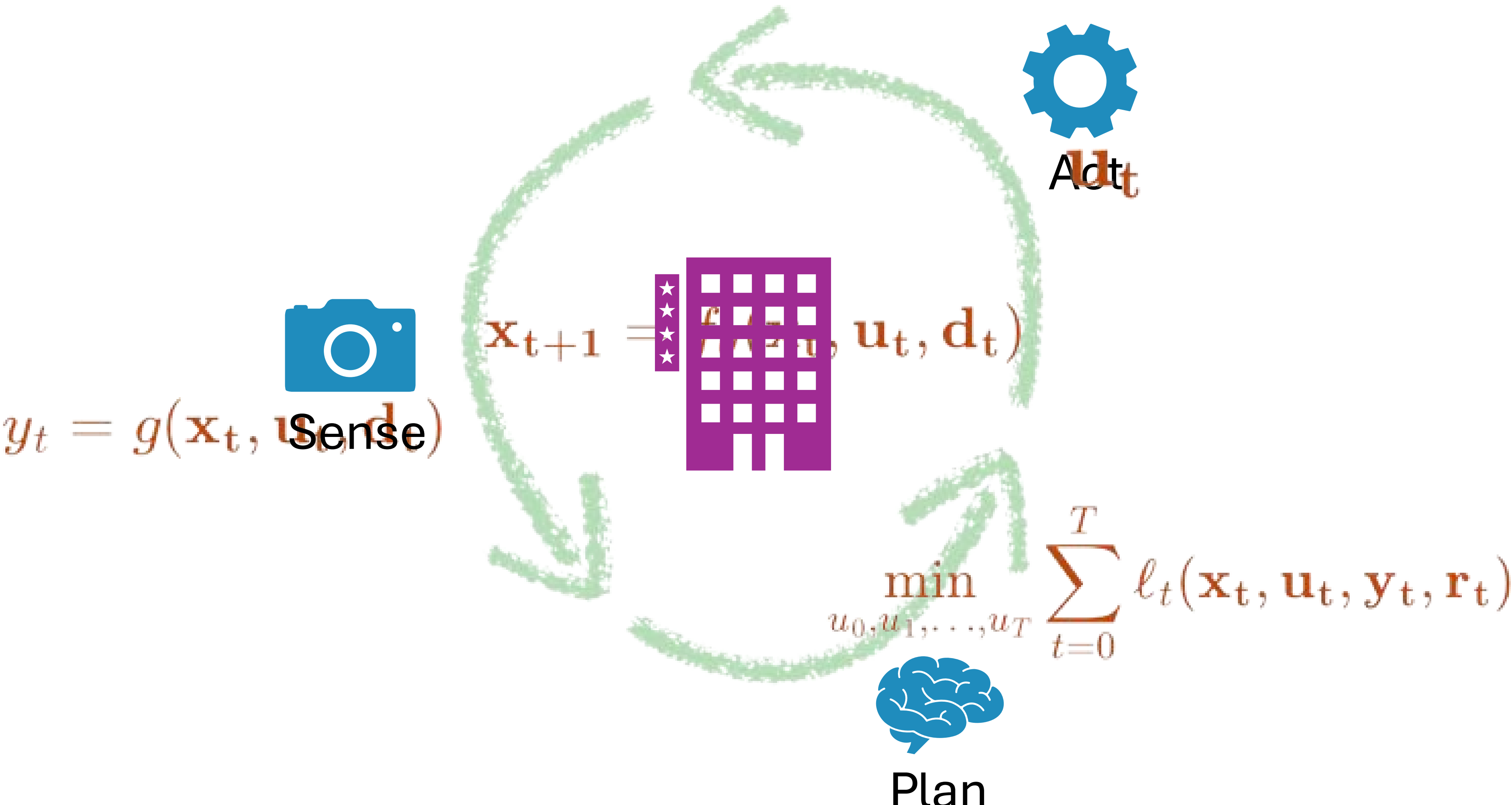
What do we design them for and how good are we at it?


Act
Sense
Plan


Act
Sense
Plan


Years


Act
Sense
Act
Sense
Plan
Years
Hours
Plan


Act
$\mathbf{u}_t$
$\mathbf{x}_{t+1} =$
$\mathbf{u}_t, \mathbf{d}_t)$
$y_t = g(\mathbf{x}_t,$
Sense
$\min_{u_0, u_1, \ldots, u_T} \sum_{t=0}^{T} \ell_t(\mathbf{x}_t, \mathbf{u}_t, \mathbf{y}_t, \mathbf{r}_t)$
Plan

What objectives can we write down? Which ones can we solve?

- We're limited by (at least)
  - The observations we have access to $\mathrm{y}(t)$
  - The states we can infer from them x_t
  - The dynamics we can learn $f_\theta(x_t, u_t)$

$$\min_{u_0, u_1, \ldots, u_T} \sum_{t=0}^{T} \ell_t(\mathbf{x_t}, \mathbf{u_t}, \mathbf{y_t}, \mathbf{r_t})$$

- Can we write down an $\ell(\ )$ function for:

Well-being, health, and opportunity

Resilience to disruption

Social capital formation


Let’s look at the case for designing your thermal comfort experience
Closed loop control over relatively short horizons and for simple objectives

Hey Siri, can you optimize my temperature while keeping my costs low?

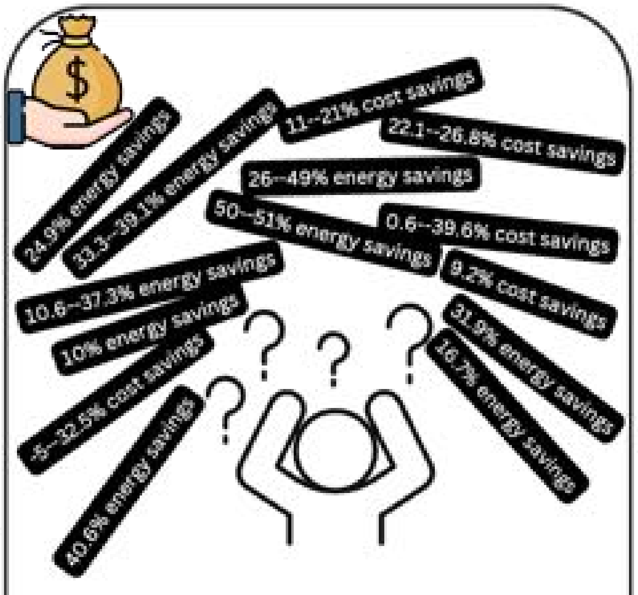
Model Predictive Control

Reinforcement Learning

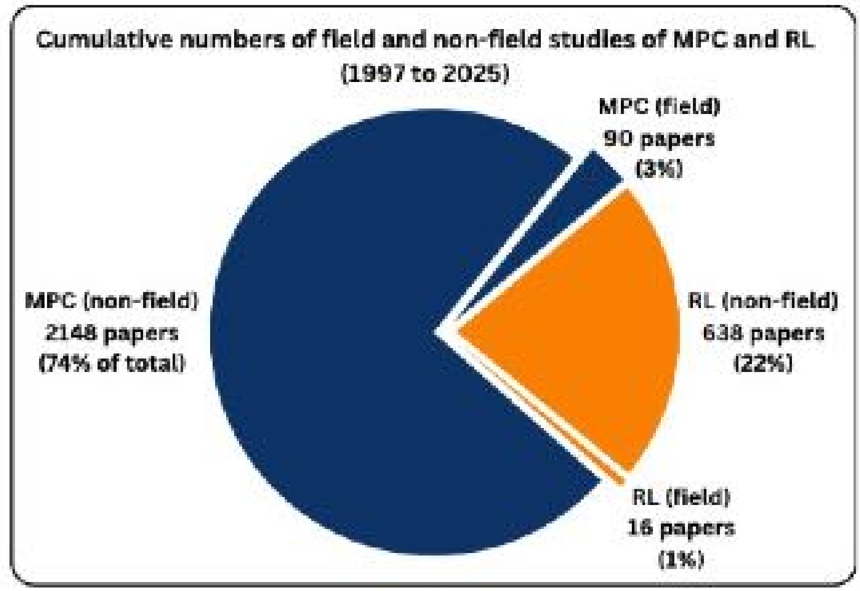
What is a practical and scalable solution for building control?

$$\ell = \sum \left(\lambda \left|\left| x_t - x^* \right|\right|_2^2 + \left|\left| u_t \right|\right|_1 \right)$$

Where are the proactive controllers?



Many promises have been made.



Many simulation and some real-world experiments have been done.

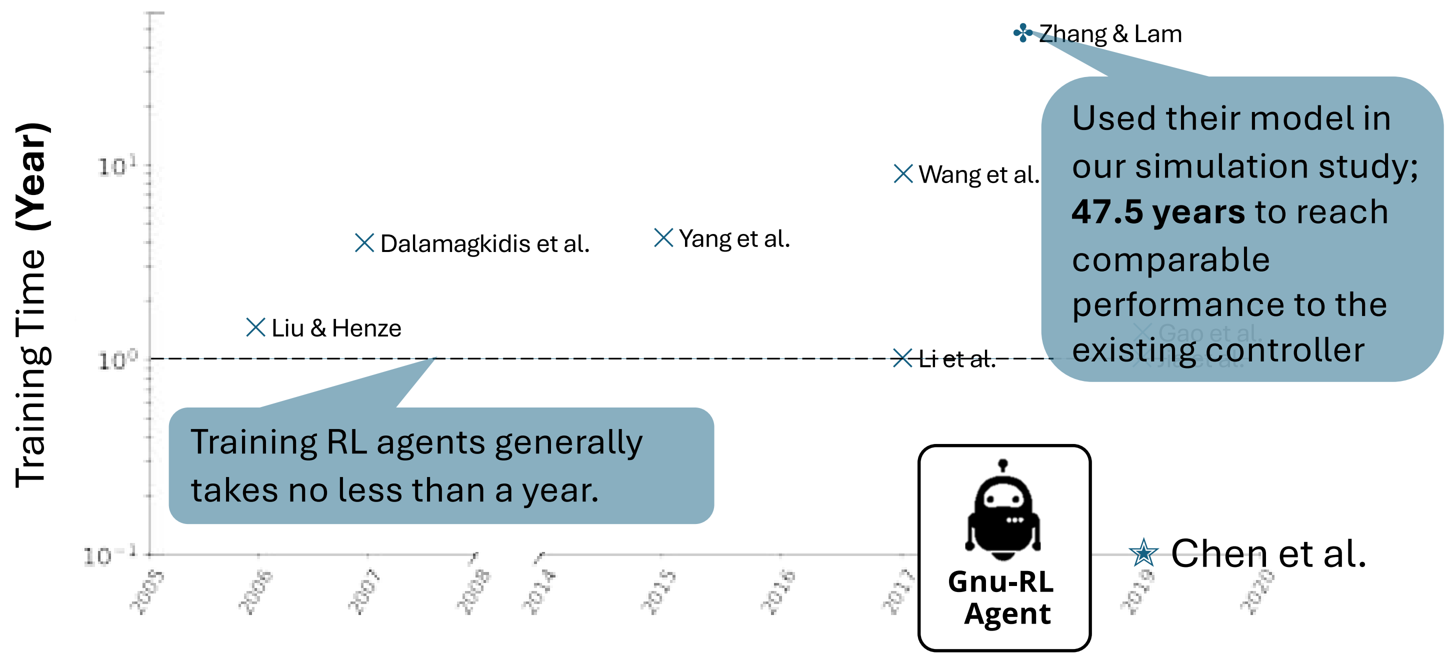
Khabbazi et al. (2025) Lessons learned from field demonstrations of model predictive control and reinforcement learning for residential and commercial HVAC: A review

What do the existing studies lack?

Checklist	CLUE(2024) DPC (2022) MB2C(2020)	DeePC (2024)	Model-free RL (2016,2020,2022)	Gnu-RL (2020)
“Scalable”	Data-hungry	Extensive tuning	Needs a simulator for training	Manual fitting of cost parameters
Interpretable	Black-box	Black-box	Black-box	Limited interpretability
Adaptive	No online learning	No online learning	Online learning	Online learning
Safe	Only in simulation	Real world experiment (5 days)	Real world experiment (13 days)	Real world experiment (21 days)

Most studies are validated only in simulation, black-box in nature or not adaptive.

