A Trip to the Neural Frontier: Neurosymbolic Sensor Fusion for Trustworthy CPS

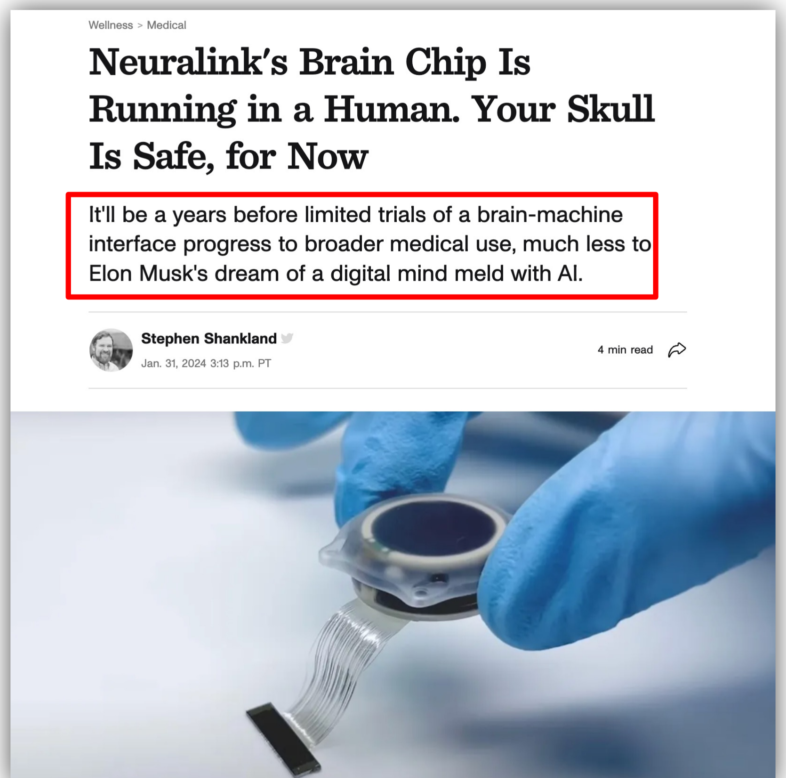
Luis Garcia
Assistant Professor
Kahlert School of Computing
University of Utah
https://lagarcia.us

A Trip to the Neural Frontier: Neurosymbolic Sensor Fusion for ~~Trustworthy~~ AI-DRIVEN CPS

EXPLAINABLE?
LESS BLACKBOXY?

Luis Garcia
Assistant Professor
Kahlert School of Computing
University of Utah
https://lagarcia.us

Welcome to the Neural Frontier...

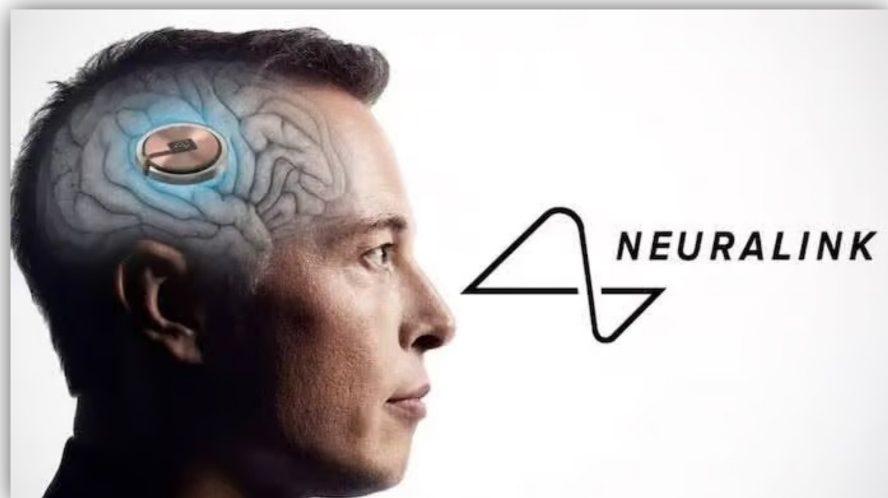


Wellness > Medical

Neuralink's Brain Chip Is Running in a Human. Your Skull Is Safe, for Now

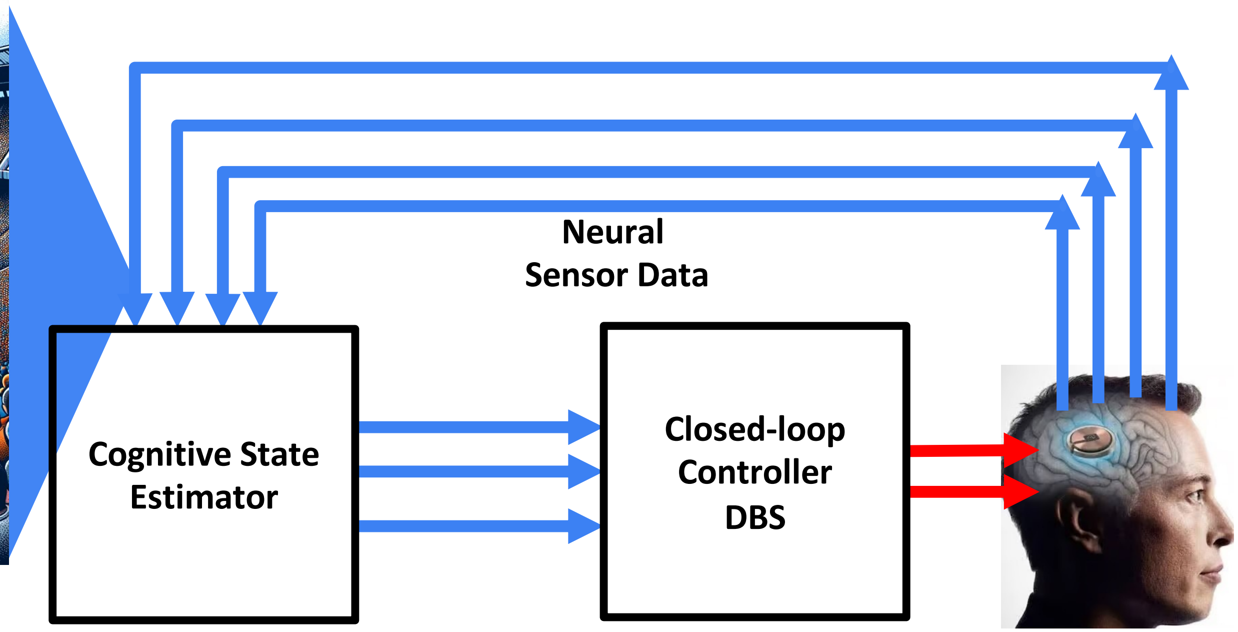
It'll be a years before limited trials of a brain-machine interface progress to broader medical use, much less to Elon Musk's dream of a digital mind meld with AI.

Stephen Shankland
Jan. 31, 2024 3:13 p.m. PT
4 min read



Is Elon's "digital mind meld with AI" so far away?

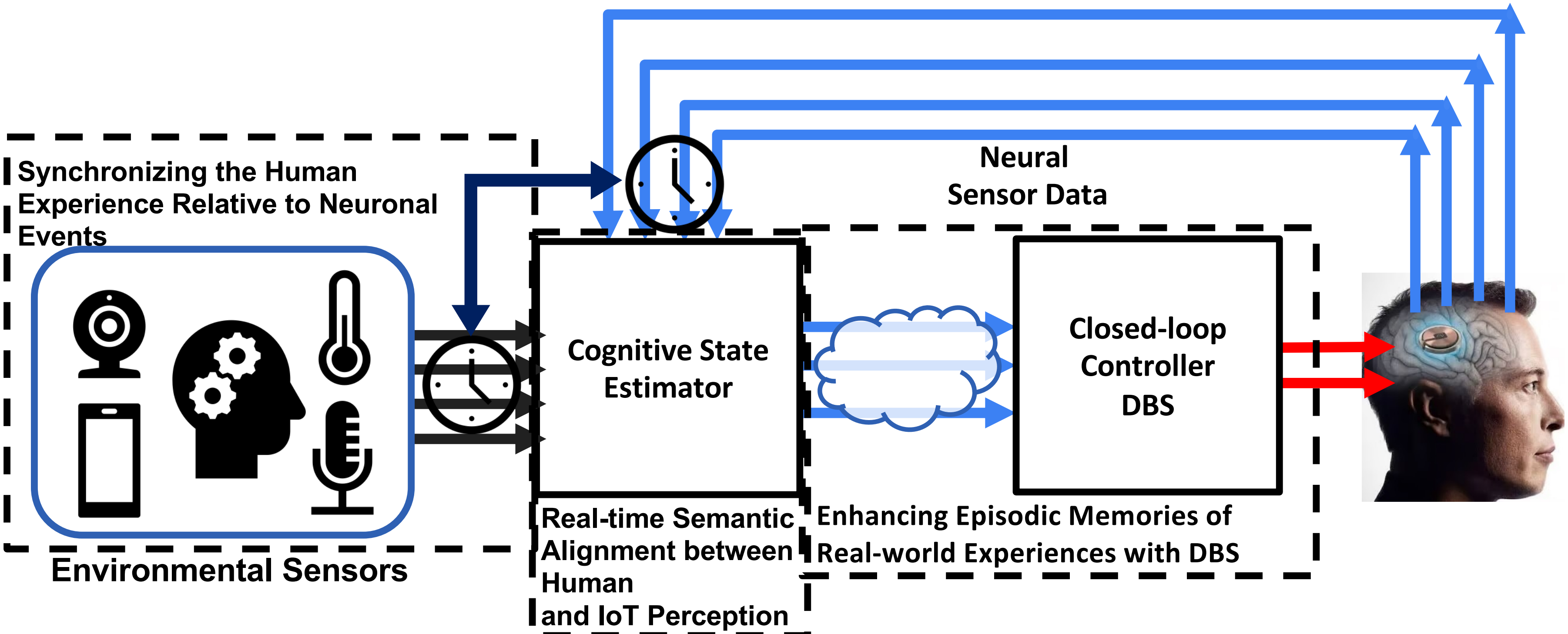
"trying to isolate a neuron signal from an electrode is like holding up a mic in a stadium to figure out who is speaking"



- Parkinson's treatment
- Seizure mitigation
- Memory enhancement
- etc.…

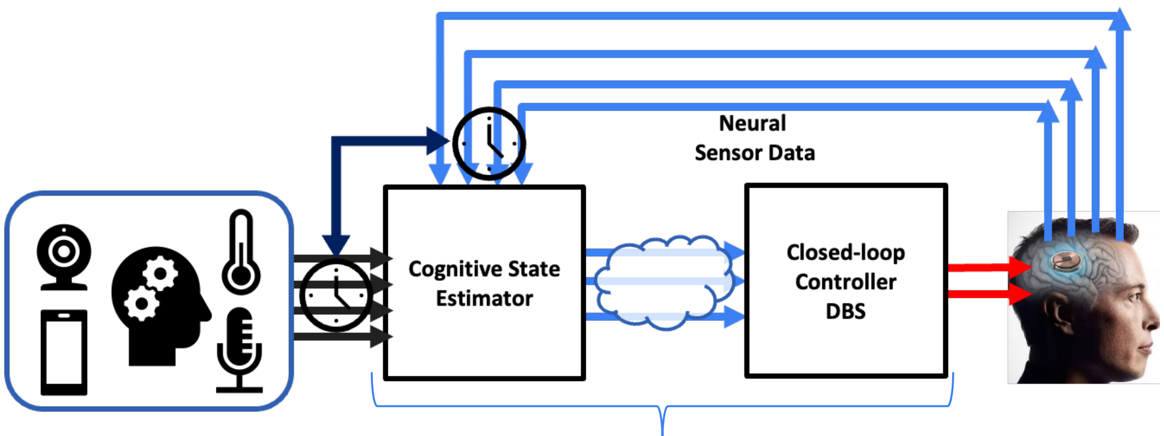
A simplified view of AI-enabled Deep Brain Stimulation

Augmenting Deep Brain Stimulation with Environmental Context



A simplified view of AI-enabled Deep Brain Stimulation

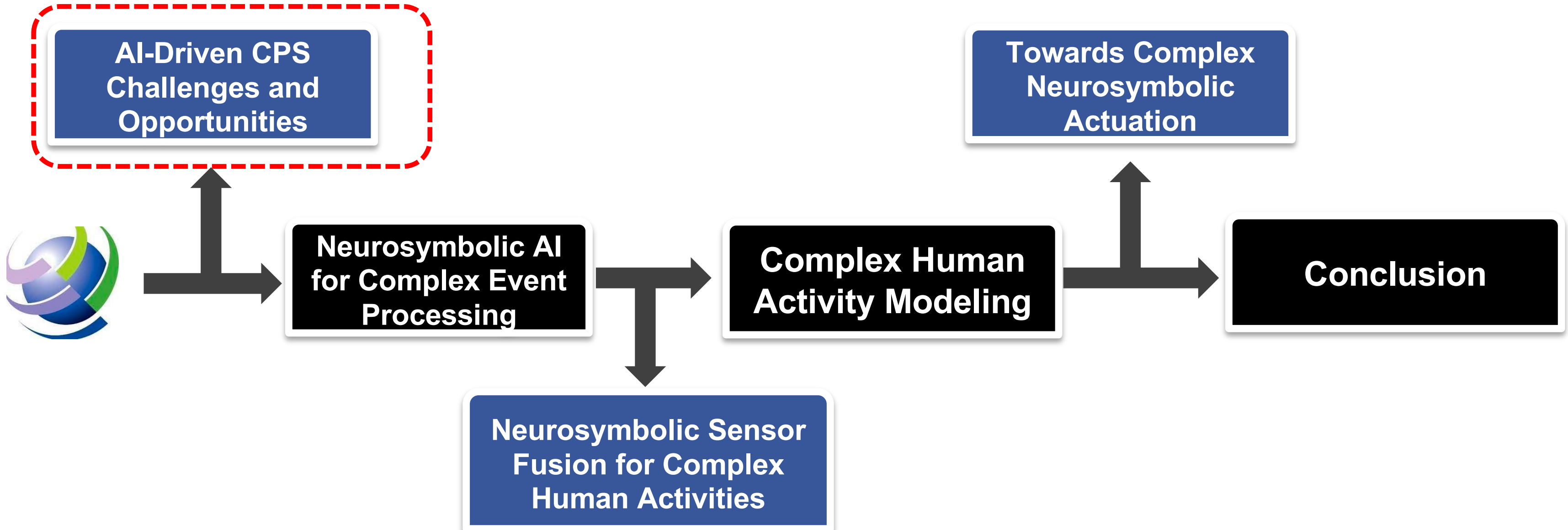
Can we maintain *explainability* and *intervenability* of AI-enabled Deep Brain Stimulation?