Training Time in Literature for RL Control of HVAC Systems



Chen, Berges, Cai, *ACM BuildSys (2019)*

We expedited the training by:

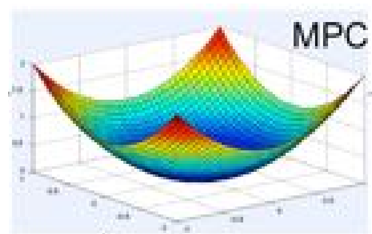
Imitation Learning

Domain Knowledge

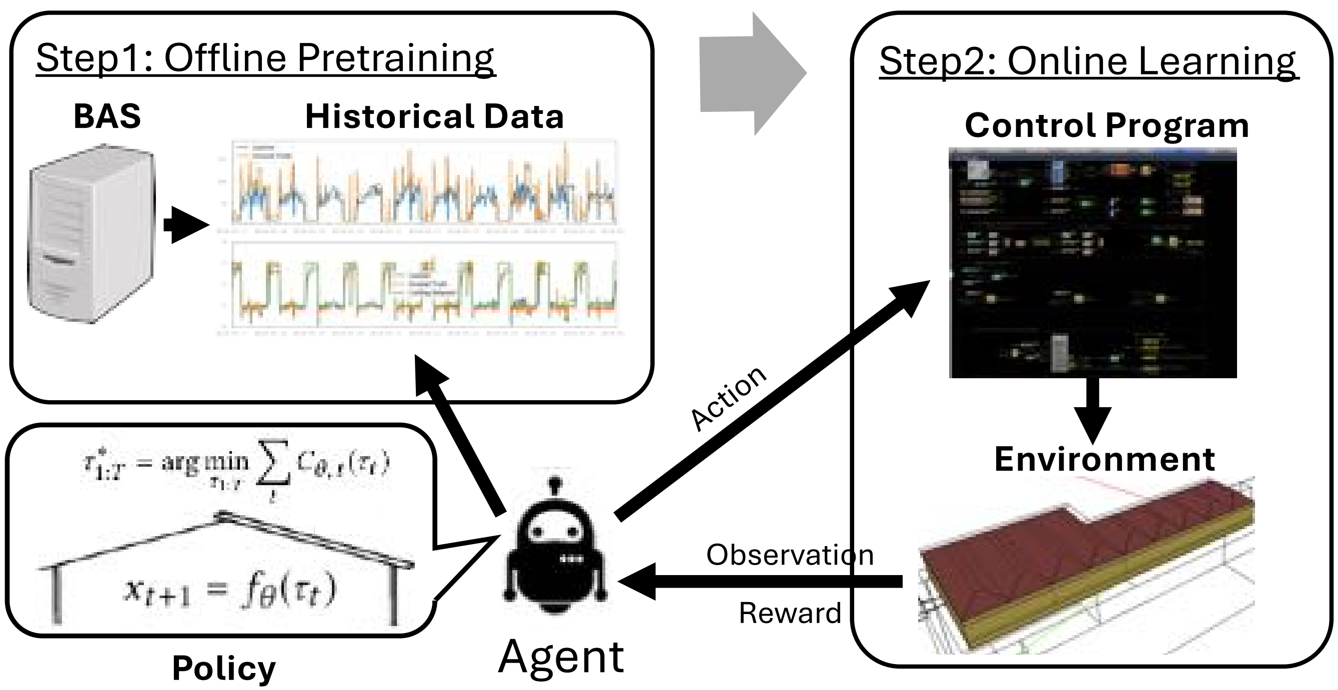
A physical-based model:

$$m\frac{dT}{dt} = Q_{\text{internal}} + Q_{\text{external}}$$

Model Predictive Control

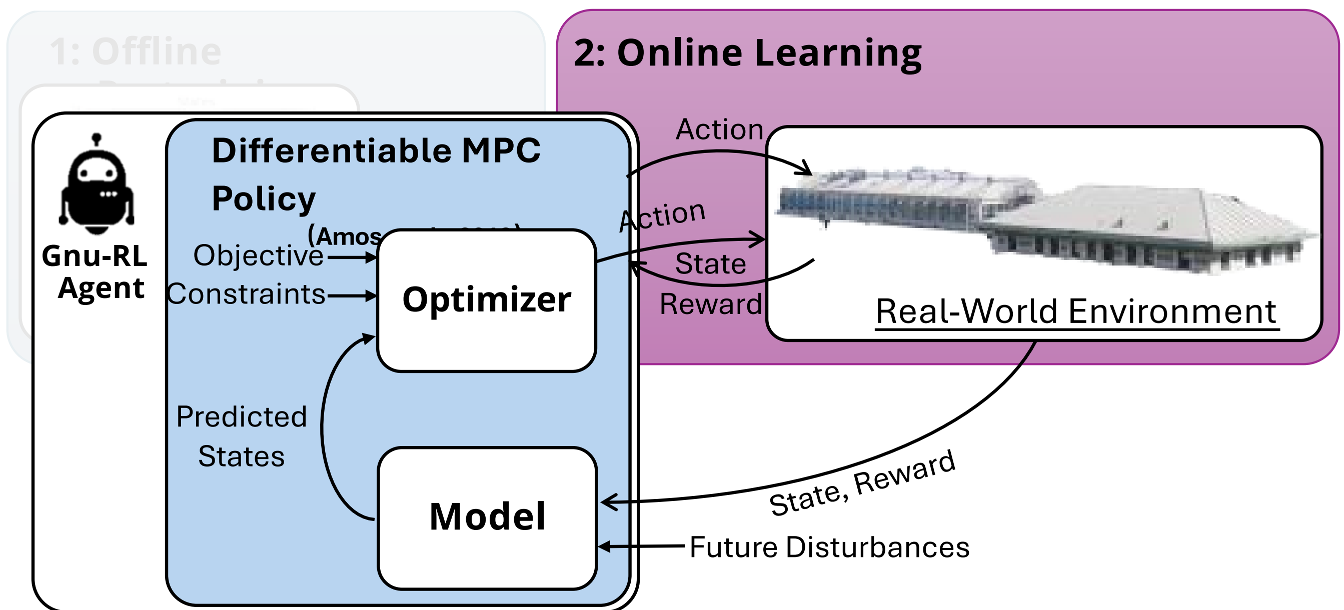


Framework



Step1: Offline Pretraining
BAS
Historical Data
Step2: Online Learning
Control Program
Environment
Action
Observation
Reward
$\tau^*_{1:T} = \arg\min_{\tau_{1:T}} \sum_t C_{\theta,t}(\tau_t)$
$x_{t+1} = f_\theta(\tau_t)$
Policy
Agent

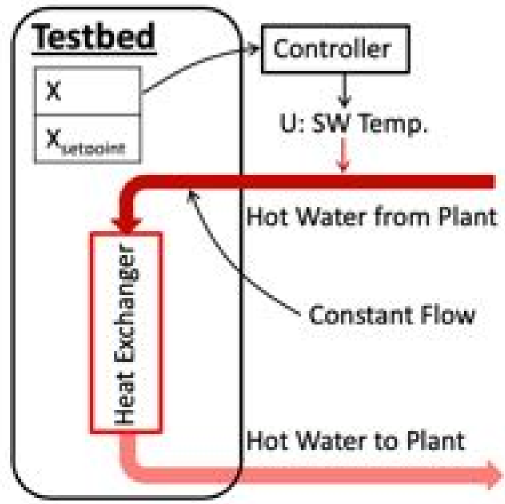
Besides imitating the existing controller, we expedite the training by using a policy that encodes knowledge on system dynamics and control.



Amos, B., Jimenez, I., Sacks, J., Boots, B., & Kolter, J. Z. (2018). Differentiable MPC for End-to-end Planning and Control. In Advances in Neural Information Processing Systems (pp. 8289-8300).

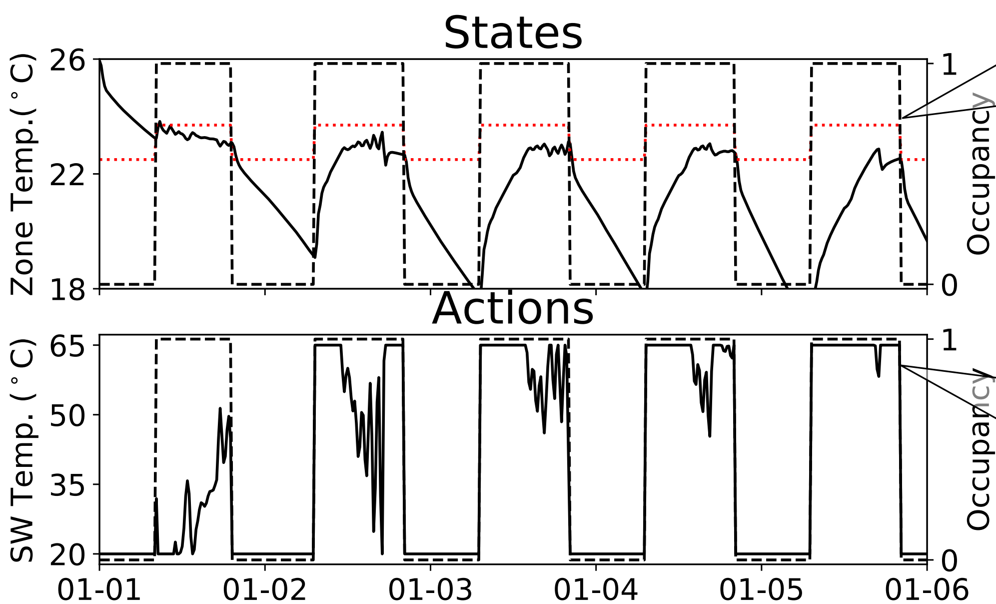
Experiment 1: Simulation Study

Hot Water Pipes



HVAC Schematic

Offline Pretraining: Gnu-RL imitates a P-controller.



The building operates under a fixed schedule

Setpoint

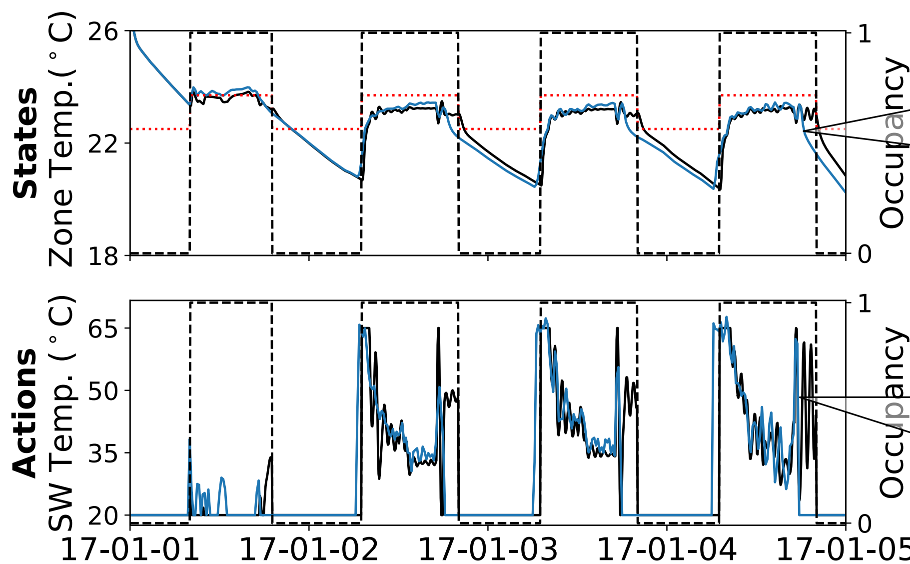
Occupancy

We used a

Baseline P-Controller

to generate historical data under the TMY3 weather sequence.

Online Learning: Gnu-RL is Precocial.



We evaluated the performance of Gnu-RL under the actual weather in 2017.

Gnu-RL performed similarly to Baseline P-Controller prior to any interaction with the environment.

Gnu-RL achieved significant energy savings without compromising thermal comfort.

- **Gnu-RL** achieved **20.6%** energy savings compared to the existing controller and **6.6%** energy savings compared to the best published RL result in the same environment.

	Total Heating Demand	Predicted Percentage Dissatisfied	
		Mean	**STD**
	(kWh)	**(%)**	**(%)**
Existing Controller	43709	**9.45**	5.59
Agent #6 (✤ Zhang & Lam, 2018)	37131	11.71	3.76
Gnu-RL	**34678**	9.56	6.39

Experiment: Real World

Carnegie Mellon University

Icons: flaticon.com

Environment: Purdue House
System: Heat pump with resistive backup heat

State (*x*): Indoor air temperature
Control Input (*u*): Power (translated to setpoint)
Disturbances (*d*): Solar gain, Outdoor air temperature

Objective: Minimize total and peak energy consumption and temperature deviation.

Horizon: 24 hours
Control Interval: 60 minutes

Training Data: 30 days
Validation Data: 15 days
Evaluation Data: 30 days

Baselines:
PID: Existing controller
MPC: (Pergantis2024)

Figure 3: Testbed House is a 208 m^2, 1920s-era house with all-electric appliances in West Lafayette, Indiana, USA.

Longest residential RL deployment without pretraining with a simulator

Limitations of Gnu-RL & D-MPC

Cognition & Action

Carnegie Mellon University

Icons: flaticon.com

$$T_{t+1} = \frac{dt}{C}\left(\frac{T_m - T_t}{R_m} + \frac{T_{oat} - T_t}{R_o} + COP * P_{hp} + \eta P_{bh} + \alpha Q_{sol}\right)$$

Dynamics in Gnu-RL:
$T_{t+1} = AT_t + B_u u_{SW} + B_d T_{oat}$

2R1C requires $u = P\ or\ Q$
Testbed only accepts $T_{setpoint}$

Choice of Controllable Action

Gnu-RL: Fitting $\{A, B_u, B_d\}$ to minimize a L when $\{O, R\}$ are fixed.
MPC: Non-quadratic cost function

Manual Configuration of Cost Function

Differentiable MPC: Expert demonstrations will fit the cost
Using existing data to fit $\{O, R\}$ results in suboptimal performance

Non-Expert Demonstrations

Can we do better?