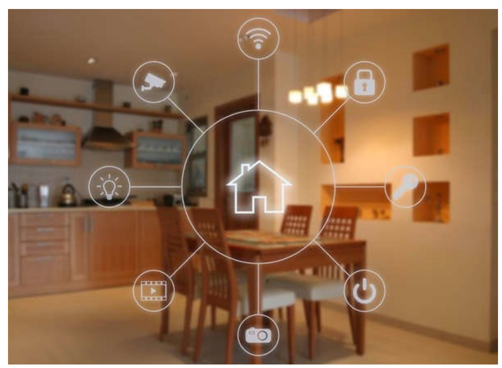
Blurry Neuroscientist/Programmer Requirements:
- Safety guarantees
- Proficiency and understanding
- Monitoring and feedback
- Adapting to patient needs
- Patient-centered design

Blurry Patient Requirements: (from Klein et. al 2016)
- Control over device function
- Meaningful consent
- Authentic self
- Relationship effects
- Safety/Security/Privacy

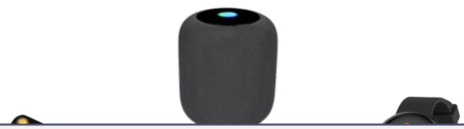
Outline for Today's Talk



Explosion of IoT Devices in Our Environment

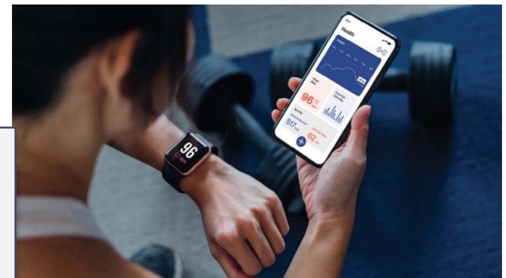


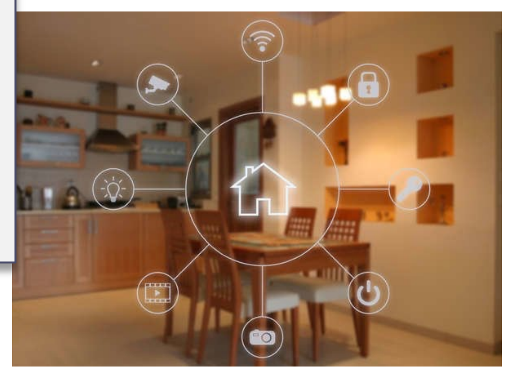
Explosion of IoT Devices in Our Environment



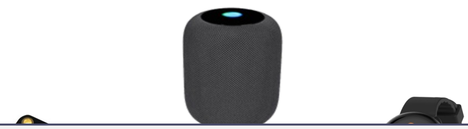
IoT Traditionally

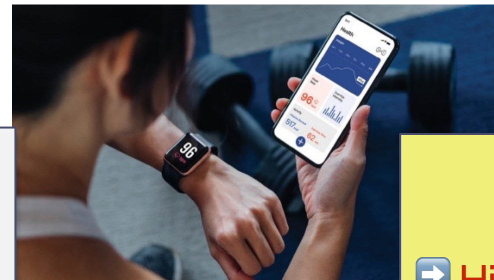
- Low-dimensional structured sensor data (e.g., temperature, humidity, etc.)
- Tasks requiring simple inferences
- Mechanistic or first-principles models, and simple data-driven models





Explosion of IoT Devices in Our Environment



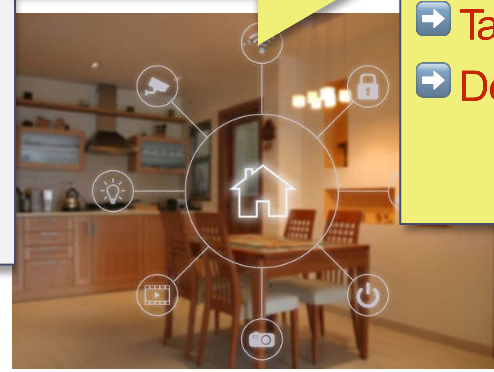


IoT Traditionally

- Low-dimensional structured sensor data (e.g., temperature, humidity, etc.)
- Tasks requiring simple inferences
- Mechanistic or first-principles models, and simple data-driven models

AI-enabled IoT

- High-dimensional unstructured sensor data (e.g., image, acoustic, lidar, etc.)
- Tasks requiring complex inferences
- Deep neural networks, and other large data-driven models



M. Srivastava, CPSWeek ‘23

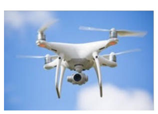
A Nexus Driven by Technology Trends


Acoustic Array

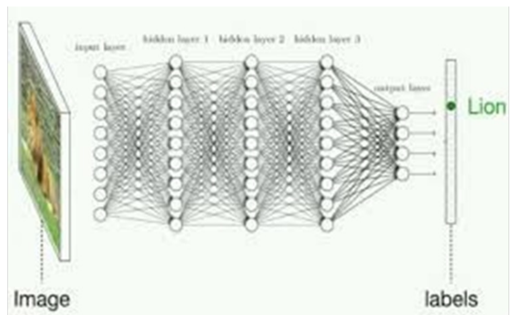

Camera

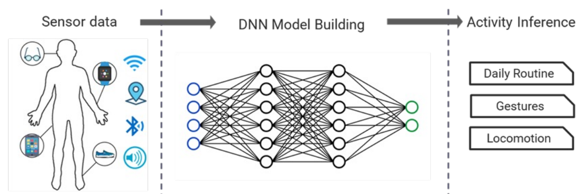

mmWave Radar


LIDAR


UAVs

Rich Sensors & Actuators



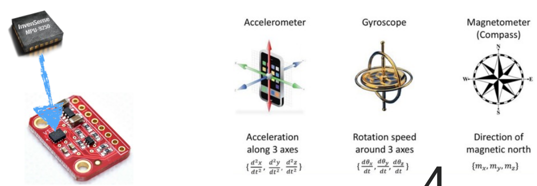


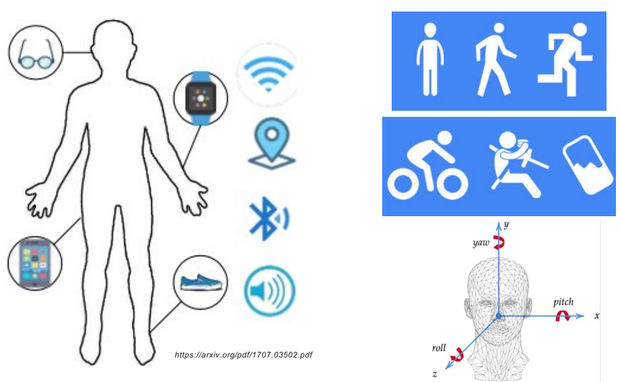
Deep Learning

Accelerators

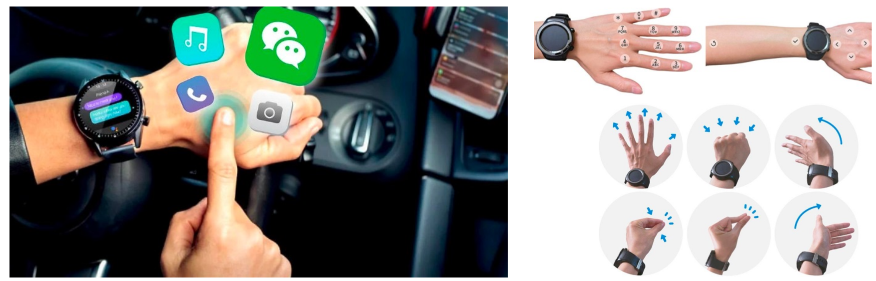
M. Srivastava, CPSWeek ‘23

Complex Inferences from Simple Sensors

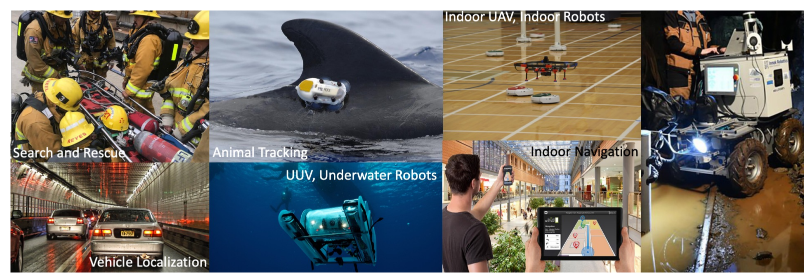




Human activity & behavior recognition



Interacting with wearable devices via on-body tapping



Accurate estimation of 3D motion trajectory

But many things are still missing…

Sensing Challenges in AI-enabled CPS

8

1
Domain Shift

2
Embedded Implementation

3
Combining Data and Knowledge

Sensing Challenges in AI-enabled CPS

1 **Domain Shift**

2 **Embedded Implementation**

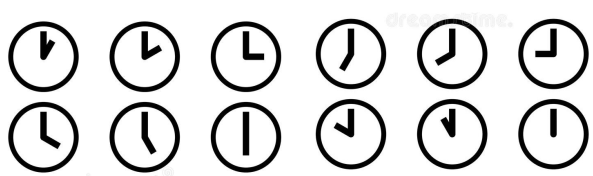
3 **Combining Data and Knowledge**

Many Forms of Domain Shifts in AI-enabled CPS

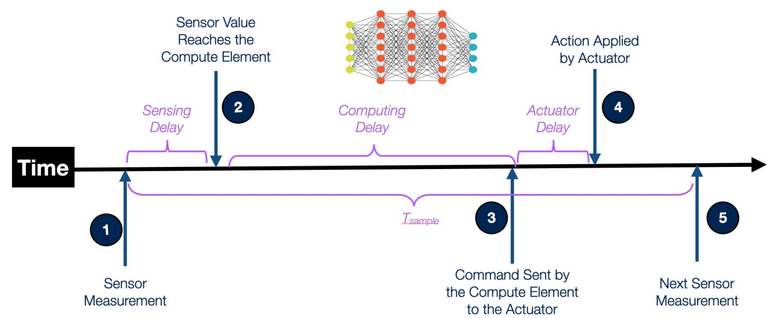
Person-to-person differences

Different environments

Variations in sensors

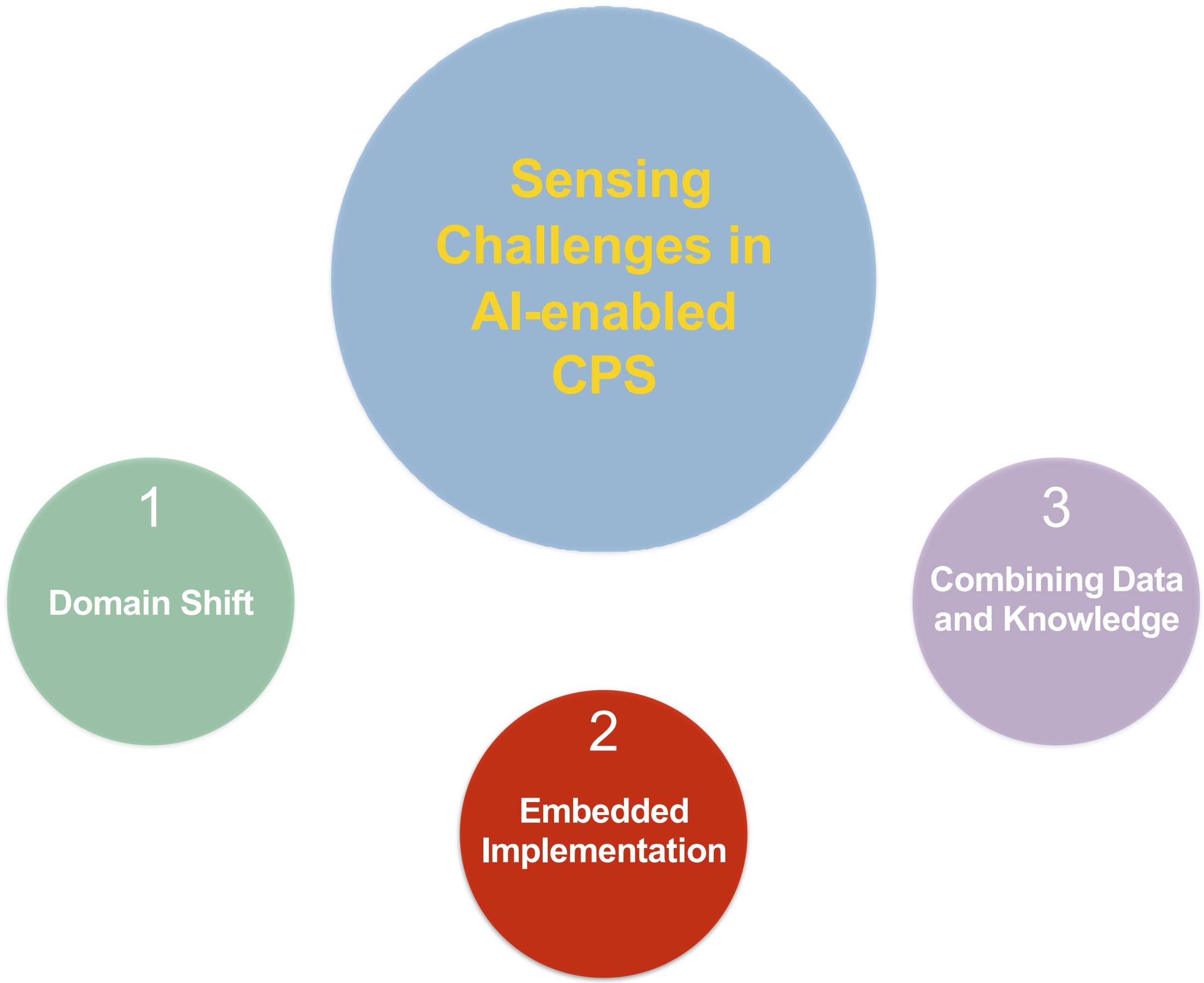


Misaligned time



Latency variations

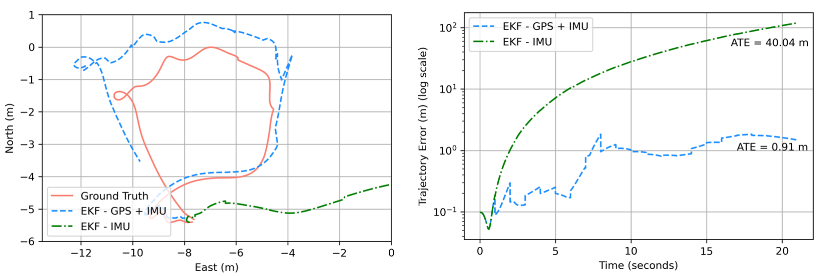
M. Srivastava, CPSWeek ‘23


Sensing Challenges in AI-enabled CPS
1
Domain Shift
2
Embedded Implementation
3
Combining Data and Knowledge

The Challenge of Embedded Implementation

- Neural network models promise better performance for many IoT applications, but due to the IoT platform resource-constraints and diversity the promise remains unrealized

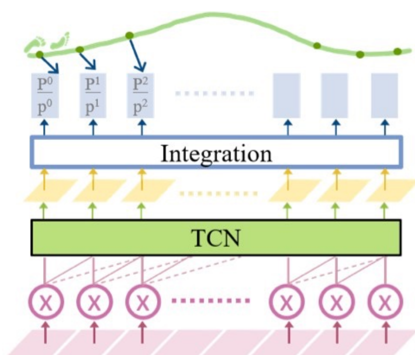
Example: Inertial Odometry on MCU-class Ultra Resource Constrained IoT



The curse of drift in inertial odomtery

Hardware	SRAM (kB)	Flash (kB)
Qualcomm CSR8670 (eSense platform)	128	16000
STM32F446RE	128	512
STM32F407VET6	192	512
STM32L476RG	128	1024
STM32F746ZG	320	1024

Ultra Resource Constrained IoT platforms



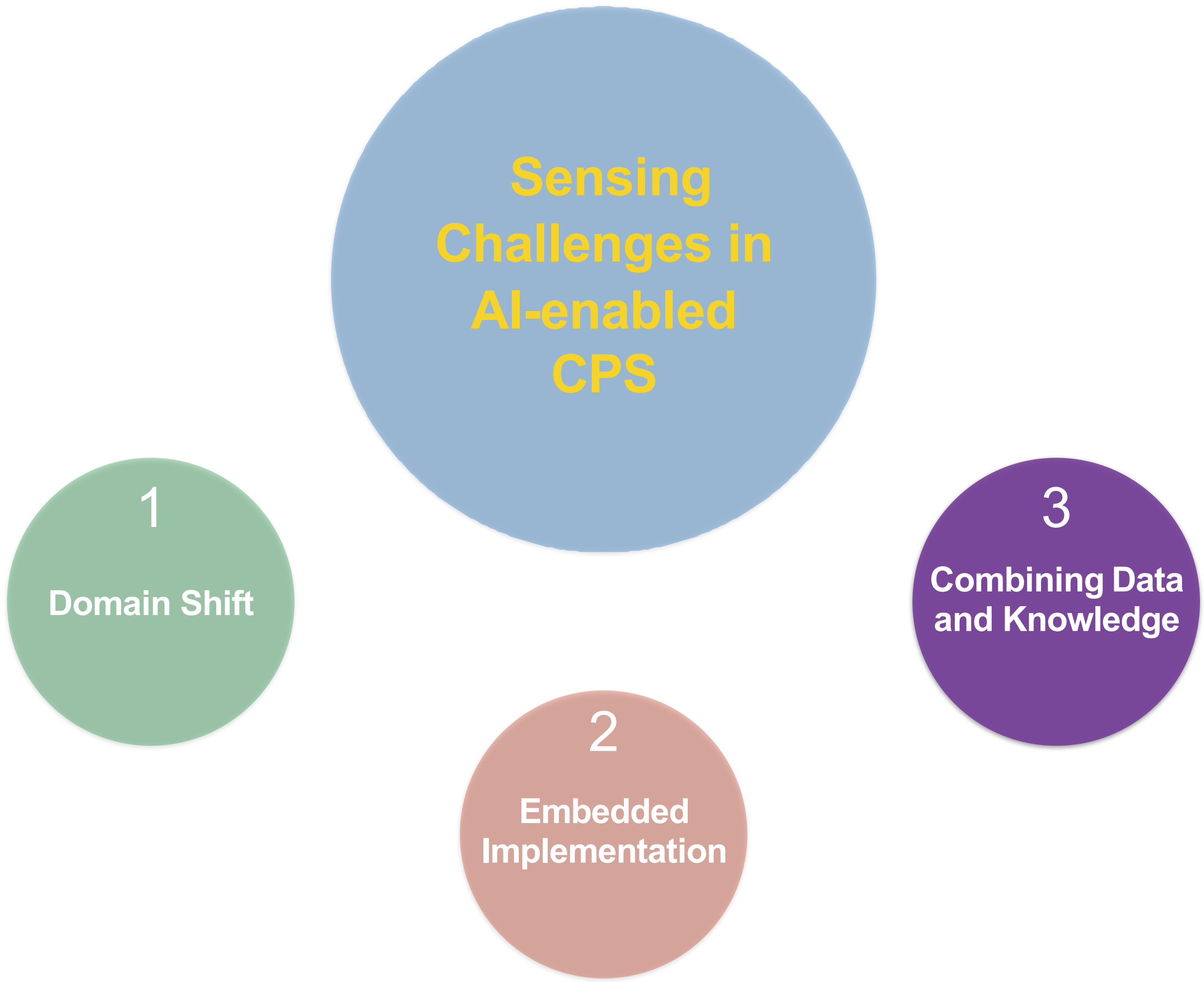
Example: RoNIN TCN

Naive double integration
SRAM=1.2kB, Flash=28.1kB
ATE=12398m, RTE=59.85m

Pedestrian Dead Reckoning
SRAM=10.8kB, Flash=49.6kB
ATE=34.81m, RTE=23.62m

RoNIN TCN
SRAM=2046.3kB, Flash=2195.5kB
ATE=4.73m, RTE=1.21m

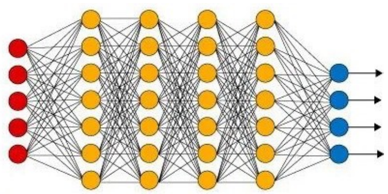
IMWUT '22



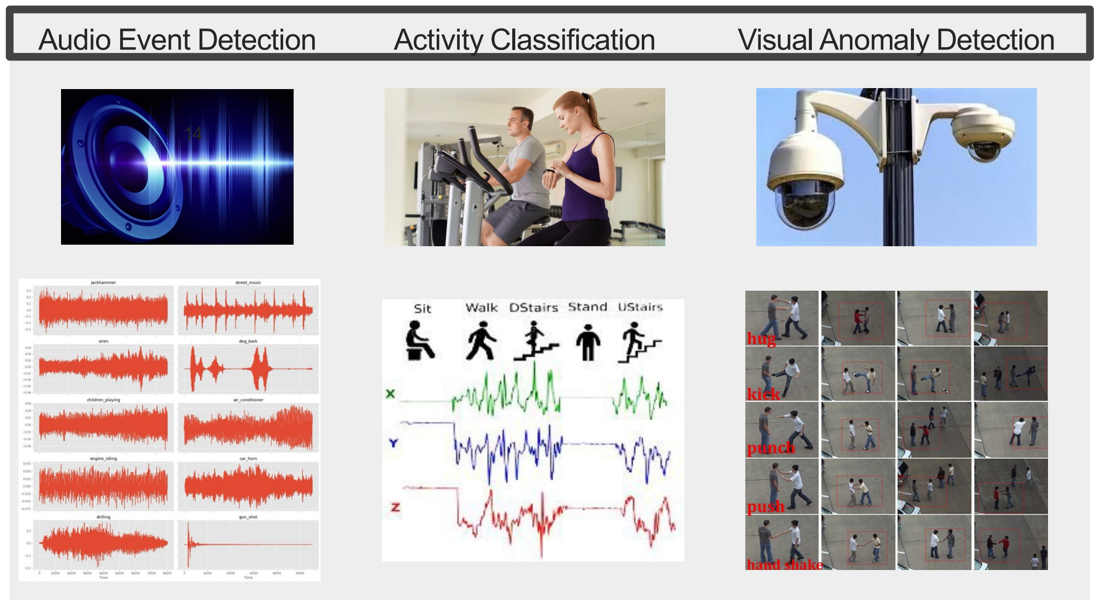
M. Srivastava, CPSWeek ‘23

Deep Learning for Perception

Excellent at detecting and classifying simple events and activities



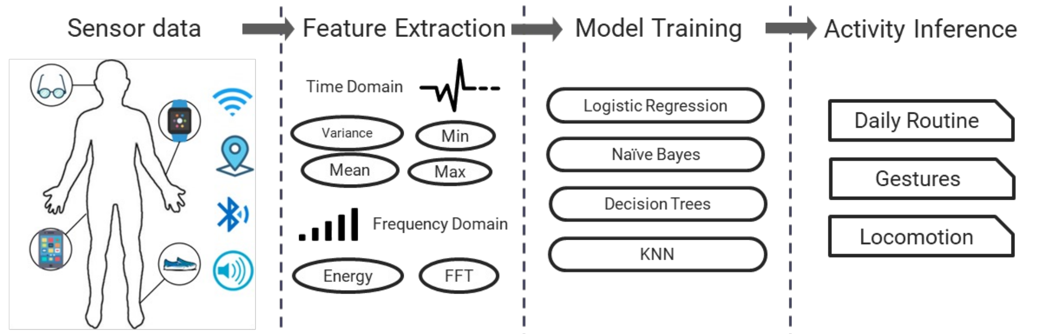
Deep Learning is **faster**, and **more accurate** than humans!