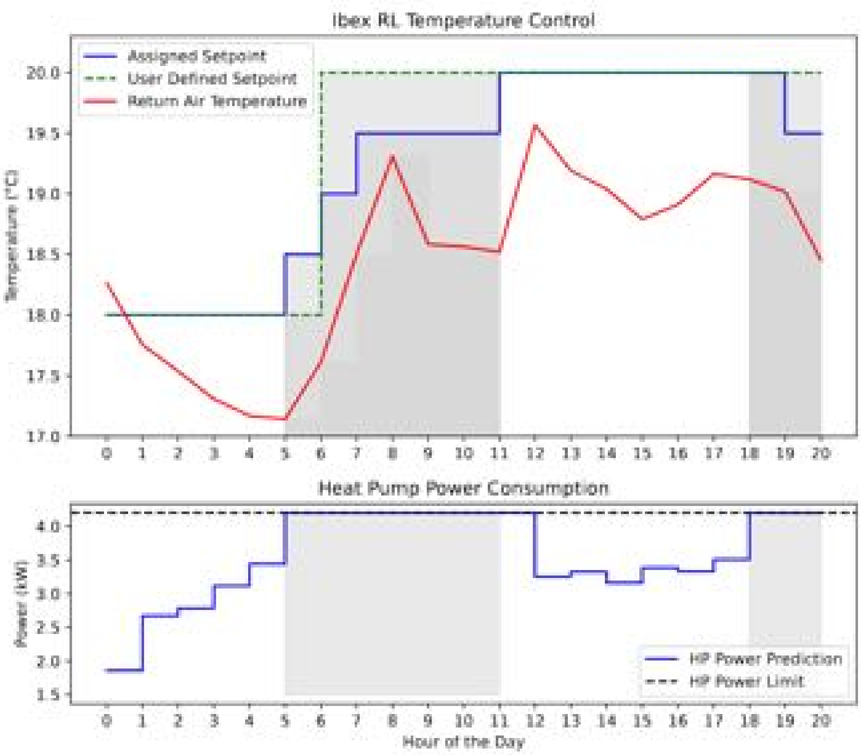
Ibex-RL Agent

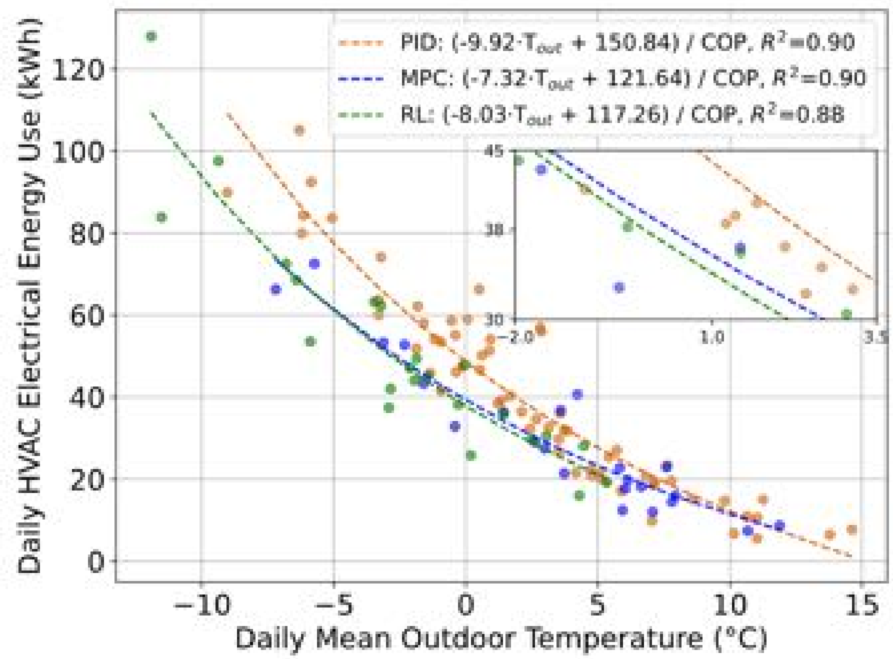
Mulayim, Bergés
ACM BuildSys (2025)

Results-Real World

Carnegie Mellon University

Icons: flaticon.com





Ibex-RL learned to do **stepped increases** to avoid using backup heat!

Ibex-RL achieved **22% improvement** in savings while MPC got 20%.

Mulayim, Bergés
ACM BuildSys (2025)

Can we change the design/control objective?

Making it closer to the types of *capital* we ultimately want to realize

How do we coordinate a large population of flexible building loads to address challenges arising from modern grid operation?

How do we ensure end-use requirements are satisfied for individual buildings given their system-specific dynamics?

COHORT: Coordination of Heterogeneous Thermostatically Controlled Loads for Demand Flexibility

Bingqing Chen
Carnegie Mellon University
Pittsburgh, PA, USA
bingqinc@andrew.cmu.edu

Jonathan Francis
Carnegie Mellon University
Pittsburgh, PA, USA
jmf1@cs.cmu.edu

Marco Pritoni
Lawrence Berkeley National Lab
Berkeley, CA, USA
mpritoni@lbl.gov

Soummya Kar
Carnegie Mellon University
Pittsburgh, PA, USA
soummyak@andrew.cmu.edu

Mario Bergés
Carnegie Mellon University
Pittsburgh, PA, USA
mberges@andrew.cmu.edu

$$g_{\text{tv}}\Big(\sum_i \mathbf{u_i}\Big) = \sum_{k=t+1}^{t+T} |P_{\text{net,k}} - P_{\text{net,k-1}}|$$

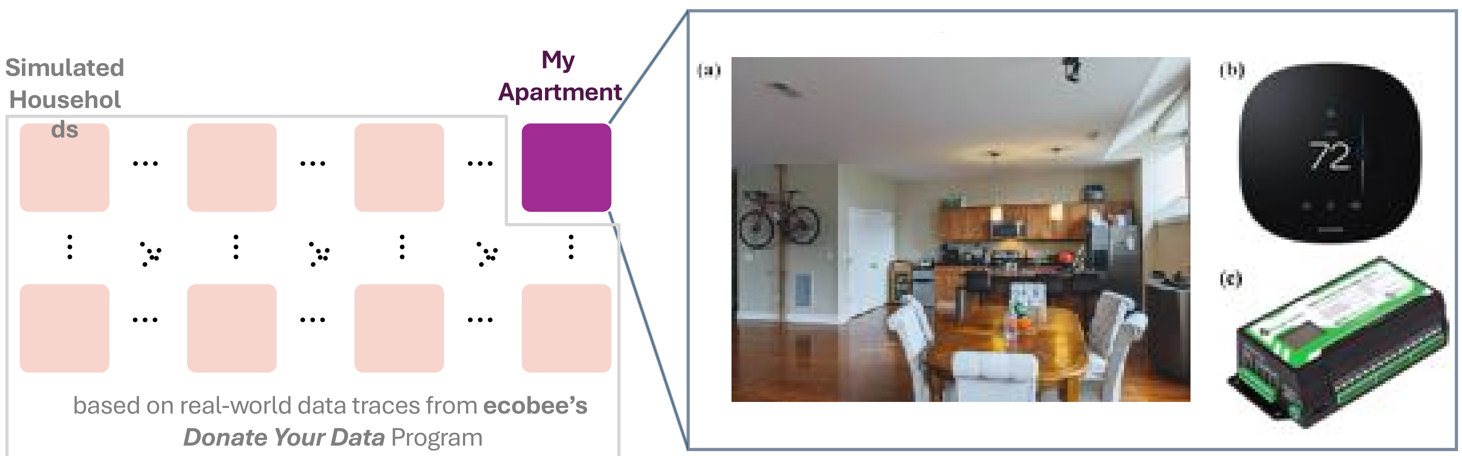
where,
$$\mathbf{P}_{\text{net}} = \mathbf{P}_{\text{total}} - \mathbf{P}_{\text{gen}}$$
$$\mathbf{P}_{\text{total}} = \mathbf{P}_{\text{non-shiftable}} + \sum_i \mathbf{u_i}$$

COHORT is a practical, scalable, and versatile solution for coordinating a large population of flexible building loads to jointly provide grid services, while ensuring the end-use requirements are satisfied at individual buildings.

[Code] https://github.com/INFERLab/COHORT
[Paper] https://doi.org/10.1145/3408308.3427980

Chen et al.
ACM BuildSys (2020)

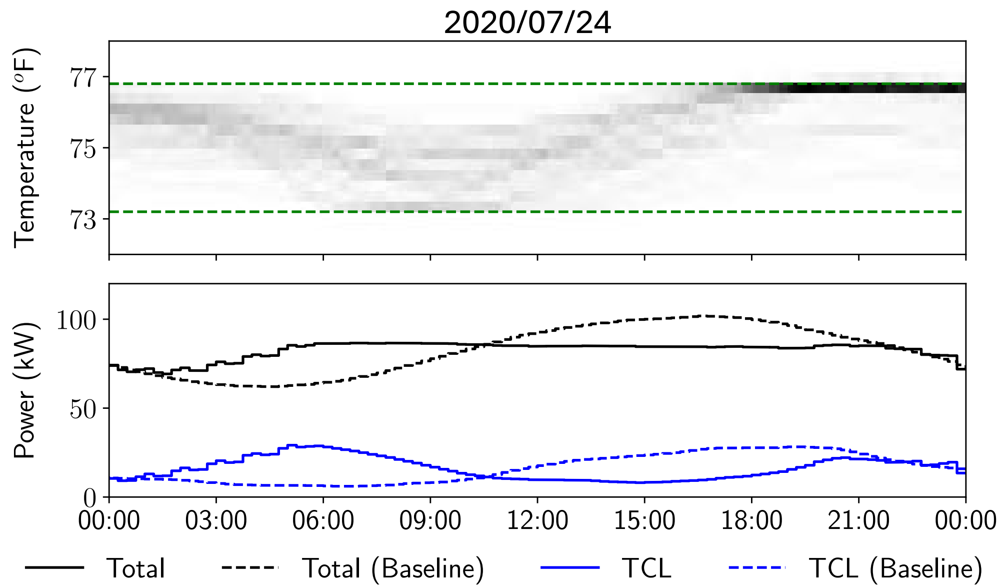
We validated that COHORT is practical for real-world systems through a hardware-in-the-loop simulation.



15-day Experiment Period: 2020/07/11-2020/07/25

Use Case 3: Peak Load Curtailment (Population)

COHORT reduced daily peak loads by an average of **12.5%**.



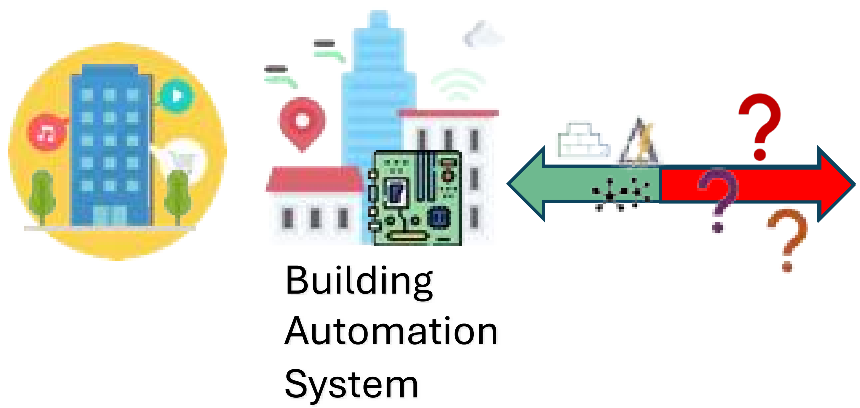
(From PJM Interconnection)

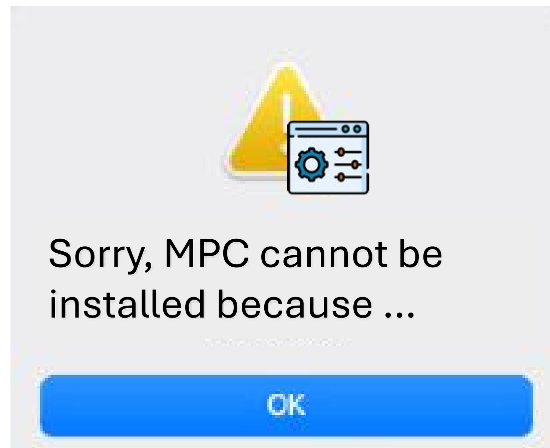
Looks great!
Can we do this for all buildings, and for other objectives?

Well, not so fast....

“*Buildings*” is a heterogenous target

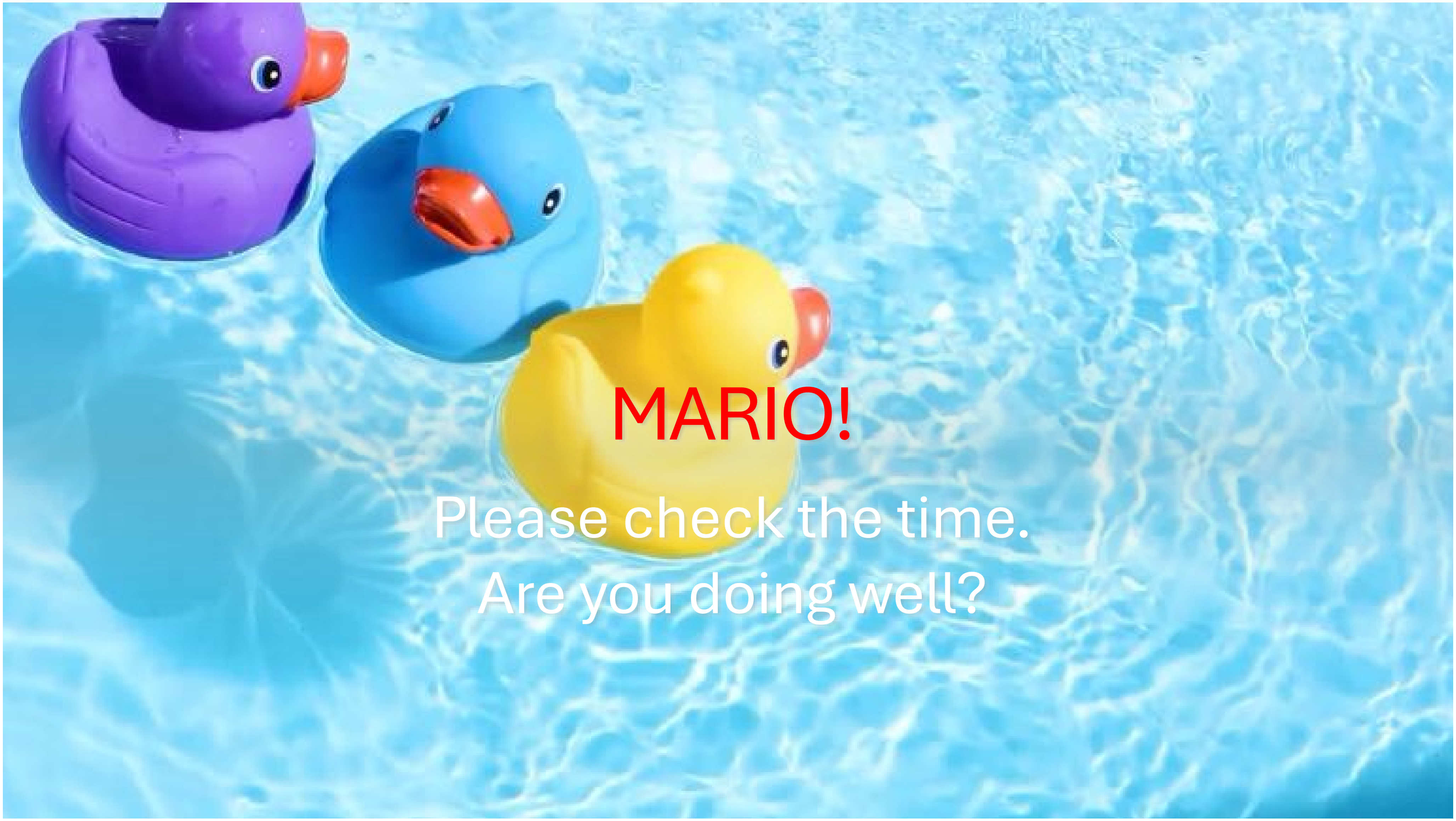
To scale these solutions we need so much more...