Traditional Methods vs. DNN's

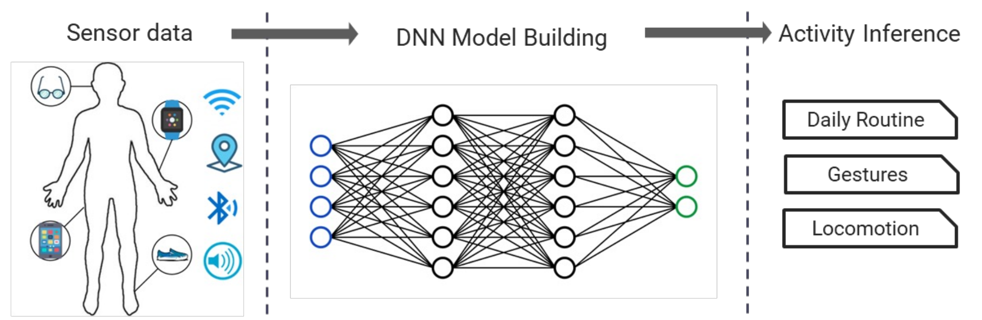
Traditional Methods

- Required Domain Expertise
- Feature Extraction
- SVM/Decision Trees
- Not scalable



Deep Neural Networks (DNN)

- Less Domain Expertise
- Applied on raw sensor data
- High Performance
- Scalable



M. Srivastava, CPSWeek '23

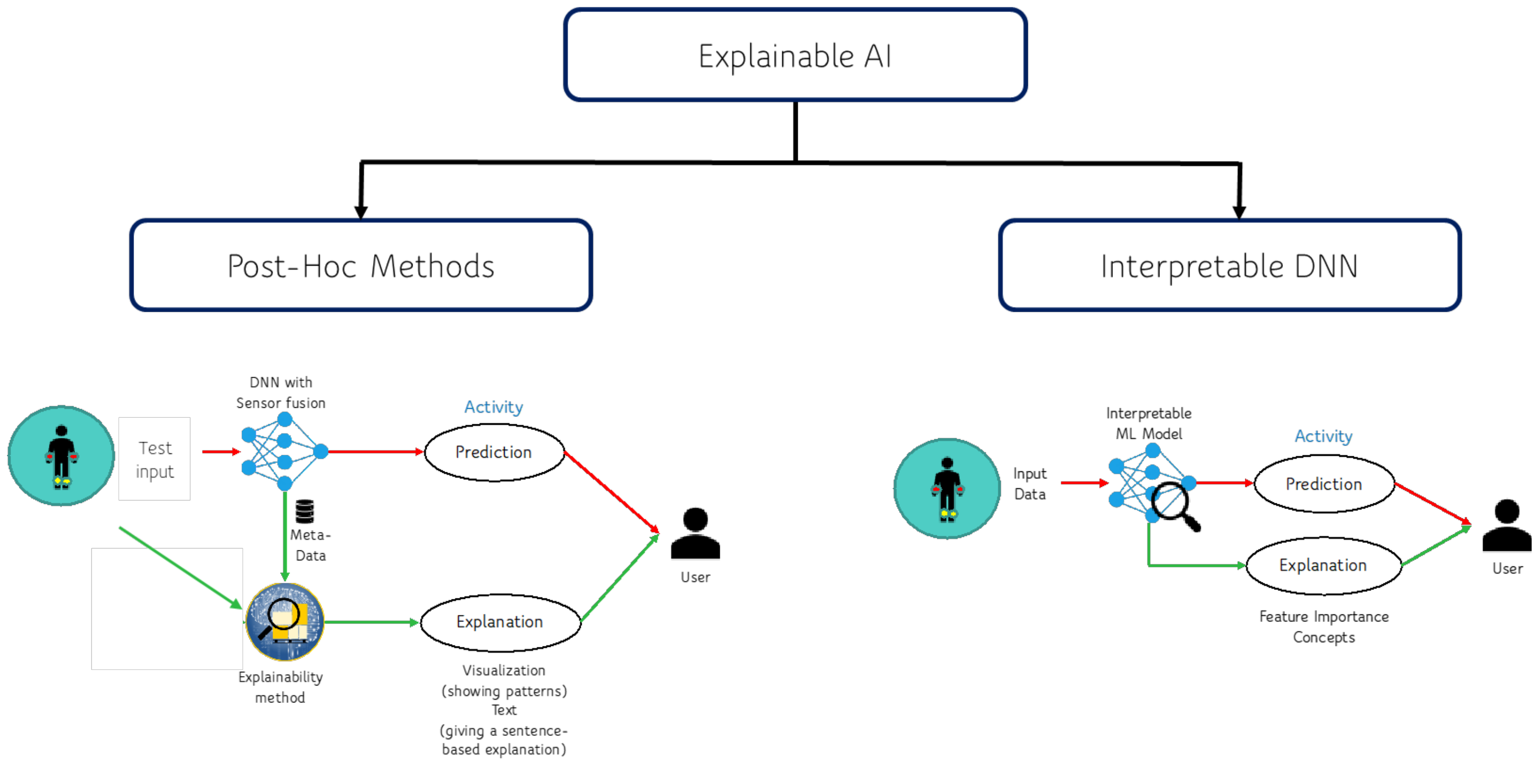
Combining Data And Knowledge
Problem #1: *Explainability* and *Tellability*

All of the above challenging with data-driven models but much easier with traditional first principles (symbolic) models.

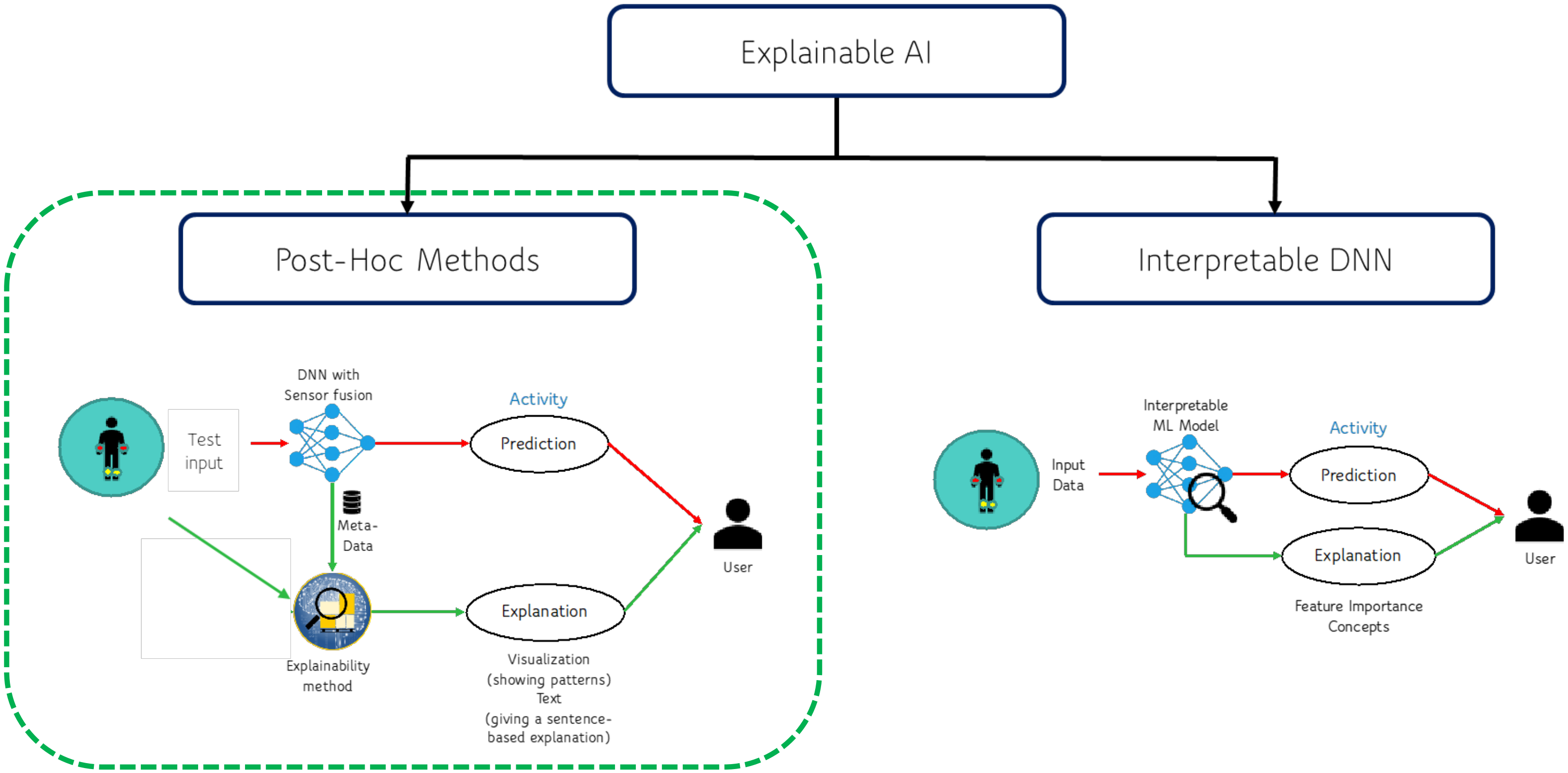
A Sea of DNN Explanation Methods

Symonian '13 Gradient	Zeiler'14 Occlusions	Zhang '16 Excitation BP	Zintgraf'17 Pred Diff	Zhang'18 Explanatory Graph	Ancona'19 Polynomial SHAP
Landecker '13 Contrib Prop	Haufe'15 Pattern	Ribeiro'16 LIME	Montavon '17 Deep Taylor	Ye'18 CNN Framelets	Goyal'19 Counterfactual
Brazen '13 Taylor	Bach '15 LRP	Shrikumar '17 DeepLIFT	Selvaraju '17 Grad-CAM	Yang'18 Recursive Partitioning	Kuo'19 Interpretable CNN
Zeiler '14 Deconv	Caruana '15 Fitted Additive	Lundberg '17 Shapley	Kindermans '17 PatternNet	Vaughan'18 Additive Index	Liantao'20 AdaCare
Springenberg '14 Guided BP	Zhou '16 GAP	Fong '17 M Perturb	Sundarajan'17 Int Grad	Caicedo '19 ISeeU	Jianbo'20 LS Tree

How should we explain DNNs?



How should we explain DNNs?



Can we use post-hoc explanations for Sensor Data?

NeurIPS ‘20

Post-Hoc Methods Considered

Perturbation Based

- LIME
  - Creates a local surrogate model
- Anchor
  - If-else rules

Cons
- Lots of hyper-parameters
- Inconsistent over multiple runs

Saliency Based

- Gradients
- GradCAM
- SHAP

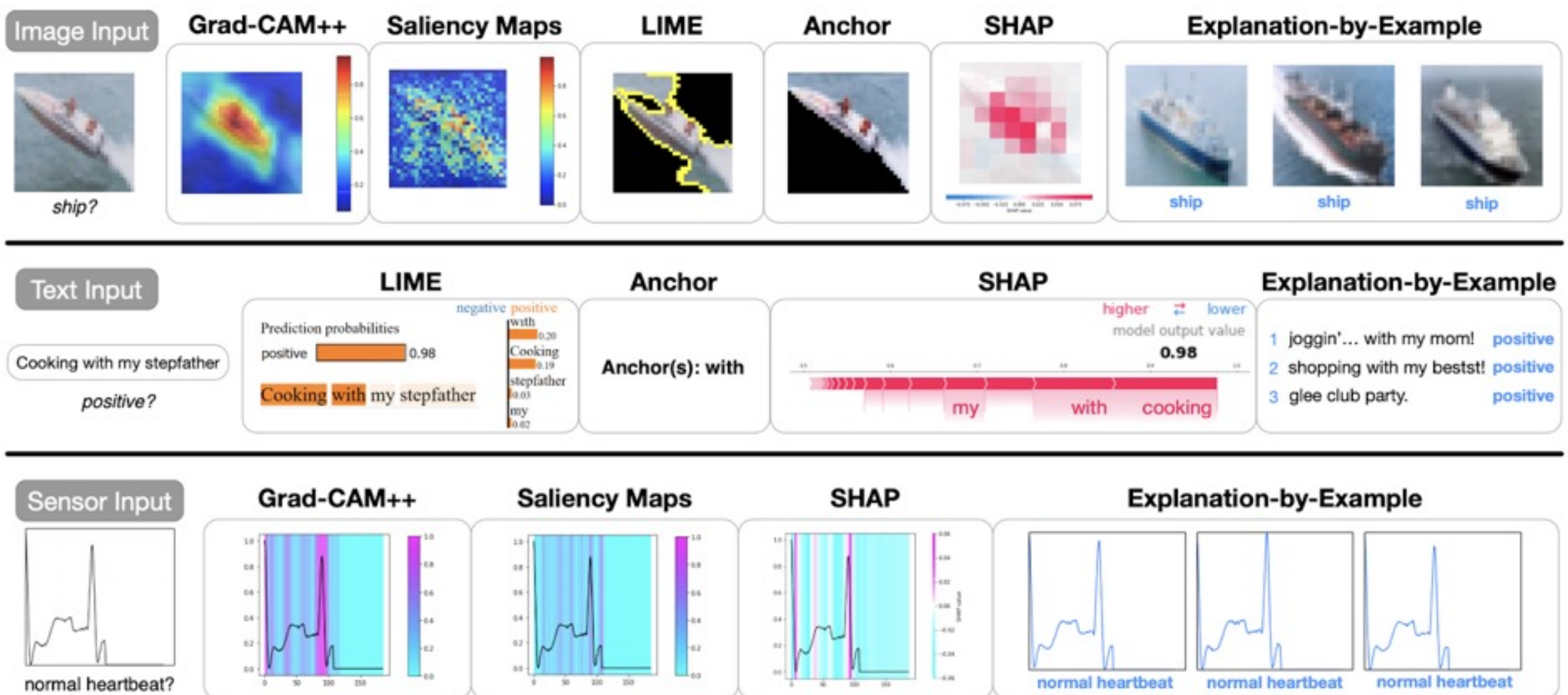
Cons
- Mainly designed for images
- Same saliency regions

Explanation by Examples

Provides a few key perceptually-relevant items from the training dataset

Cons
- Requires Training data
- Privacy concerns

Post-hoc Explanations



Results

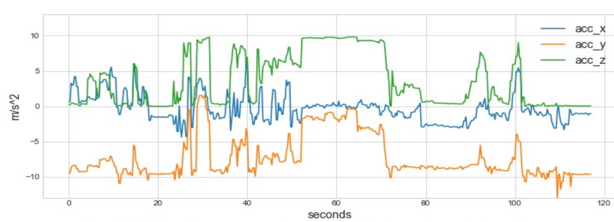
Identify the Human Preferred Explanation Method

Explanation Method	Image Study	Text Study	Audio Study	ECG Study
LIME	47.7 ± 4.5%	**70.4 ± 3.6%**	-	-
Anchor	38.9 ± 4.3%	25.8 ± 3.5%	-	-
SHAP	33.7 ± 4.3%	59.9 ± 3.8%	34.7 ± 4.8%	32.8 ± 3.3%
Saliency Maps	39.4 ± 4.3%	-	46.1 ± 5.1%	40.4 ± 3.5%
GradCAM++	50.8 ± 4.5%	-	48.1 ± 5.3%	42.0 ± 3.5%
Explanation by Examples	**89.6 ± 2.6%**	43.7 ± 3.9%	**70.9 ± 4.7%**	**84.8 ± 2.5%**

Results indicate the rate by which users selected a particular method when it is an available explanation, with 95% bootstrap confidence intervals

What did we learn from our study?