Operational Heterogeneity: There are multiple ways to sense, model and actuate for each building.

What are the inputs available for your model?
What modeling paradigm should be used?
What actuations are available?


MARIO!
Please check the time.
Are you doing well?

And there are many other unknowns

To design, we need to understand and predict well

Human behavior makes predictions harder

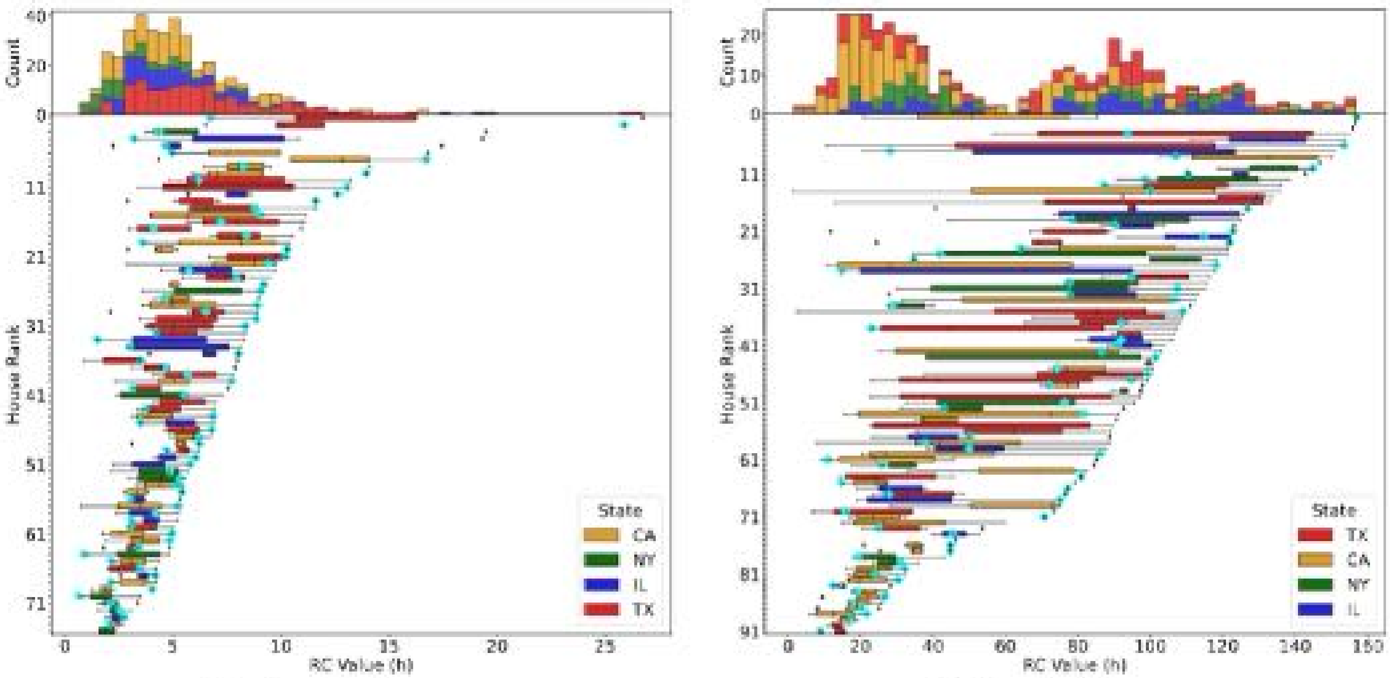
- Reinforcement Learning controllers trained on deterministic occupancy patterns break down when tested on stochastic occupancy.

Mulayim, Bergés
ACM e-Energy (2025)

- The parameters of thermal dynamics models of buildings change drastically over time and vary significantly across rooms in homes.

Mulayim, Bergés
ACM BuildSys (2023)

Gray-box models can help us understand the causes

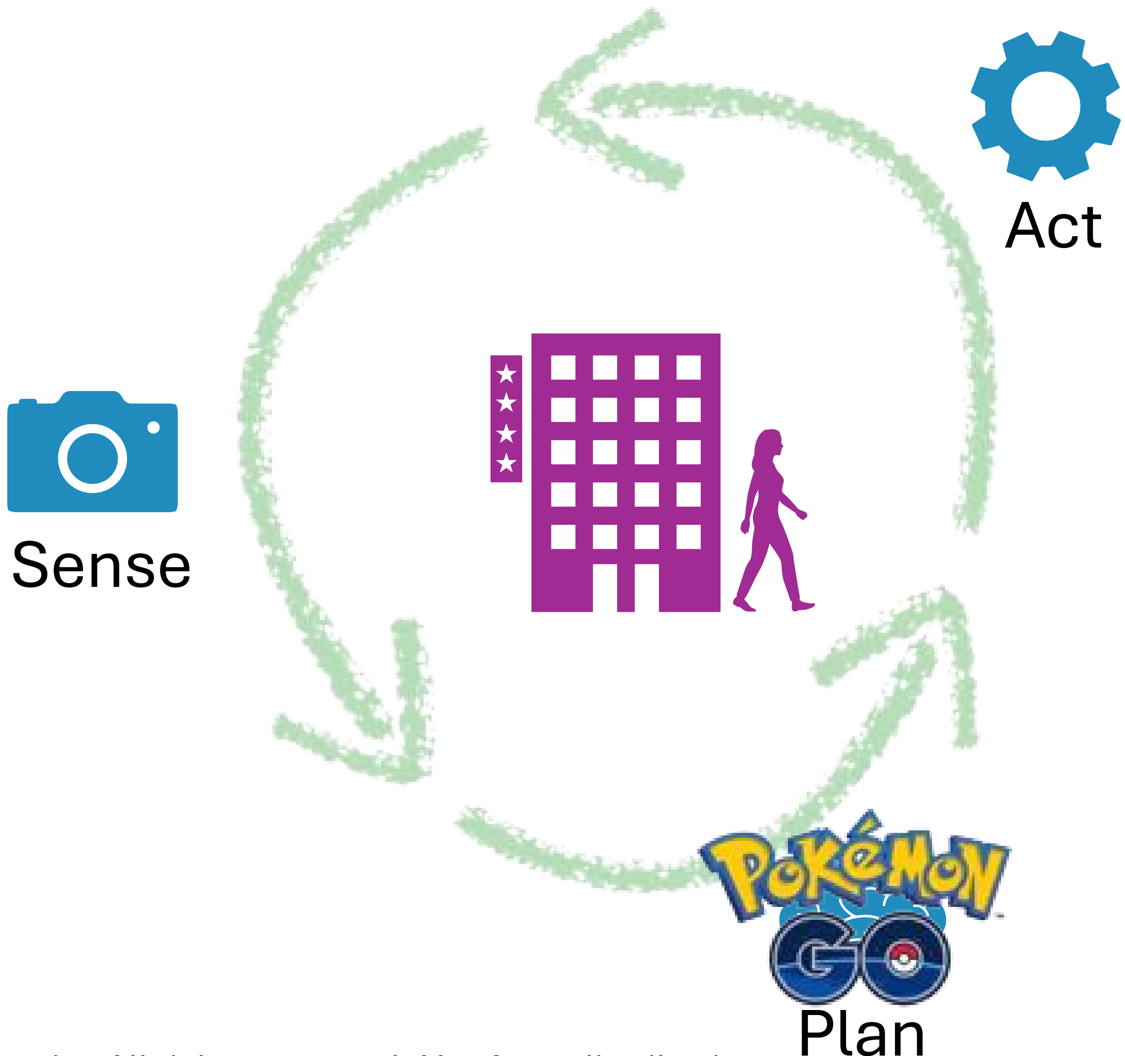


(a) Cooling Season

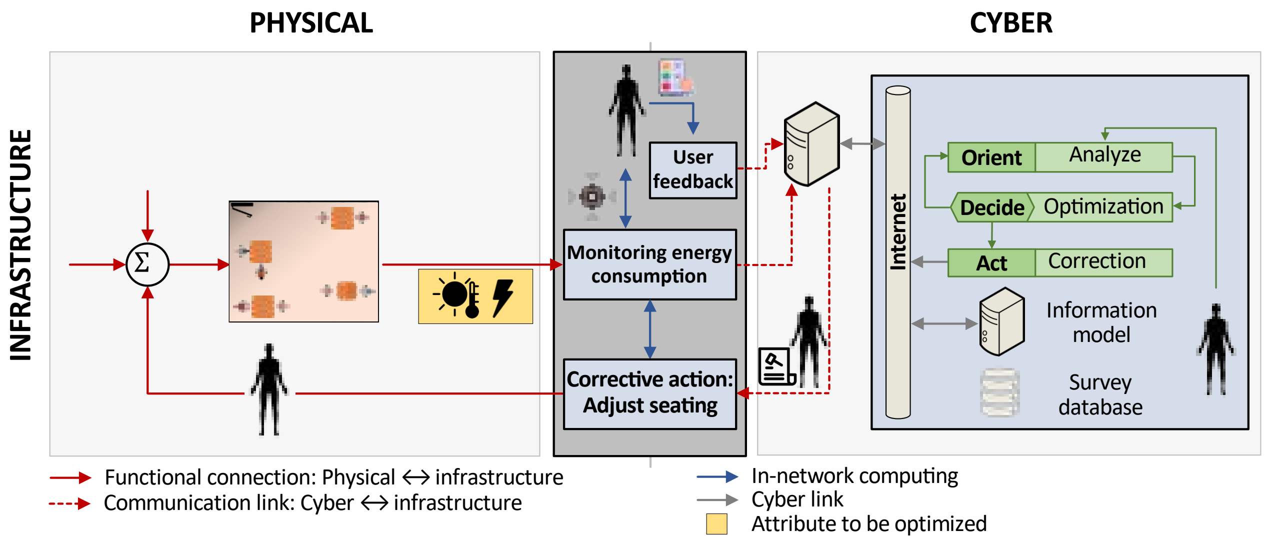
(b) Heating Season

Figure 8. *This histogram depicts the collective distribution of RC values (top), accompanied by boxplots for individual room distributions (bottom) for cooling and heating seasons. Markers indicated in light blue represent the RC values for the room where the thermostat is located..*

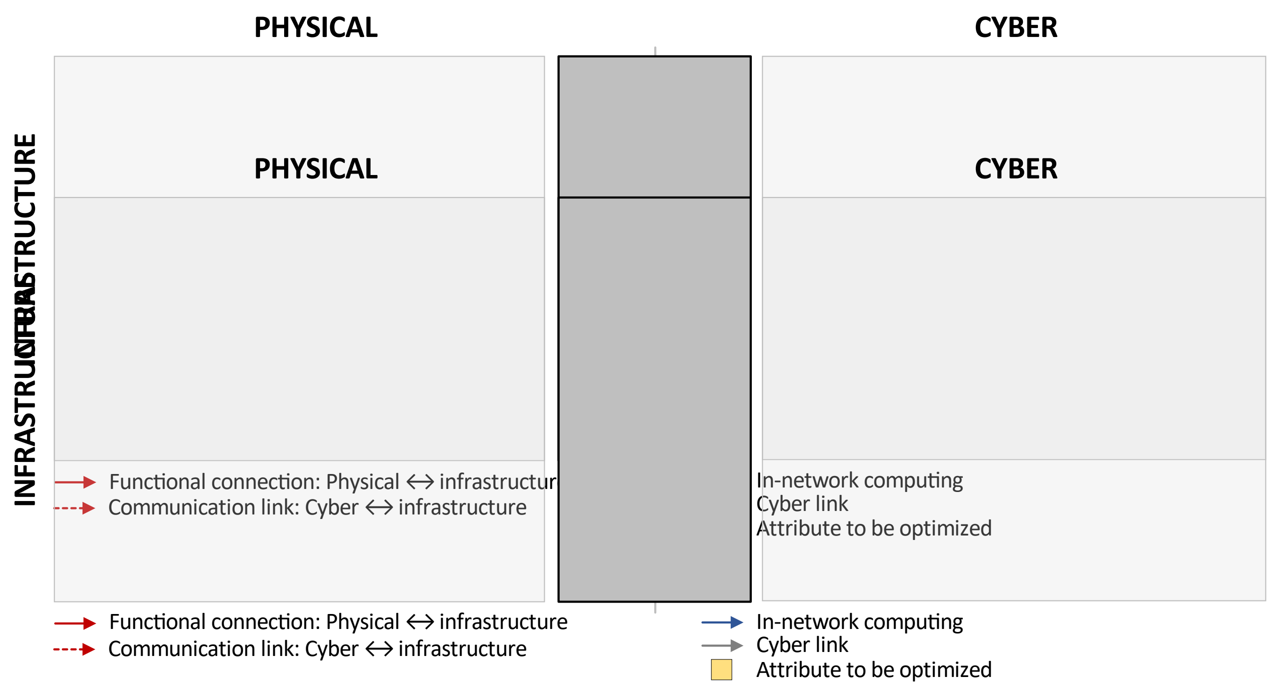
Mulayim, Bergés, Severnini
DCE Journal (2024)