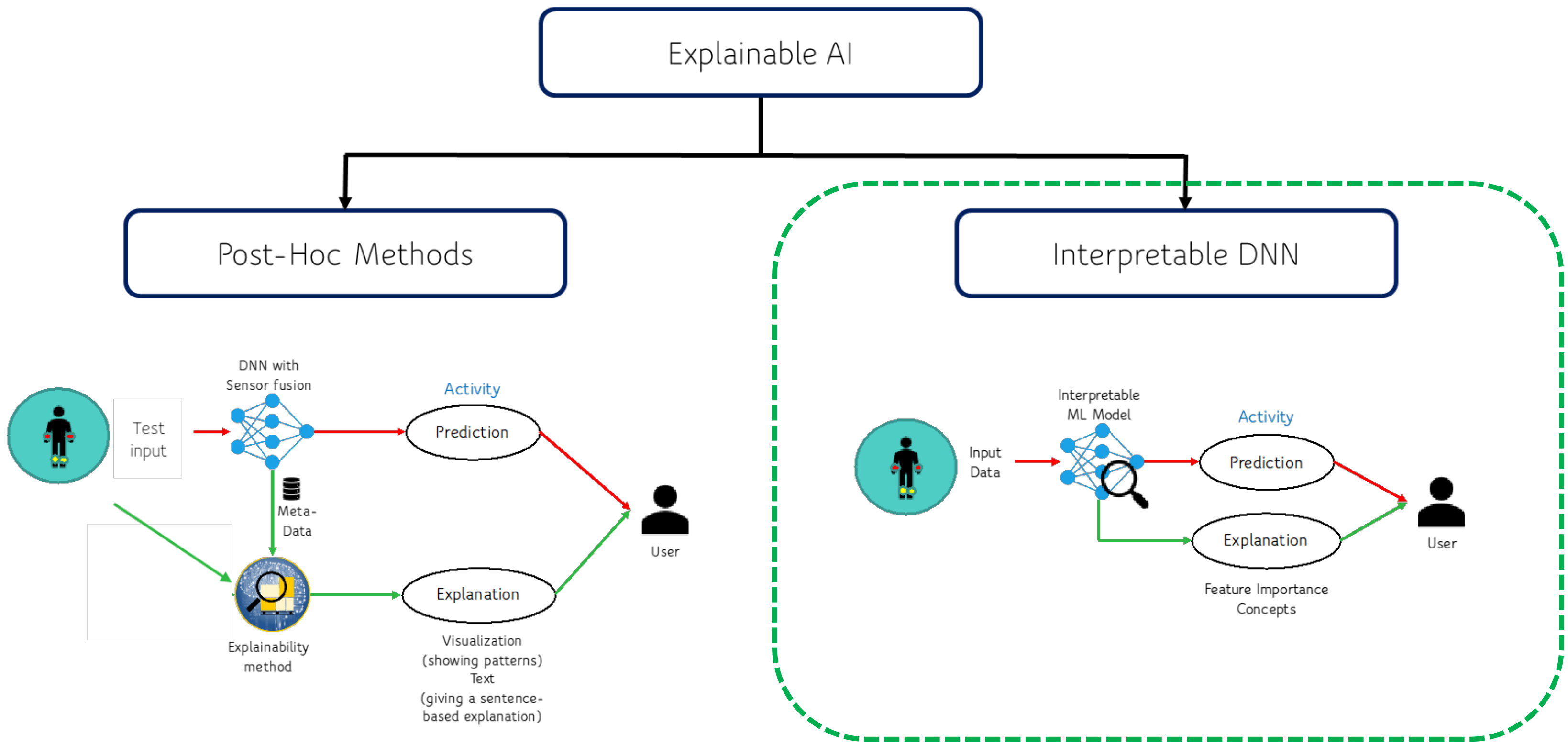
- Most of these methods are designed for images and text
- The explanations are not reliable
- Although explanation by examples is preferred, it is not suitable for multivariate time-series data
    - E.g., IMU data or videos



Predicted Activity: Using Restroom

Predicted Activity: In Play

How should we explain DNNs?



Concept-based explanations

Concept-based Interpretable DNNs

Force the DNN to Learn Interpretable Representations at hidden layers

Concepts differ from traditional feature engineering:

- Concepts are high-level and are human understandable
- Feature engineering constructs low-level features that can be computed by functions

Properties
- Stable
- Relative Faithfulness
- Easy to comprehend

Concept Bottleneck Model (CBM)

Supervised Training :

- The Dataset has the concepts labeled
- Intermediate layer bottlenecks on human-specified concepts
- Model first predicts the concepts, then uses only those predicted concepts to make a final prediction (x -> c -> y)

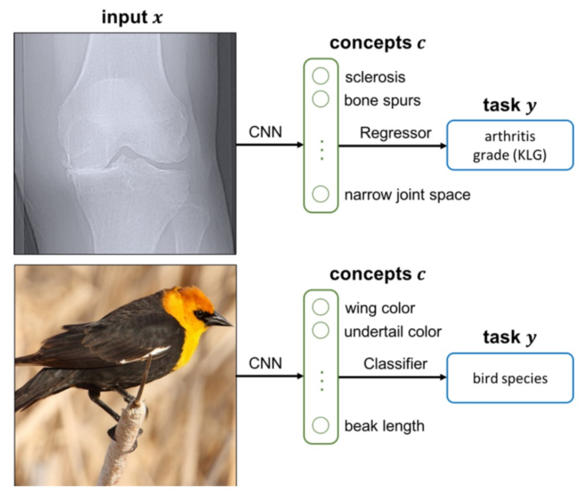


Figure 1. We study concept bottleneck models that first predict an intermediate set of human-specified concepts c, then use c to predict the final output y. We illustrate the two applications we consider: knee x-ray grading and bird identification.

Pang et.al. “Concept-Bottleneck Models”, ICML 2020

Limitations of CBM

- CBMs are designed for Image classification tasks
- Concepts are simple with the same level of abstraction, e.g., visual elements present in a single image.
- The concepts are assumed to be given a priori by a domain-expert in the dataset
  - This may not result in a necessary and sufficient set of concepts
  - Time consuming to annotate data with all the concepts
- For complex tasks like video activity classification, the concepts can represent relationships between objects spanning multiple frames
- They don't capture the temporal relationships between concepts

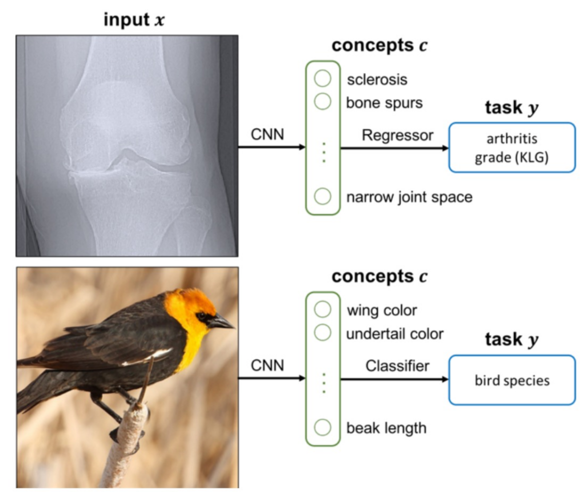
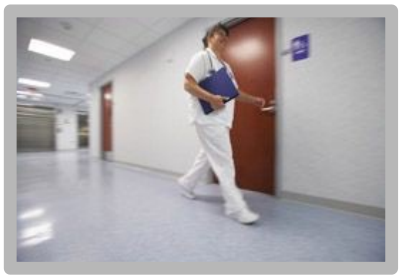
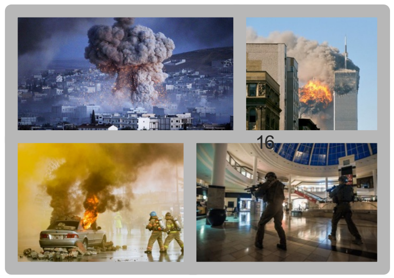


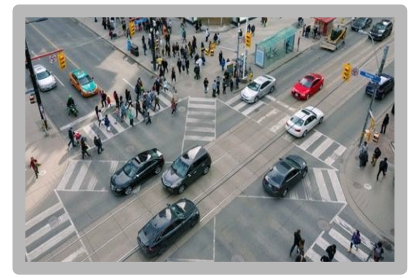
Figure 1. We study concept bottleneck models that first predict an intermediate set of human-specified concepts c, then use c to predict the final output y. We illustrate the two applications we consider: knee x-ray grading and bird identification.

Combining Data And Knowledge
Problem #2: *Complex Events*


Unsanitary Operation


Coordinated Attack

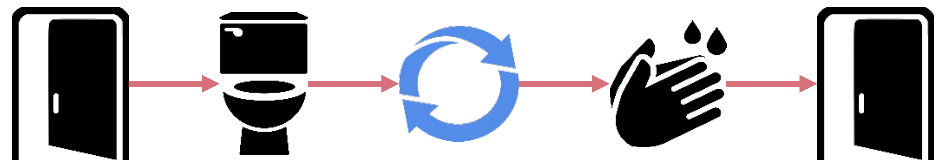

Unattended Bag


Traffic Rule Violation

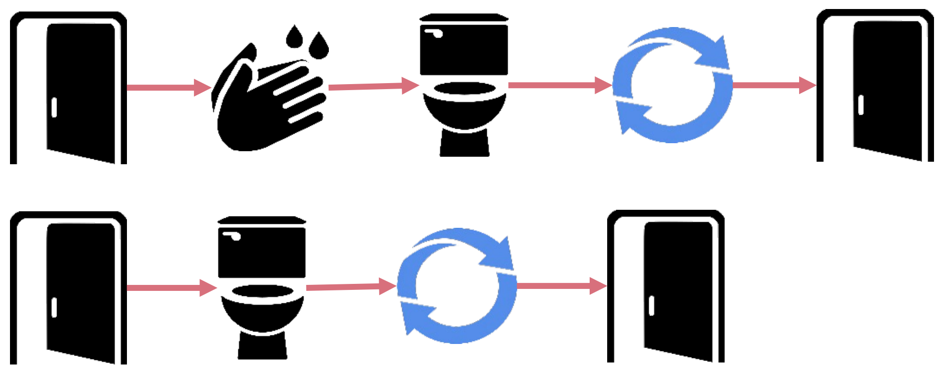
- **Connect the dots across atomic events**
  - At different locations, by different actors, across arbitrary intervals of time
- Require (i) **Perception of atomic events** from unstructured, high-dimensional, noisy, and possibly multimodal data, and (ii) **High-level reasoning** over the atomic events

Complex Activity Example

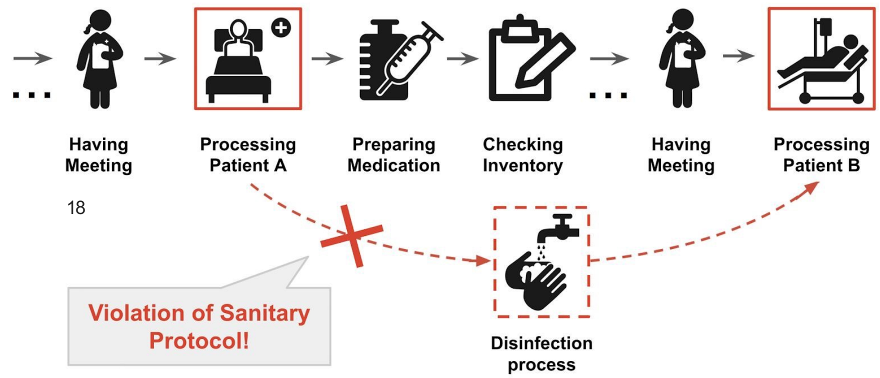
Using Restroom (Hygienic)



Using Restroom (Unhygienic)



Complex Events are challenging for Deep Learning models



A nurse forgets to wash their hands between processing different patients.