Act
Sense
Plan

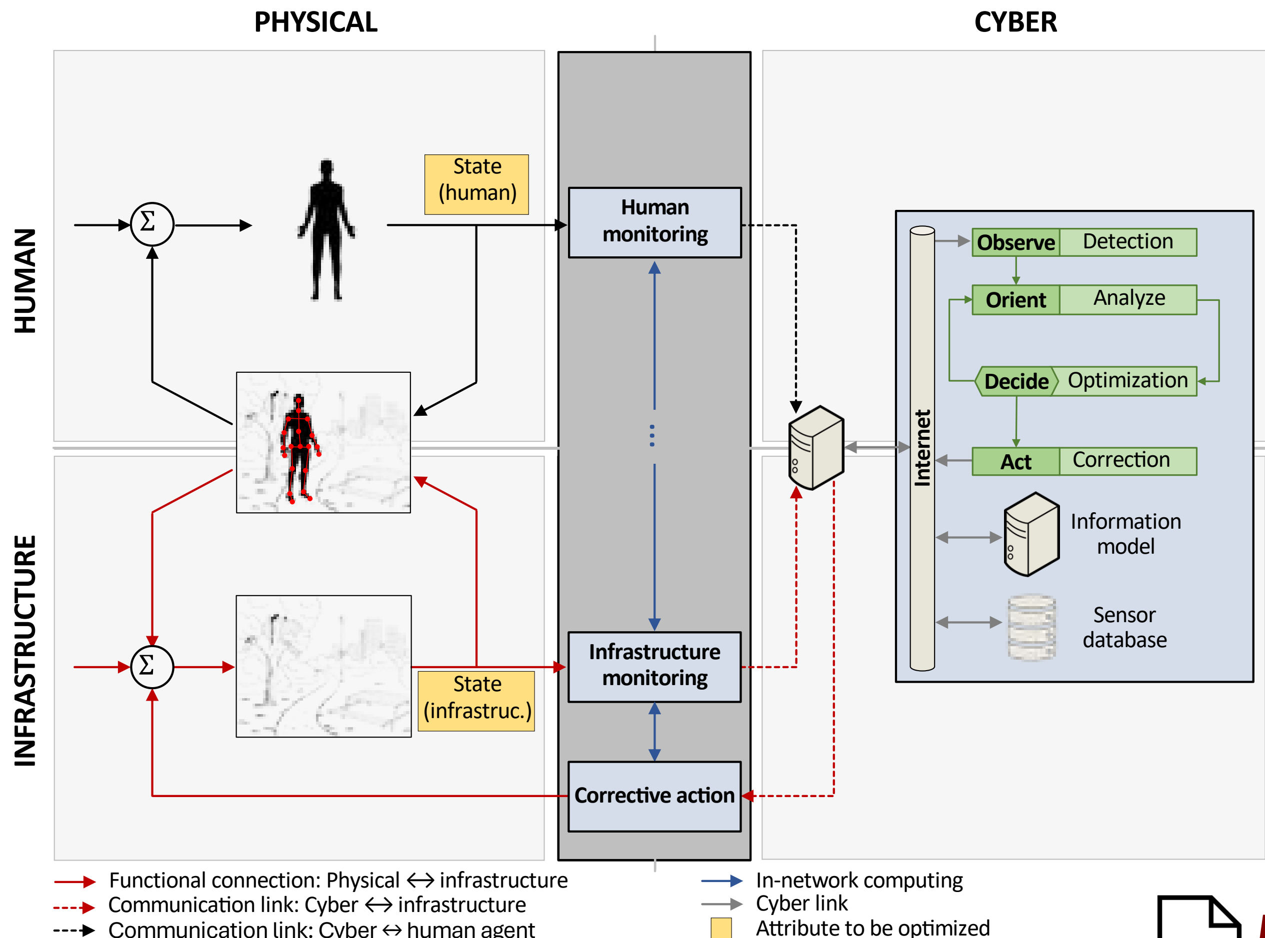
Integrating Humans into CPS



Integrating Humans into CPS



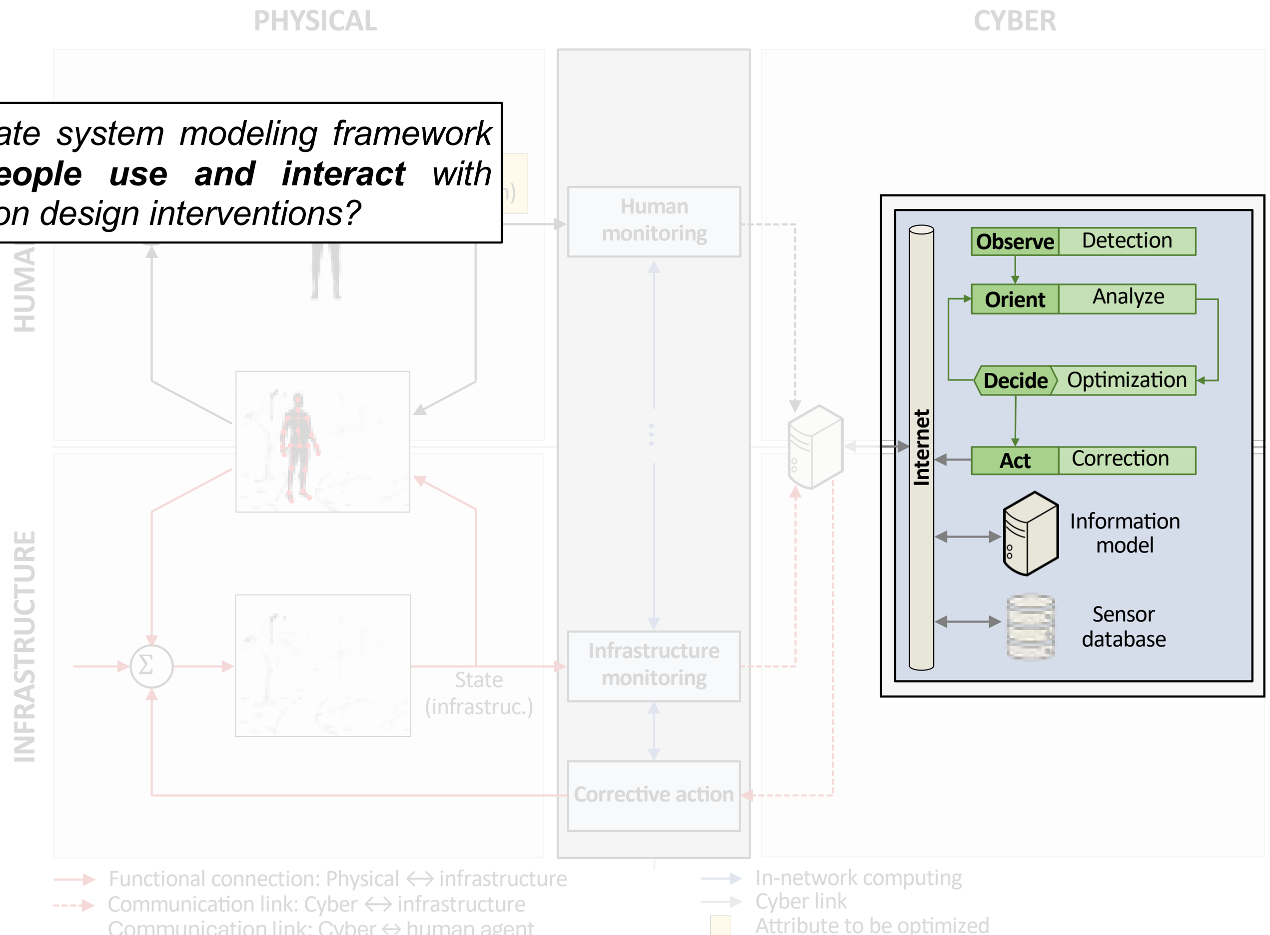
CPSIS Framework



Doctorarastoo et al.
BuildSys (2023)

CPSIS Challenges

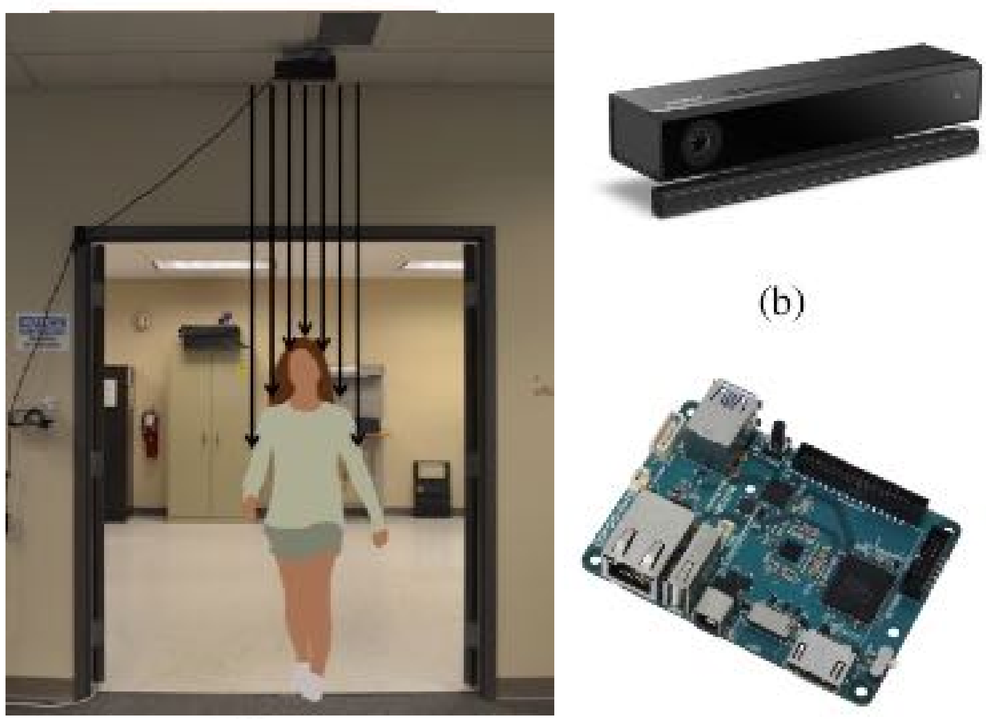
What is an appropriate system modeling framework to ***predict how people use and interact*** *with infrastructure based on design interventions?*



Challenges

- Defining Objectives
- Measuring Objectives
- User Privacy
- **Modeling Humans**
- Computational Complexity
- Real-time Actuation
- Actuation Downtime

My humble beginnings…



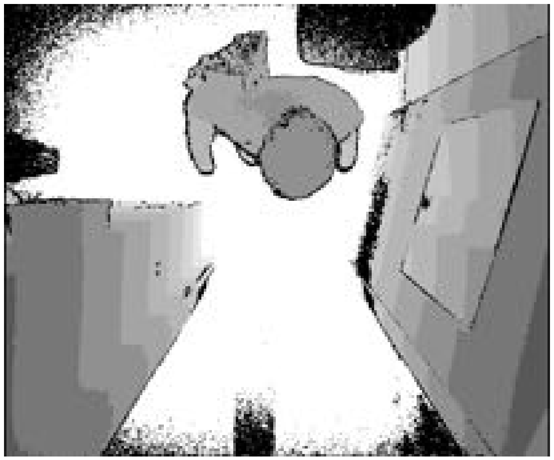
(a) (c)

(a) Placement of a Kinect sensor on ceiling tile.