- **Needle in the haystack** problem
  - Pattern in atomic events taking place over long spans of time
  - Involve atomic events from many different sensors
- The ***effective context size*** **is limited** in deep neural networks for purposes of complex event sensing (high rate, long time spans), even with new transformer architectures

Modeling Long-term Dependencies Requires Memory

Models	Related Work	Effective Context Size
RNN / LSTM and Variants	Bi-LSTM [Singh et al. CVPR'16] CRNN [Cakir et al.]	Around 200-400 time steps with large LSTM model A few seconds (4-10) on visual & audio analytics tasks
Convolution Based	TCN [Lea et al. ECCV'16]	A larger receptive field of about 10s on video-based action classification
Transformer/Attention	TransformerXL [Dai et al. Arxiv'19], BERT, GPT model, Informer [Zhou et al. AAAI'21]	Time-series forecasting on hundreds to 1K of steps. NLP: sentence → paragraph → article

Detecting complex events with sampling rates of typical sensors require vastly larger context sizes

Bridging Deep Learning and Symbolic Models in AI-Driven CPS

Deep Learning Models

- Accelerator-friendly computation
- Excel at extracting complex short timescale events from unstructured, high-dimensional, sensory data
- Data-hungry
- Lack transparency and interpretability
- Poor at incorporating domain knowledge

Symbolic Models

- Work well at reasoning with structured data in human understandable ways
- Represent complex spatial & temporal dependencies efficiently and effectively
- Assured performance while incorporating domain knowledge
- Not accelerator friendly
- Can't handle unstructured & noisy data

Bridging Deep Learning and Symbolic Models in AI-Driven CPS

Deep Learning Models

- Accelerator-friendly computation
- Excel at extracting complex short timescale events from unstructured, high-dimensional, sensory data
- Data-hungry and poor at capturing Css
- Lack transparency and interpretability
- Poor at incorporating domain knowledge

Symbolic Models

- Work well at reasoning with structured data in human understandable ways
- Represent complex spatial & temporal dependencies efficiently and effectively
- Assured performance while incorporating domain knowledge
- Not accelerator friendly
- Can't handle unstructured & noisy data

Perception
(System 1)

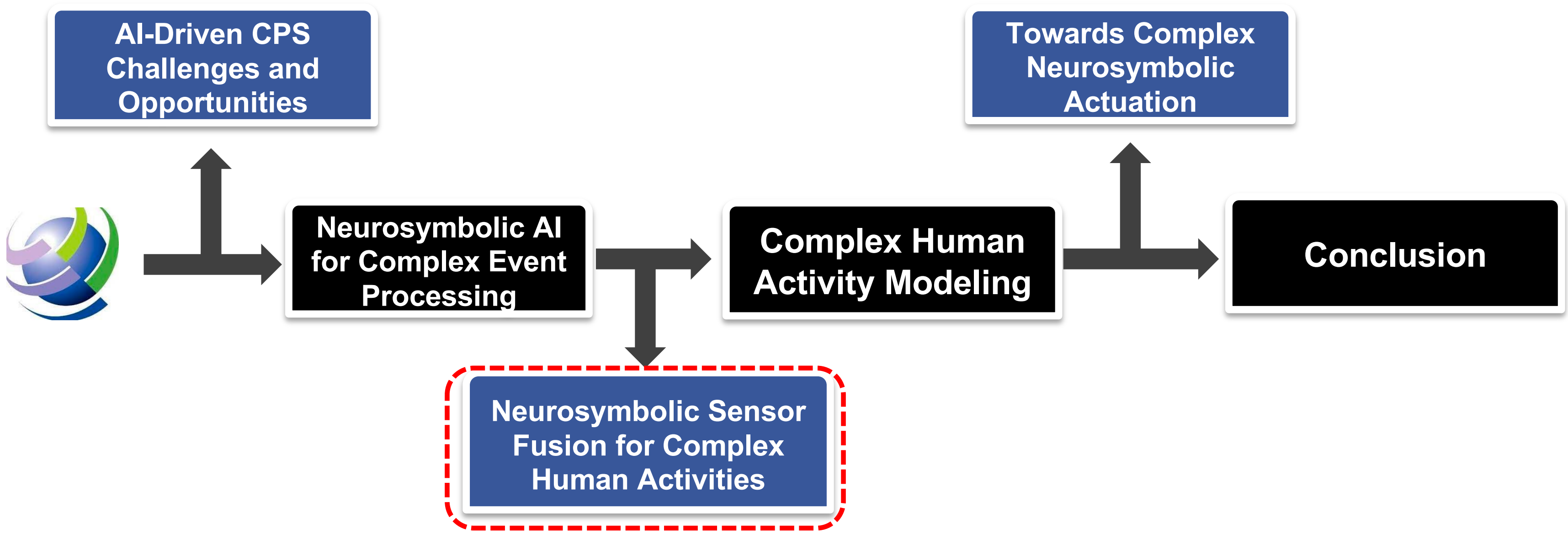
Reasoning
(System 2)

A hybrid "Neurosymbolic" approach?

- Inspired by how human process CE
- Combine the power of the DL & Logic approaches.

M. Srivastava, CPSWeek '23

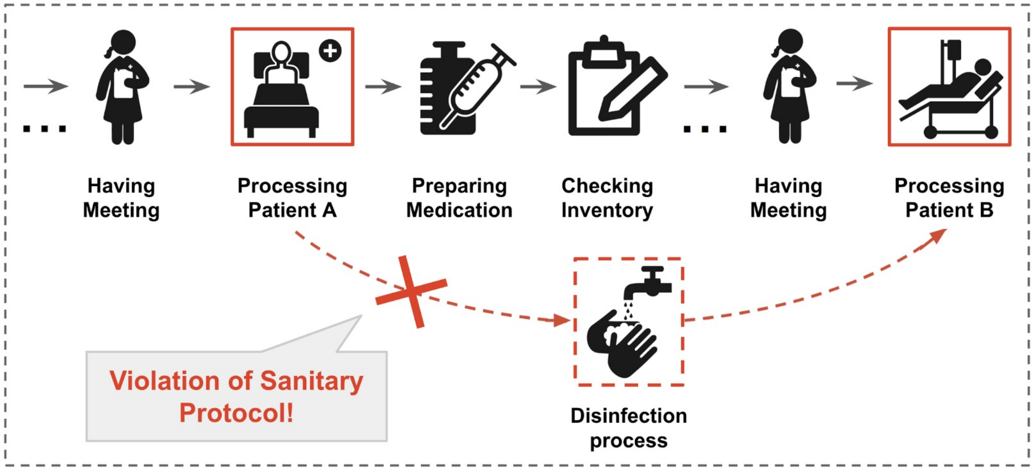
Outline for Today's Talk



Neuroplex: Learning to Detect Complex Events in Sensor Networks Through Knowledge Injection

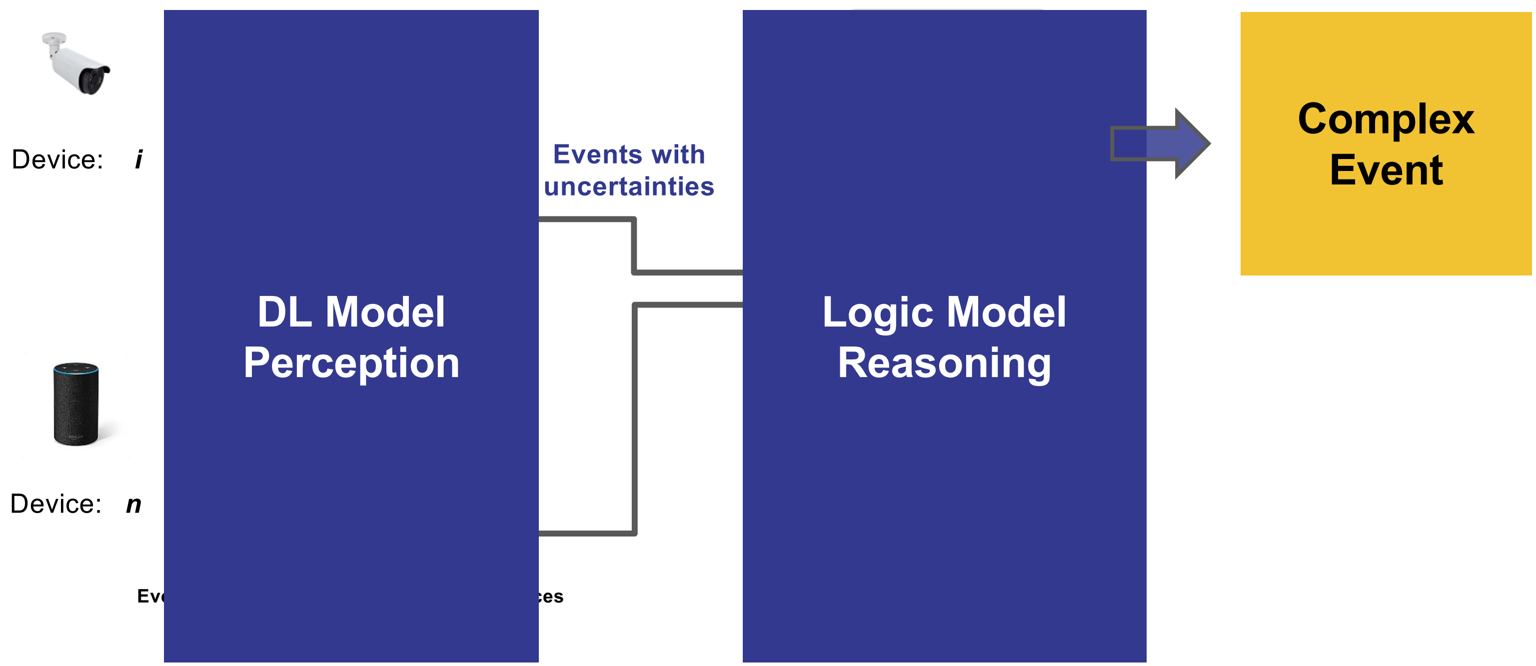
SenSys '20

Complex Event Detection



Simple Events compose Complex Events

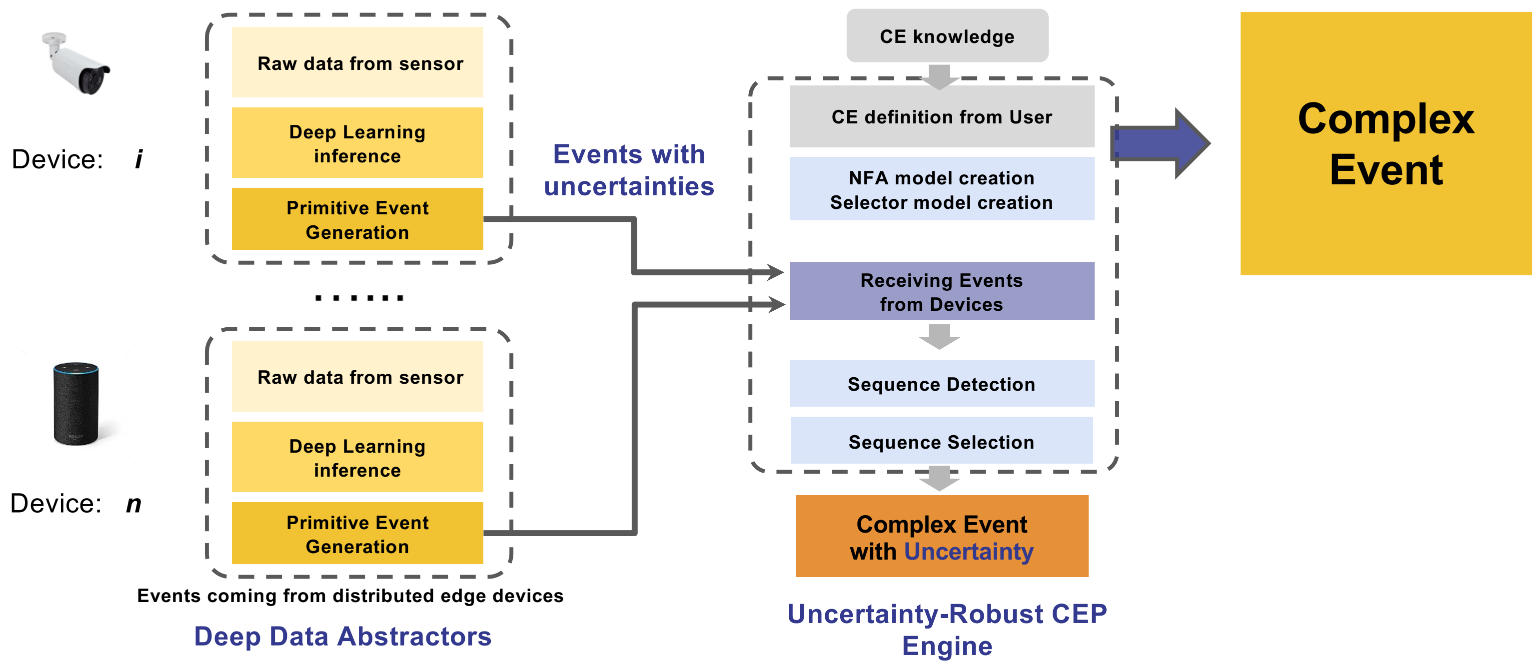
Neuroplex Inference: Deep Learning Perception + Logical Reasoning