(b) Kinect sensor

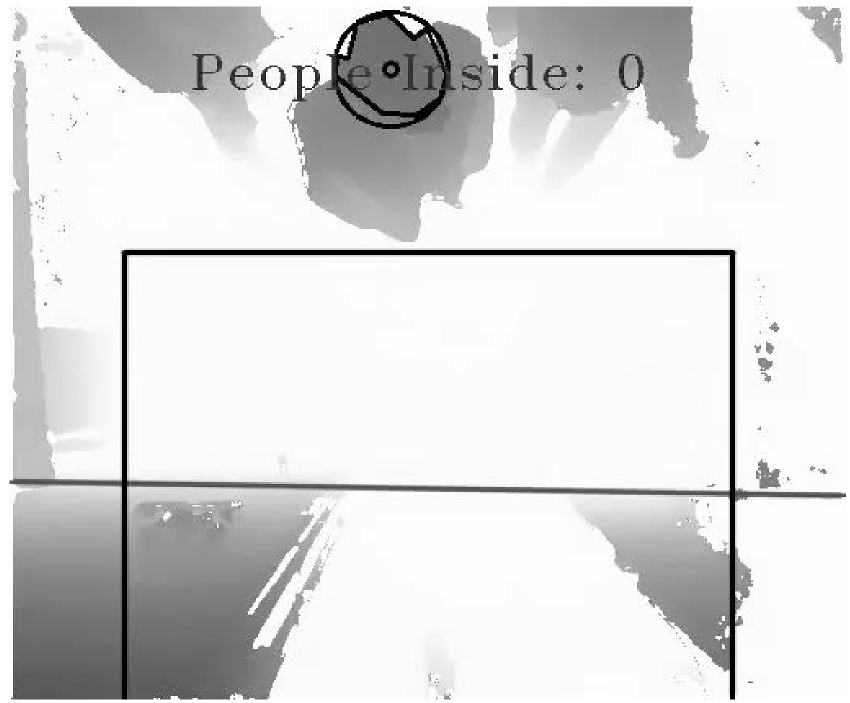
(c) Embedded computer: Odroid-XU4

Some sample data



Depth Map

RGB Data



Munir et al.
IEEE ICCCPS (2017)

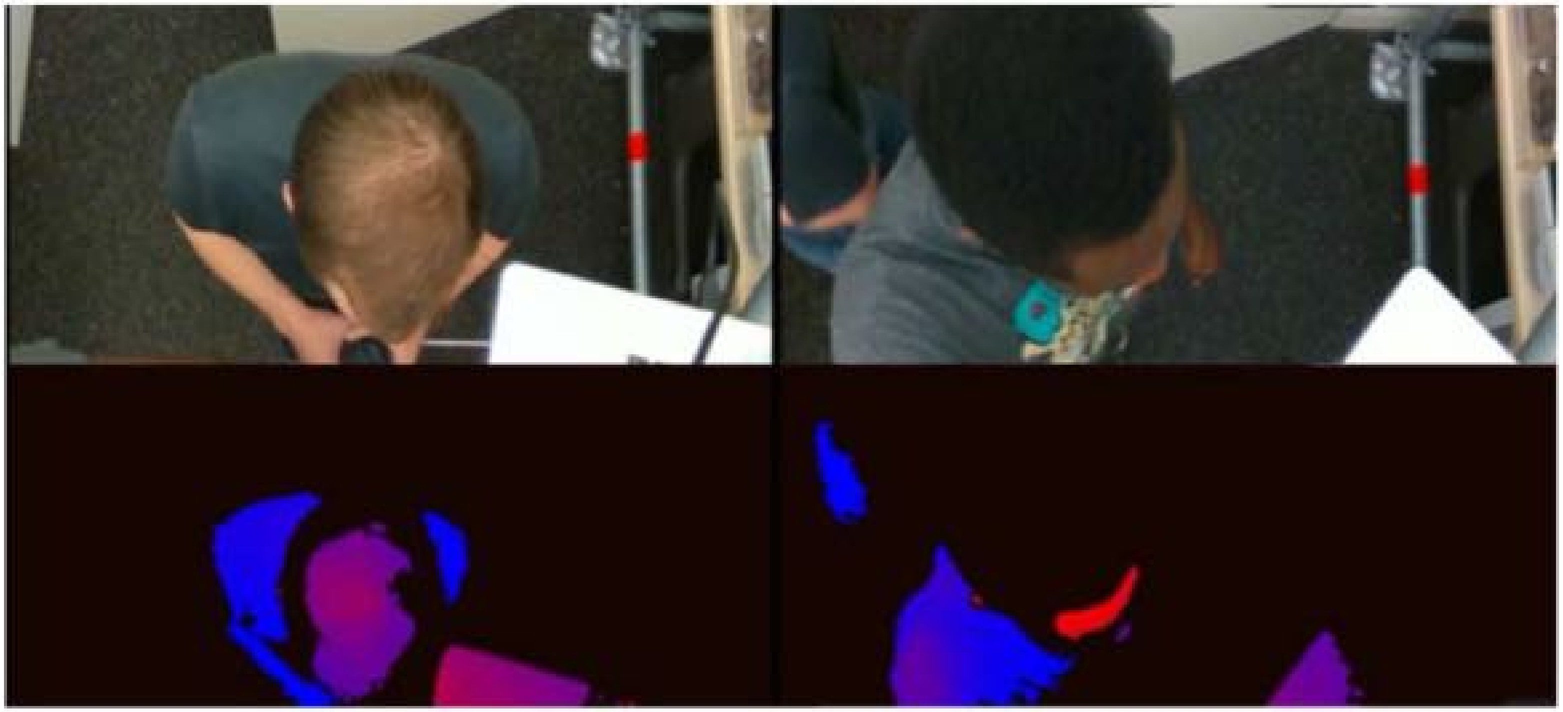


Figure 7. Color (top) and depth (bottom) image of two human subjects with different hair types as measured by an active infrared stereo camera. The depth map is color-coded such that darker (more black) is farther away from the camera while lighter (more red) is closer.

Becerik-Gerber et al.
Nature Scientific Reports (2022)

Data-driven agent based models

How can we model human behavior in a generalizable way?

AGENT-BASED MODELING: PROMISE AND PITFALLS

Bottom-up power
Simulates individuals and groups, allowing complex patterns to emerge

Broad adoption
Used in urban planning, evacuation modeling, crowd dynamics

Appealing vision
Could capture heterogeneity, unlike top-down models

Need:
Data-driven behavior modeling

Oversimplified agents
Rely on static, rule-based behaviors ("if X then Y")

Need:
Transition from survey and census data to in-situ, interaction-level data

Lack of empirical grounding
Rules often theoretical or arbitrary, leading to unrealistic dynamics

Need:
Embed contextual, person-to-person, temporal, and scenario-based generalizability

Poor generalizability
Models tuned to specific scenarios but fail in new layouts, contexts, or populations

Need:
Hierarchical decision making governing strategic and execution planning

Gap in representation
Does not reflect humans as dynamic, adaptive, and context-sensitive

EMPIRE

Empirical Modeling of People in Responsive Environments

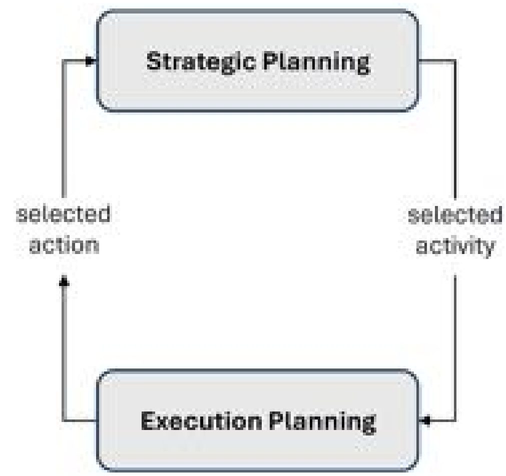
A hierarchical, data-driven modeling framework for predicting group-level human spatio-temporal behavior in dynamic physical environments, with a focus on scenario-based generalizability.

Flanigan et al.
Under Review

EMPIRE

Empirical Modeling of People in Responsive Environments

A **hierarchical**, data-driven modeling framework for predicting group-level human spatio-temporal behavior in dynamic physical environments, with a focus on scenario-based generalizability.

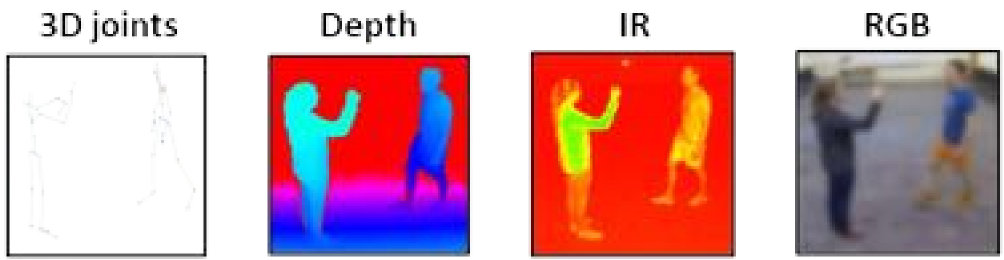


Flanigan et al.
Under Review

EMPIRE

Empirical Modeling of People in Responsive Environments

A hierarchical, **data-driven** modeling framework for predicting group-level human spatio-temporal behavior in dynamic physical environments, with a focus on scenario-based generalizability.



EMPIRE

Empirical Modeling of People in Responsive Environments

A hierarchical, data-driven modeling framework for predicting **group-level** human spatio-temporal behavior in dynamic physical environments, with a focus on scenario-based generalizability.

Physiological signals

Individual or group cognition

Population dynamics

EMPIRE

Empirical Modeling of People in Responsive Environments

A hierarchical, data-driven modeling framework for predicting group-level human spatio-temporal behavior in dynamic physical environments, with a focus on **scenario-based** generalizability.