DeepCEP SMARTCOMP ‘19
NeuroPlex SenSys ‘20

Neuroplex Inference: Deep Learning Perception + Logical Reasoning

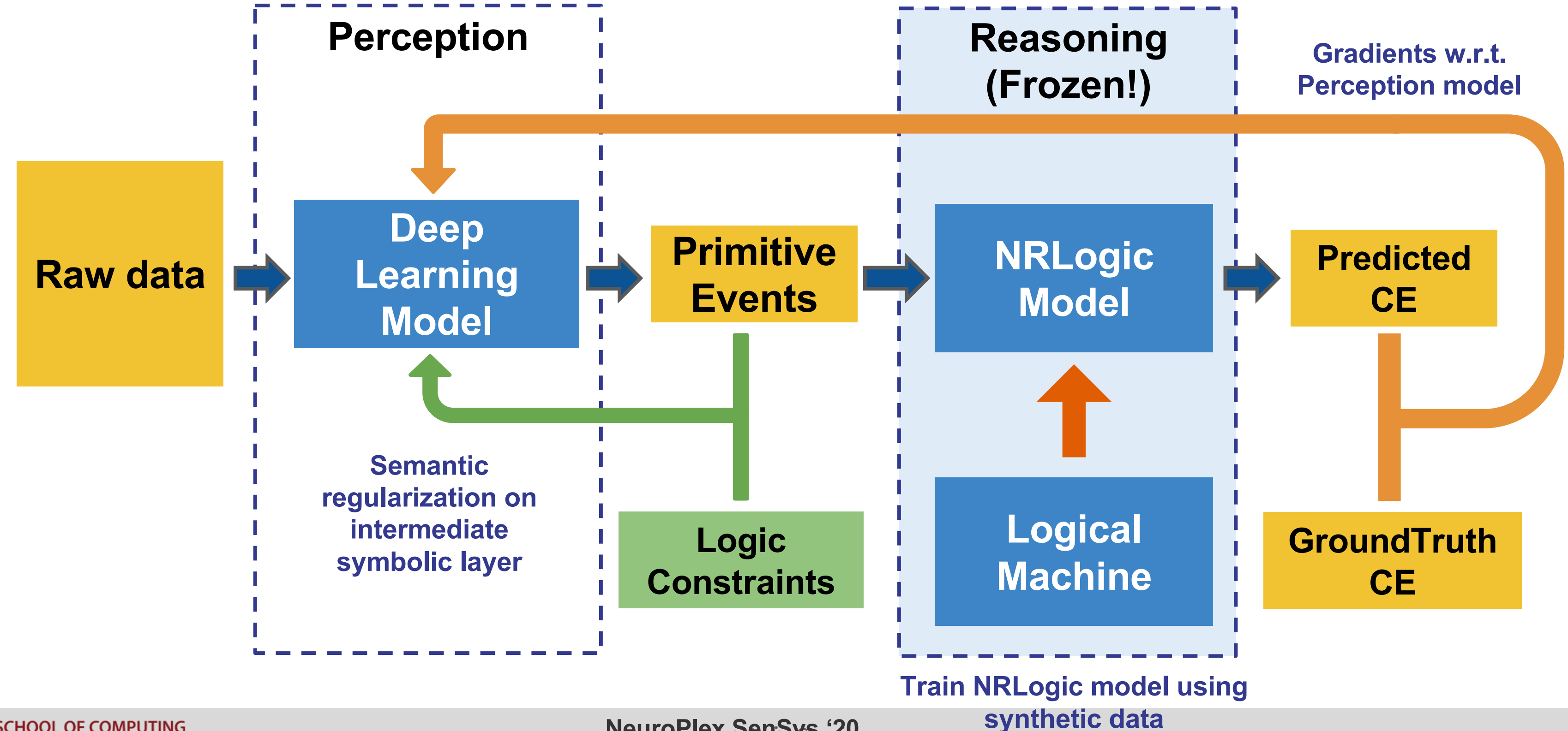
Leverage the Power of Deep Learning + Logic for Complex Event Reasoning



DeepCEP SMARTCOMP '19
NeuroPlex SenSys '20

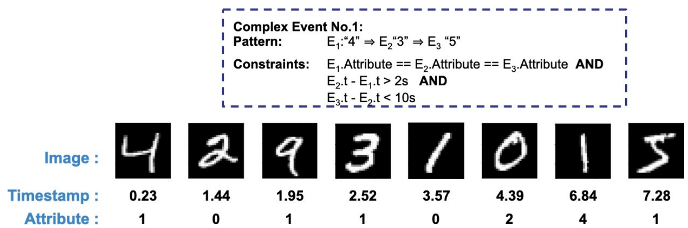
Neuroplex: End-to-end Training

We can fine-tune both deep learning perception and complex event pattern detection



Neuroplex: Performance

CE over irregular time series of images



Performs much better than DL-only baselines

	Oracle	NEUROPLEX	NEUROPLEX (w/o)	CRNN	C3D
Perception Acc	99.19%	**98.87%**	70.55%	10.09%	NA
Validation MAE	0.002	**0.013**	0.065	0.523	0.176
Converted Acc	99.85	**99.39%**	96.02%	69.98%	88.47%

CE over images

	NEUROPLEX	ConvLSTM	ConvLSTM-2	LSTM-Attention
Perception Acc	**77.59%**	1.72%	NA	NA
Validation MAE	**0.0027**	0.1430	0.1860	0.6245
R-Square	**1.000**	0.882	0.807	0.002
Converted Acc	**100%**	93.67%	89.28%	78.81%

CE over IMU

CE over nurse activities (IMU)

	Complex Nursing Event Name	Complex Nursing Event logic
Complex Event	Physiological Measurement	Vital sign ⇒ blood glucose measure ⇒ blood collection
	Indwelling Drip	Vital sign ⇒ Indwelling drip
	Patient Cleaning	Oral care ⇒ Diaper exchange
Protocol Violation	Unsanitary Operation No.1	Diaper exchange ⇒ blood collection
	Unsanitary Operation No.2	Area cleaning ⇒ blood glucose measure
	Unsanitary Operation No.3	Diaper exchange ⇒ indwelling drip

CE on audio stream

	Event types	Length	Num
CE 1	cooking ⇒ eating ⇒ dishwashing	3	1213
CE 2	social_activity ⇒ cooking ⇒ eating	3	1198
CE 3	working ⇒ other	2	2898
CE 4	watching_tv ⇒ vacuum_cleaner	2	2904
CE 5	absence ⇒ eating	2	2844
CE 6	dishwashing ⇒ cooking	2	2888
CE 7	absence ⇒ social_activity	2	2919

Event types: 9 . Avg length: 2.29. Dataset size: 16162

Scales with context length

Methods	Sim 1	Sim 2	Sim 3	Sim 4	Sim 5	Sim 6
Time window (minutes)	10	20	30	40	50	60
R-square						
Neuroplex	**1.00**	**0.99**	**1.00**	**0.90**	**0.88**	**0.85**
ConvLSTM	0.88	0.90	0.66	0.32	0.33	0.35
ConvLSTM-2	0.81	0.76	0.80	0.76	0.75	0.70
AttentionNet	0.02	0	0	0	-0.01	-0.02
Converted Accuracy						
Neuroplex	**100%**	**98.90%**	**100%**	**83.59%**	**79.00%**	**79.63%**
ConvLSTM	93.67%	83.29%	67.75%	40.79%	39.03%	37.47%
ConvLSTM-2	89.28%	80.08%	75.70%	60.30%	45.83%	39.48%
AttentionNet	78.81%	2.60%	0.62%	0.50%	0.11%	0.02%

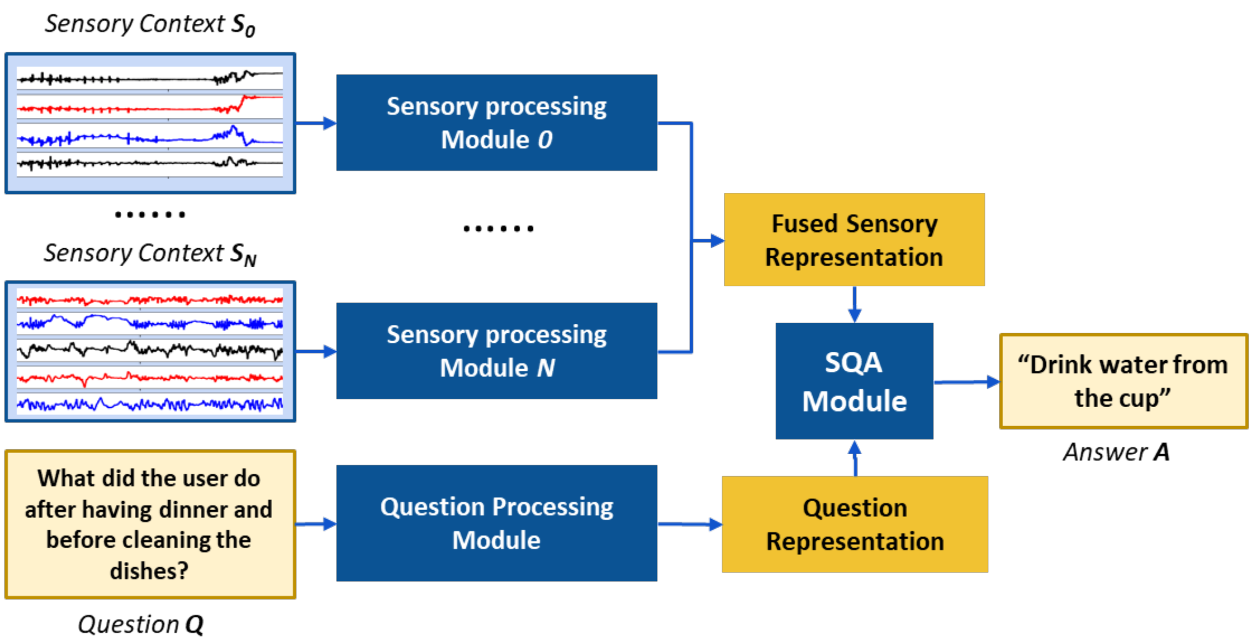
Neuroplex: End-to-end Training

We can fine-tune both deep learning perception and complex event pattern detection

Key Takeaways

1. Neurosymbolic models allow incorporating human knowledge
2. Help with learning efficiency and long context length
3. Also help with domain shift, explainability, constraints, etc.
4. Open issues relating to training, efficiency, and robustness

Follow-up:
DeepSQA: Generalized Sensor Question Answering (SQA) Framework



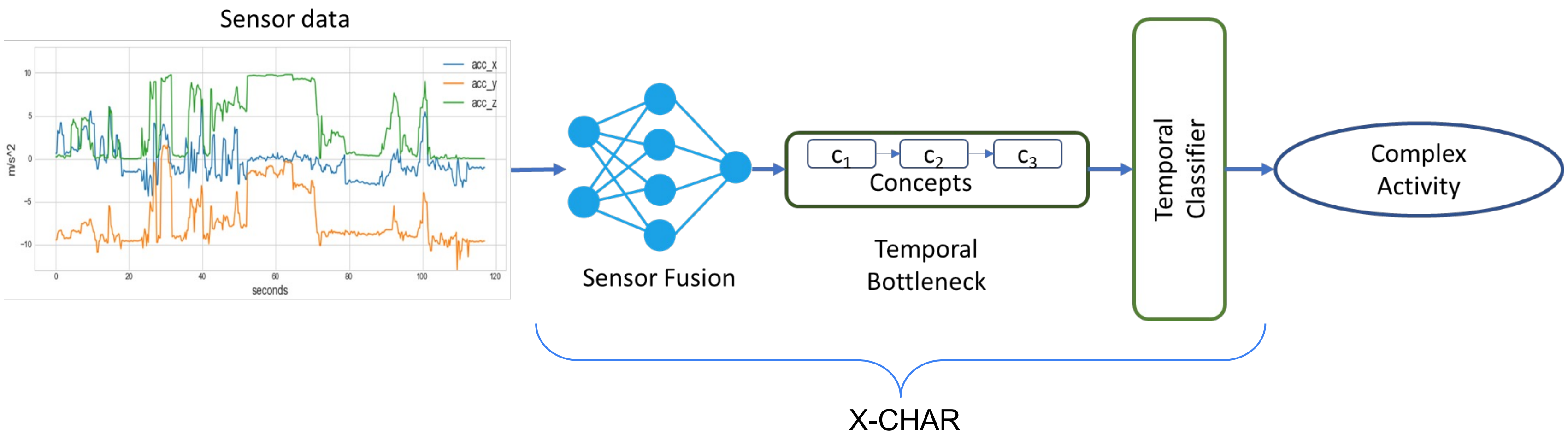
Generalized SQA framework.

Enable Flexible Querying (via Questions) for Complex Sensor Data

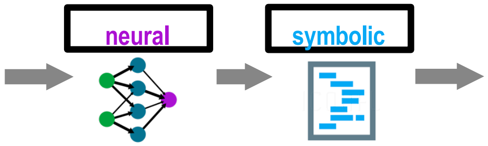
"DeepSQA: Understanding Sensor Data via Question Answering." IoTDI '21

Follow-up: Explainable Complex Human Activity Recognition (XCHAR)

- X-CHAR: an Interpretable DNN architecture for Complex activity recognition
- X-CHAR has a Temporal Concept Bottleneck layer
  - Use Connectionist Temporal Classification (CTC) loss to learn the concepts
- Use a classification model after the temporal bottleneck to get the complex activity