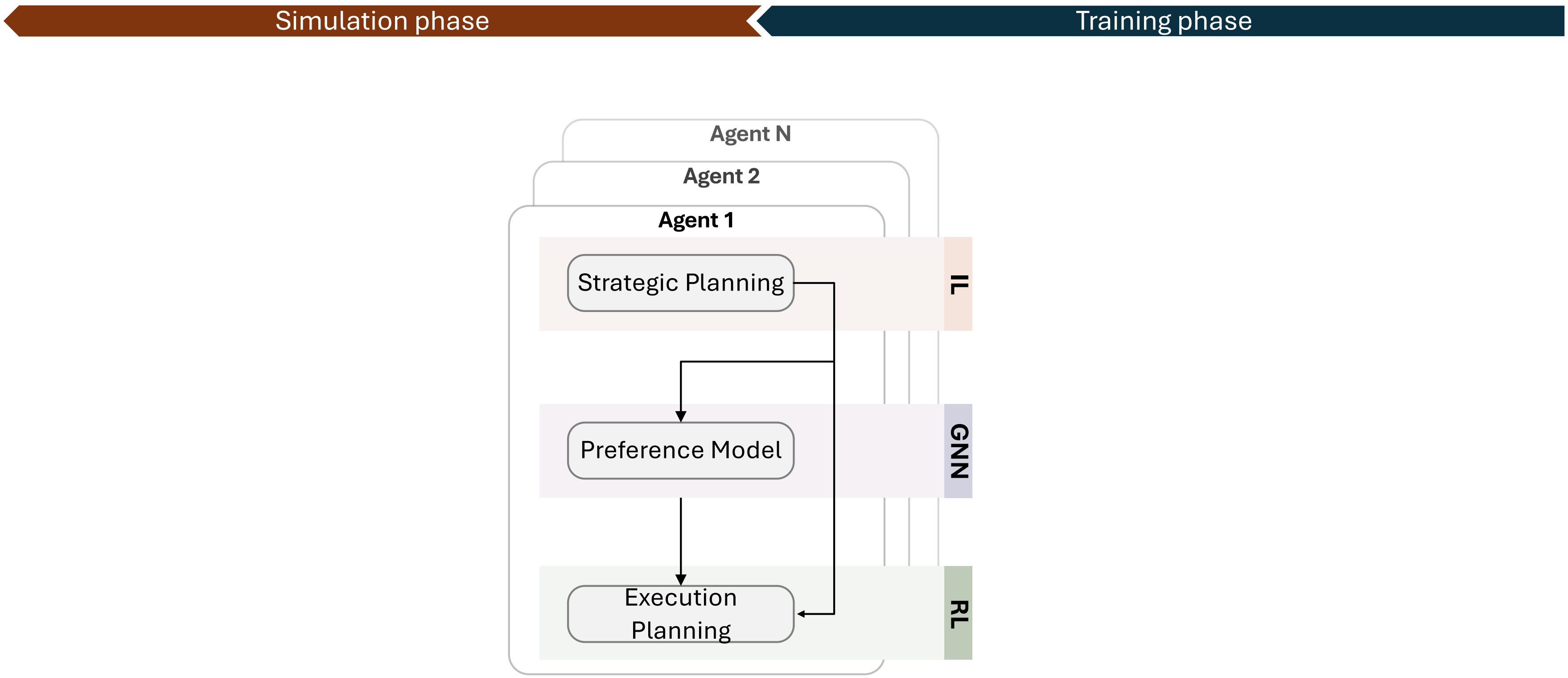
Contextual generalizability

Person-to-person generalizability

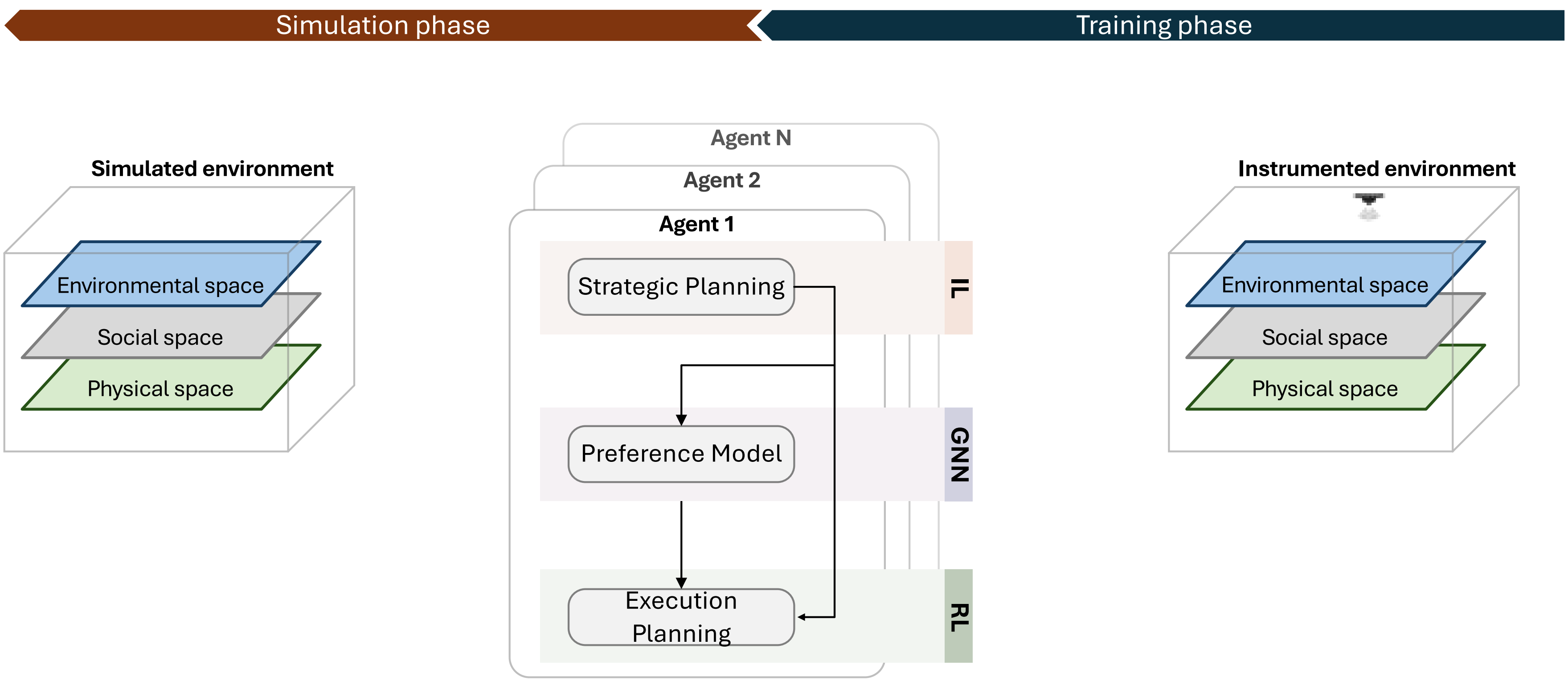
Temporal generalizability

Scenario-based generalizability

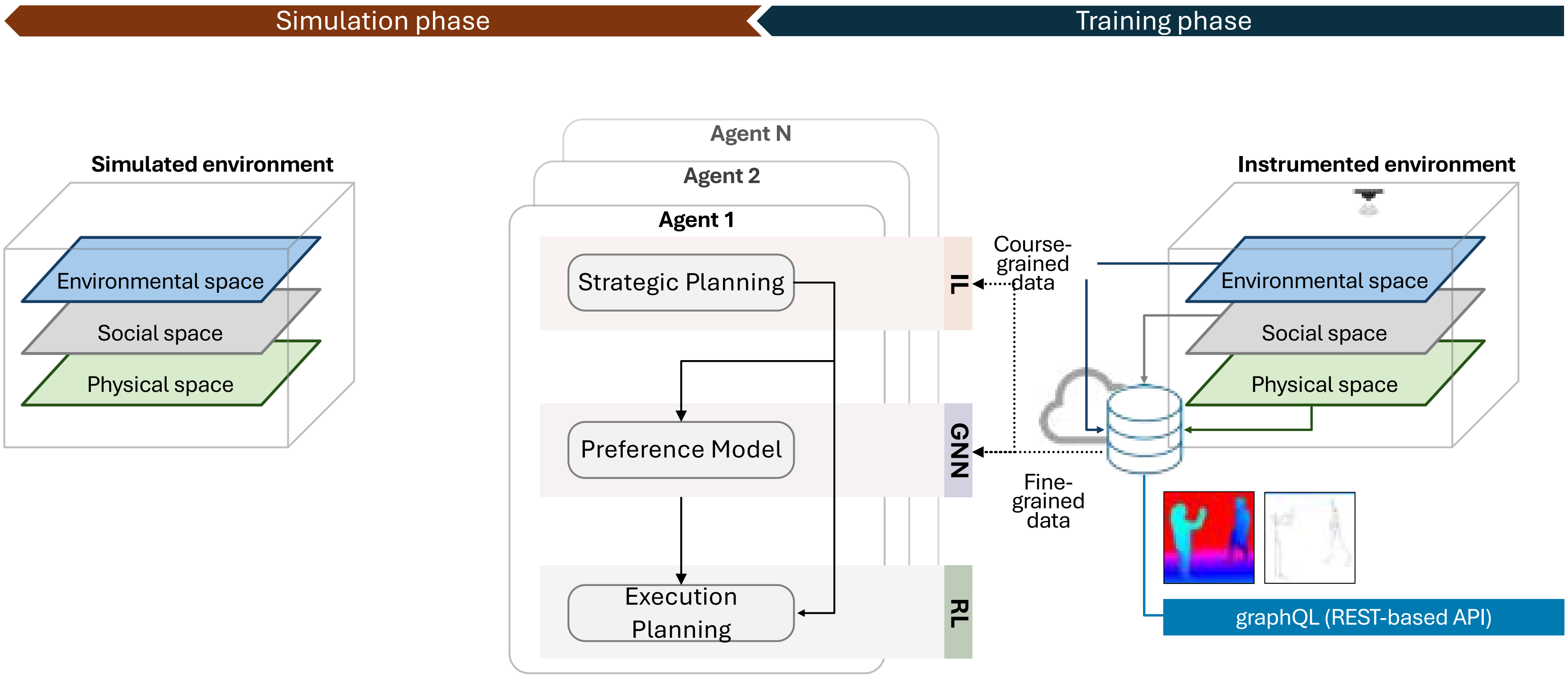
EMPIRE HIGH-LEVEL ARCHITECTURE



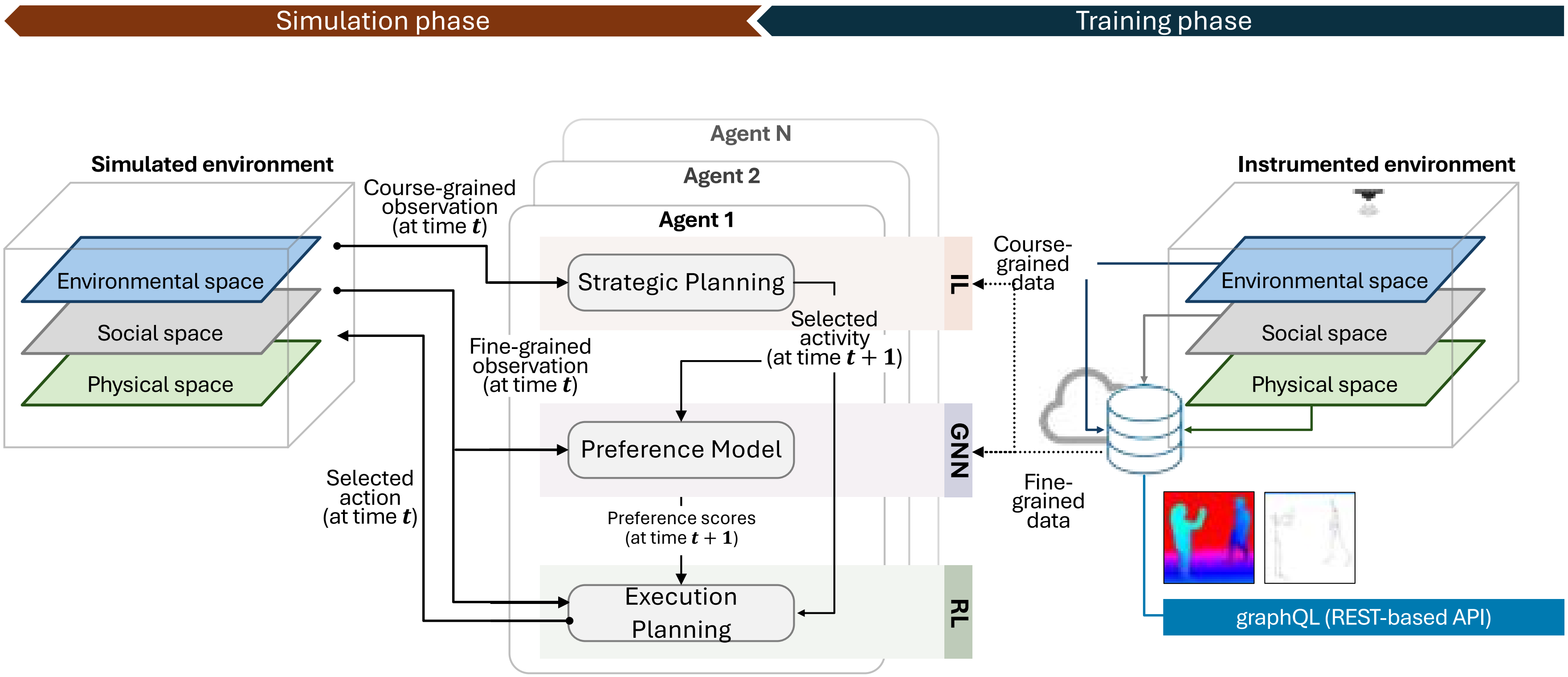
EMPIRE HIGH-LEVEL ARCHITECTURE



EMPIRE HIGH-LEVEL ARCHITECTURE

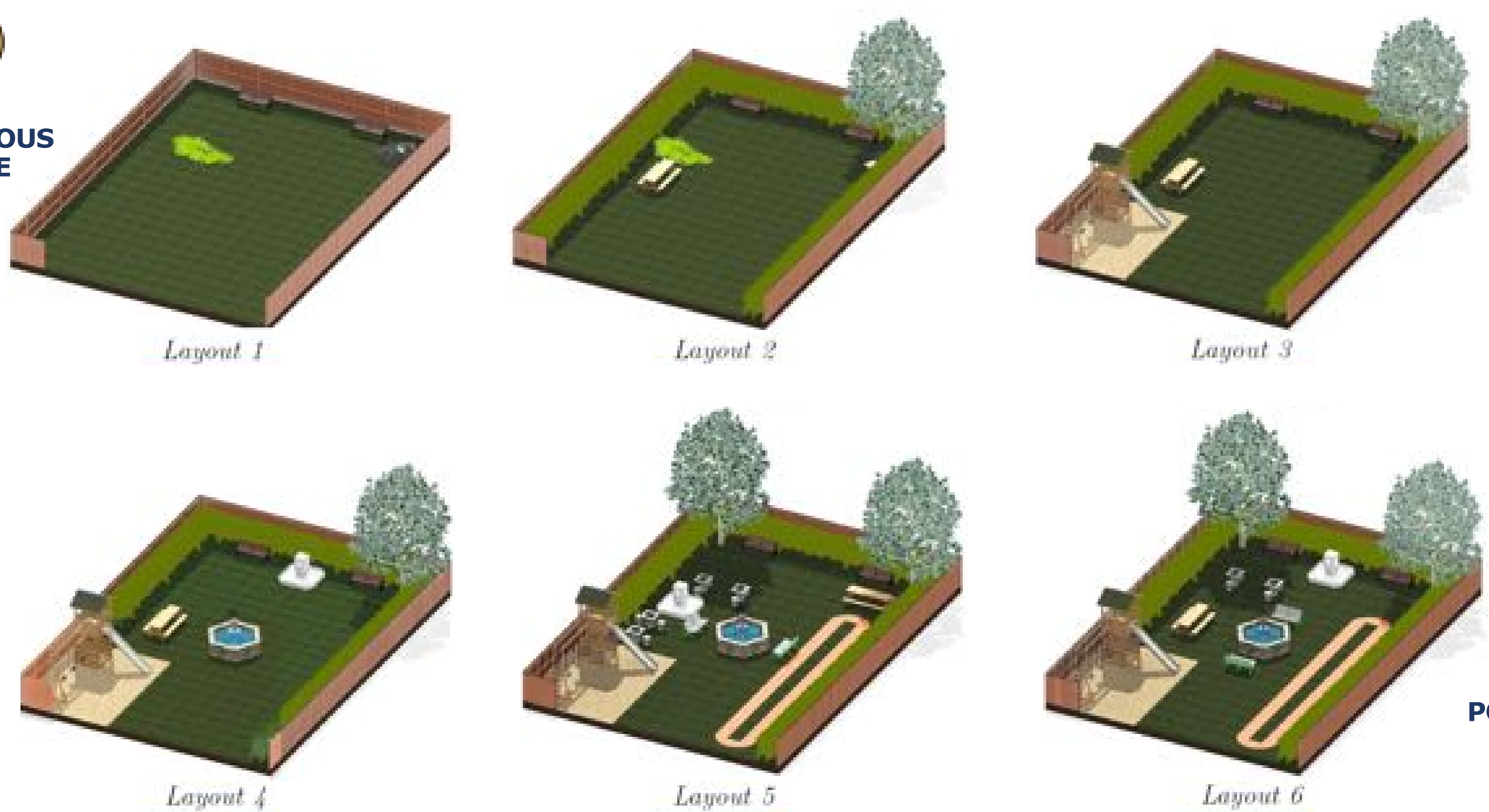


EMPIRE HIGH-LEVEL ARCHITECTURE



ILLUSTRATIVE CASE STUDY

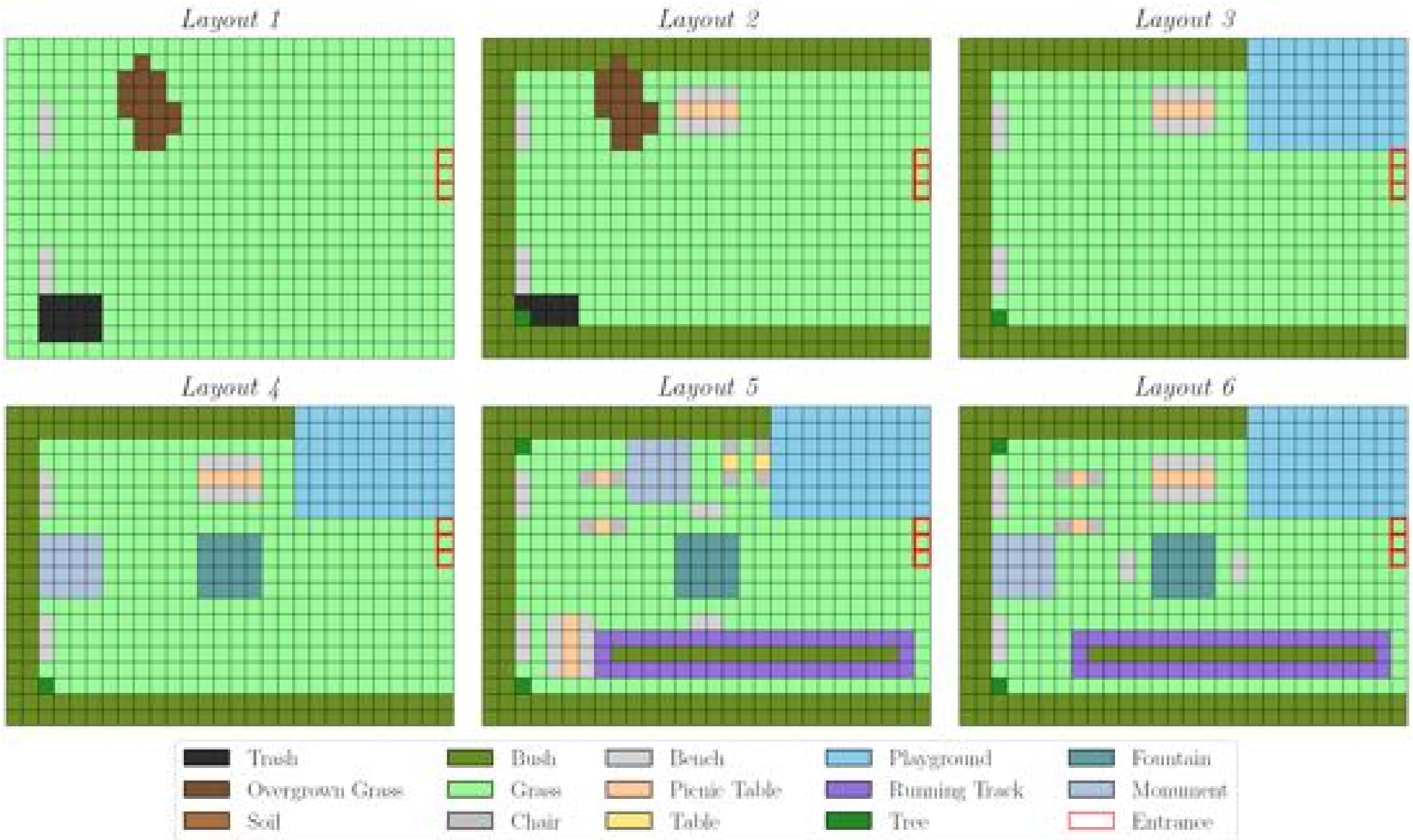
AMBIGUOUS SPACE



POSITIVE SPACE

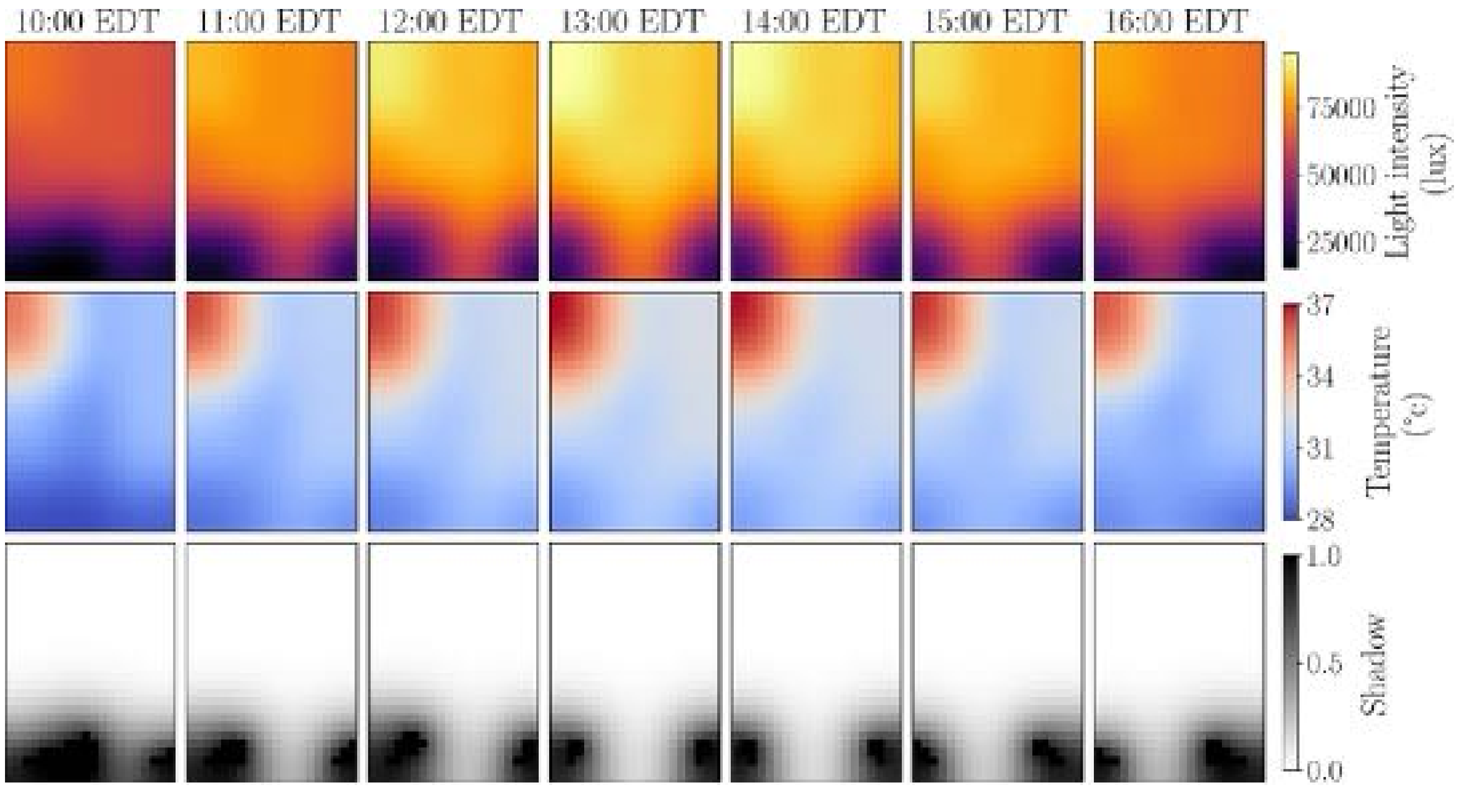
ILLUSTRATIVE CASE STUDY

Physical layer



ILLUSTRATIVE CASE STUDY

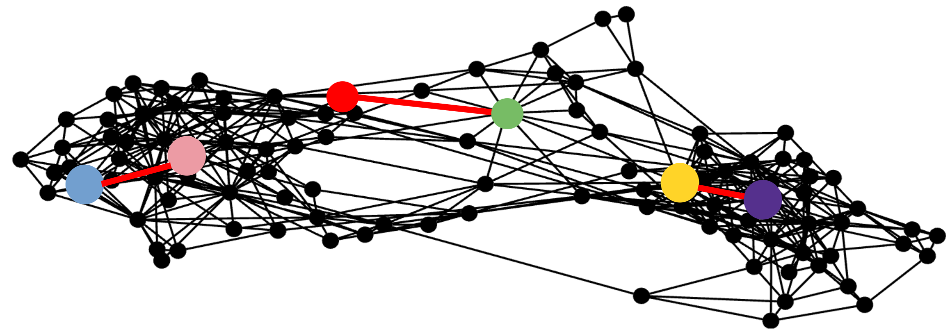
Preference layer



OK, what about even higher social objectives?

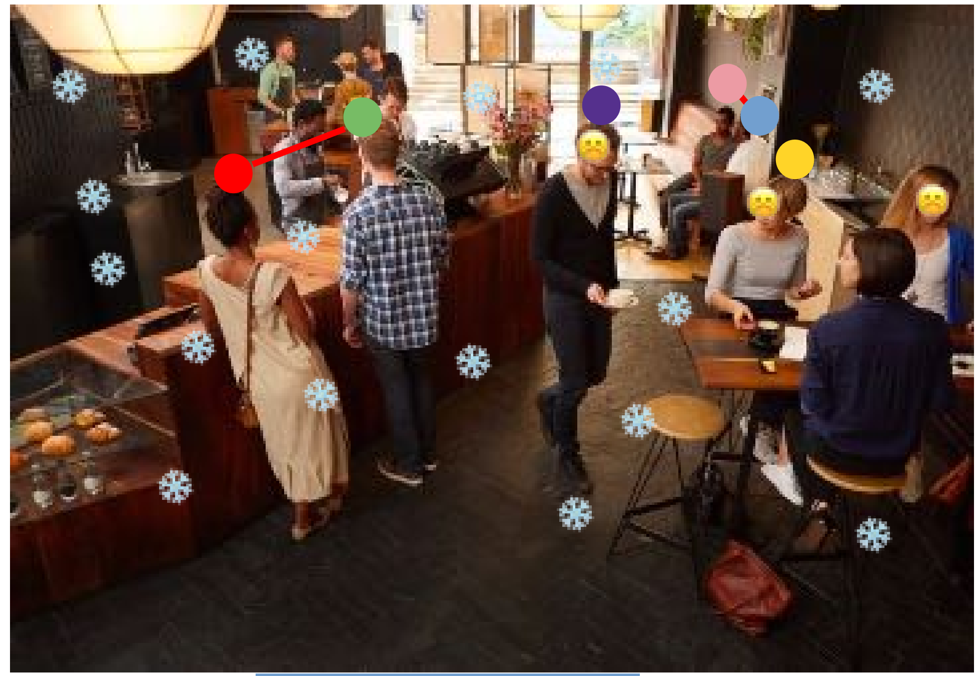
Can we infer social capital creation, for example?

To accomplish this, measuring social interactions in social infrastructure



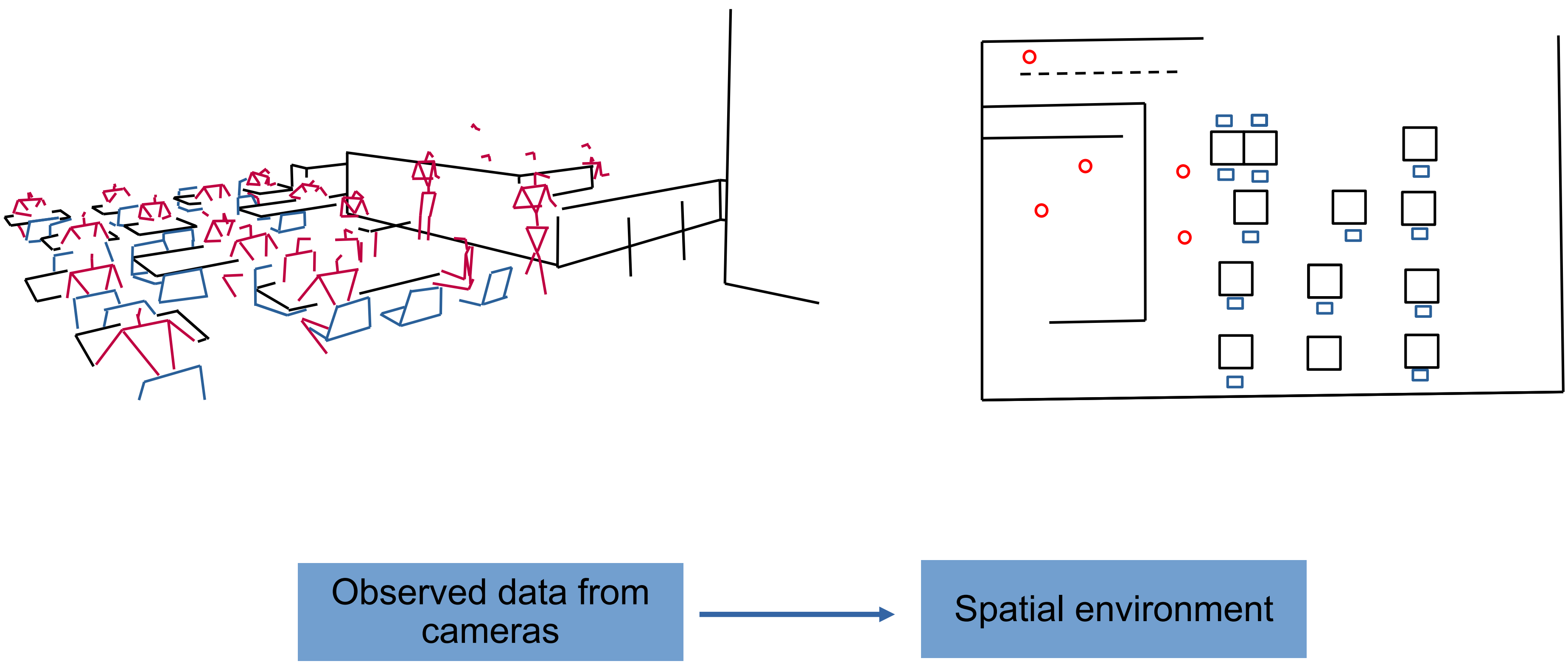
Social interactions → Social capital

Additional goal: measure influences of design changes on social interactions



Social interactions → Design change → Social interactions

Privacy invasive image → Privacy preserving aggregate


Observed data from cameras
Spatial environment

Conclusions

- There is a whole new world of design/control spaces to explore
- We are getting closer to explicitly designing for social objectives
- We still need better models and solvers to unlock it
- Data is becoming less of a problem, though privacy and ethics need to be considered
- Let's boldly go where no engineers/designers have gone before and directly optimize for the objectives we care about!

Thanks! Questions?

https://inferlab.org

marioberges@cmu.edu