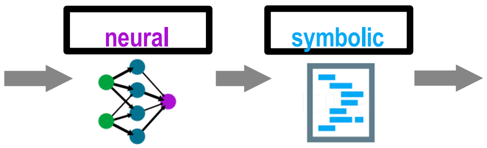


A Rich Neurosymbolic Landscape

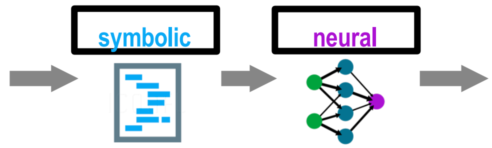


Symbolic-after-Neural
e.g., structured reasoning over natural sensor inputs

A Rich Neurosymbolic Landscape

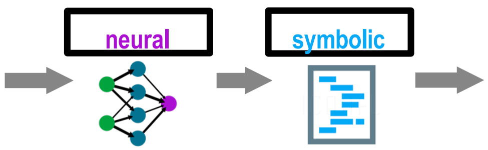


Symbolic-after-Neural
e.g., structured reasoning over
natural sensor inputs

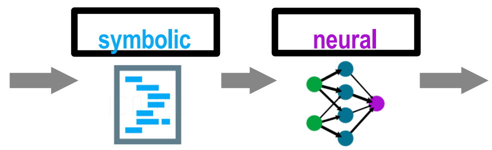


Neural-after-Symbolic
e.g., deep learning over
pre-processed inputs

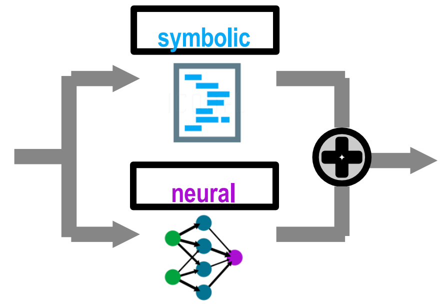
A Rich Neurosymbolic Landscape



Symbolic-after-Neural
e.g., structured reasoning over natural sensor inputs

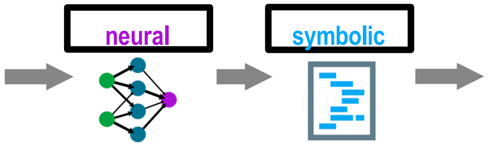


Neural-after-Symbolic
e.g., deep learning over pre-processed inputs

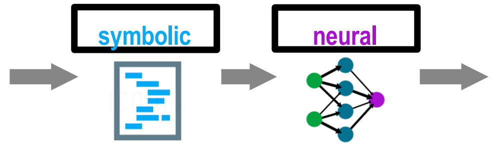


Aggregate / Fuse
e.g., DNN models errors in symbolic, symbolic polices DNN

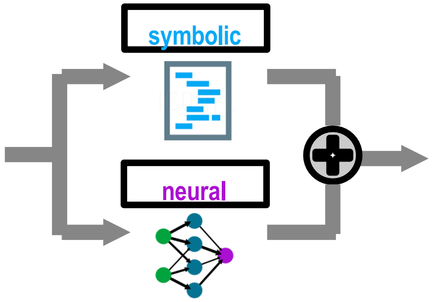
A Rich Neurosymbolic Landscape



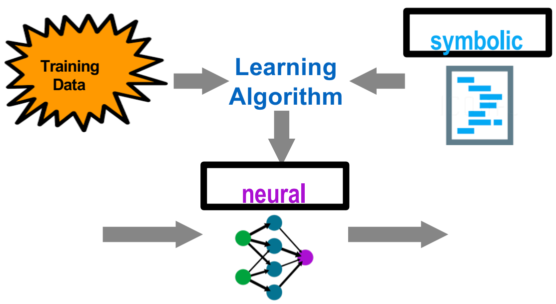
Symbolic-after-Neural
e.g., structured reasoning over natural sensor inputs



Neural-after-Symbolic
e.g., deep learning over pre-processed inputs

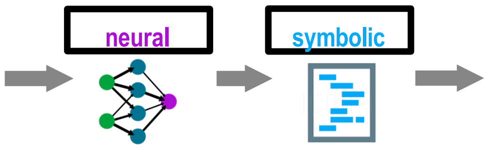


Aggregate / Fuse
e.g., DNN models errors in symbolic, symbolic polices DNN

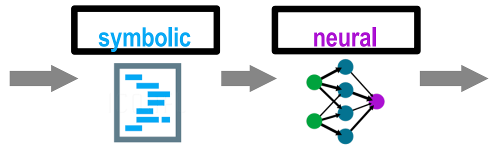


Symbolically-constrained Neural
e.g., DNN trained to follow constraints, norms and rules

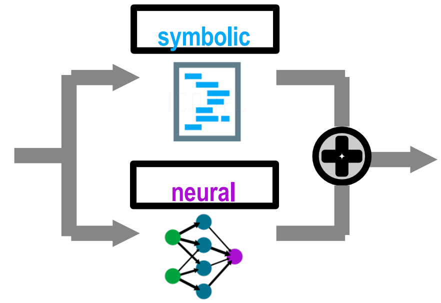
A Rich Neurosymbolic Landscape



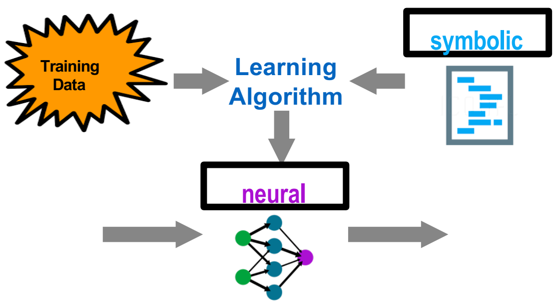
Symbolic-after-Neural
e.g., structured reasoning over natural sensor inputs



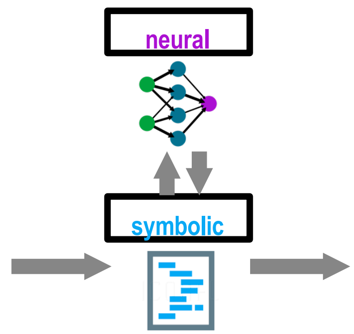
Neural-after-Symbolic
e.g., deep learning over pre-processed inputs



Aggregate / Fuse
e.g., DNN models errors in symbolic, symbolic polices DNN

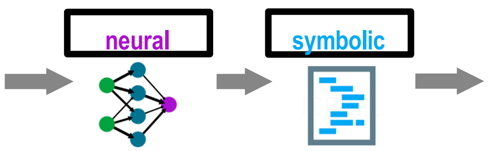


Symbolically-constrained Neural
e.g., DNN trained to follow constraints, norms and rules

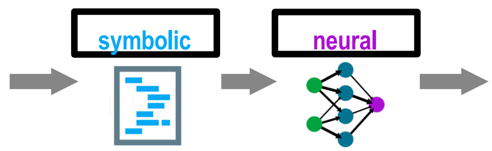


Neurally-accelerated Symbolic
e.g., neural network models errors in symbolic model

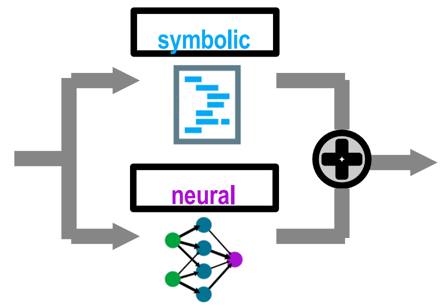
A Rich Neurosymbolic Landscape



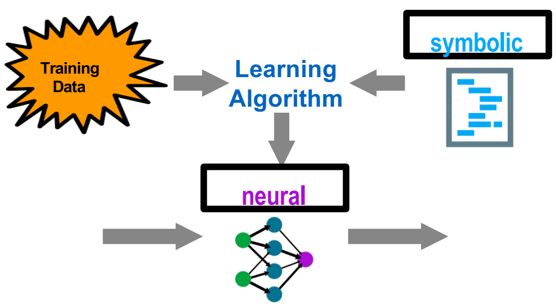
Symbolic-after-Neural
e.g., structured reasoning over natural sensor inputs



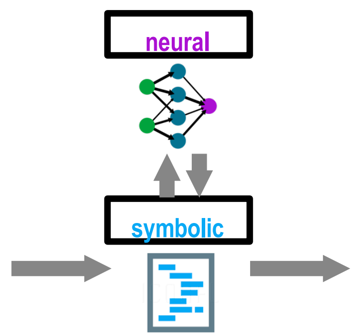
Neural-after-Symbolic
e.g., deep learning over pre-processed inputs



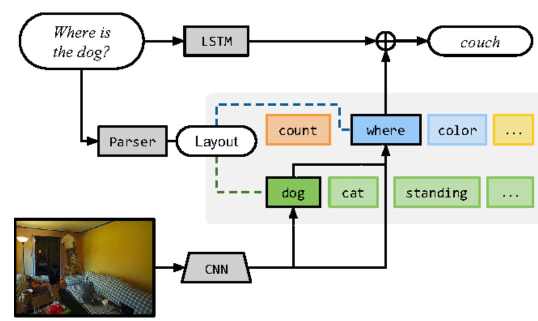
Aggregate / Fuse
e.g., DNN models errors in symbolic, symbolic polices DNN



Symbolically-constrained Neural
e.g., DNN trained to follow constraints, norms and rules



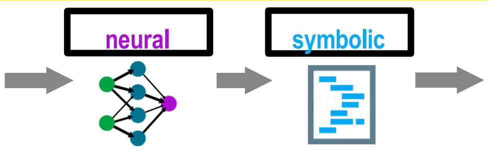
Neurally-accelerated Symbolic
e.g., neural network models errors in symbolic model



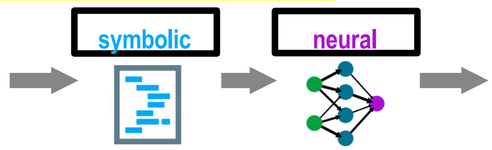
Neural Module Networks
e.g., dynamically synthesized compositions of modular neural networks

A Rich Neurosymbolic Landscape

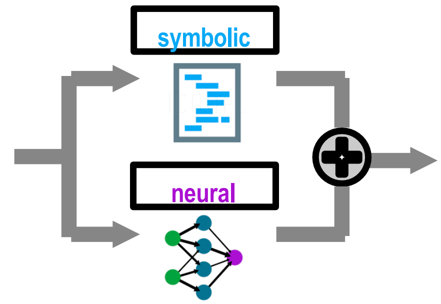
Recommended Reading:
Neurosymbolic Programming by Chaudhuri et al



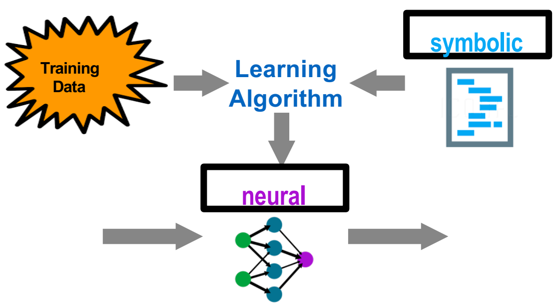
Symbolic-after-Neural
e.g., structured reasoning over natural sensor inputs



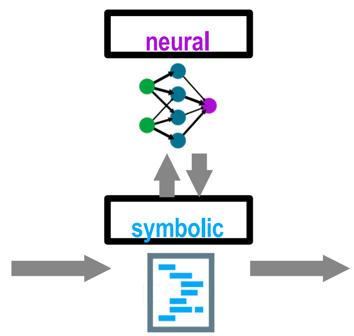
Neural-after-Symbolic
e.g., deep learning over pre-processed inputs



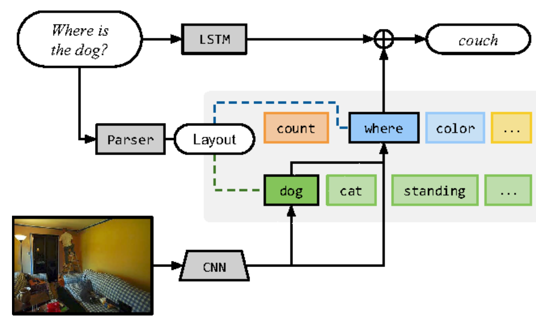
Aggregate / Fuse
e.g., DNN models errors in symbolic, symbolic polices DNN



Symbolically-constrained Neural
e.g., DNN trained to follow constraints, norms and rules

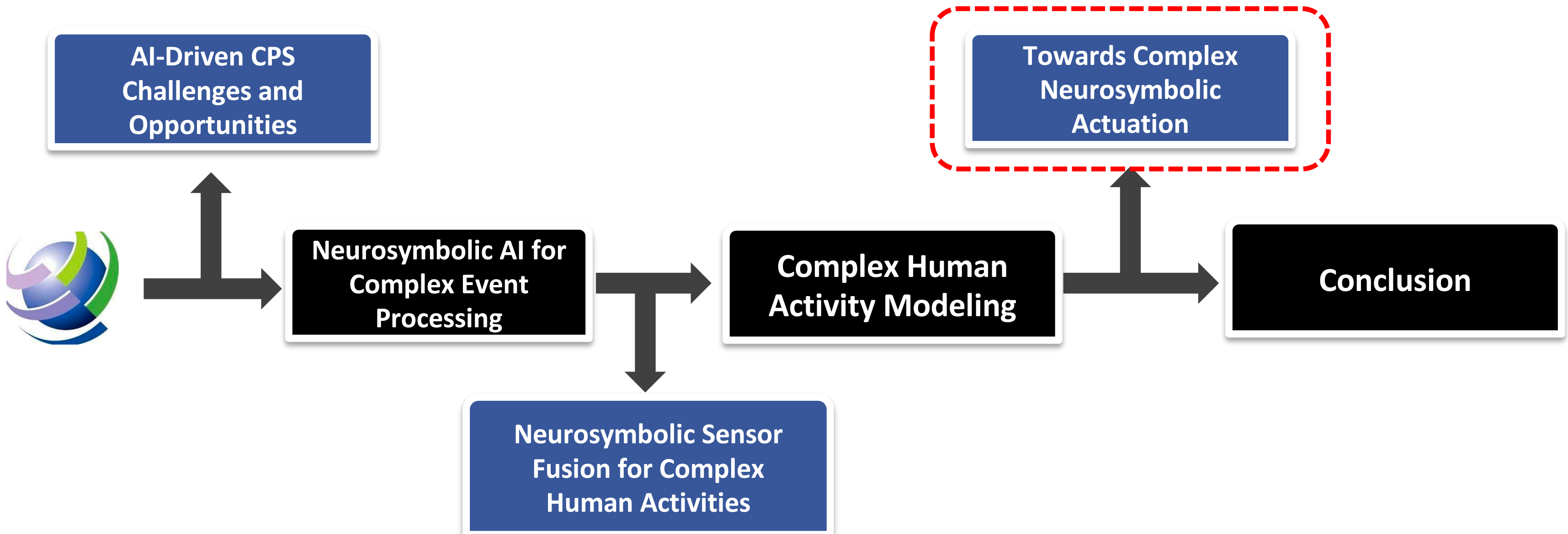


Neurally-accelerated Symbolic
e.g., neural network models errors in symbolic model



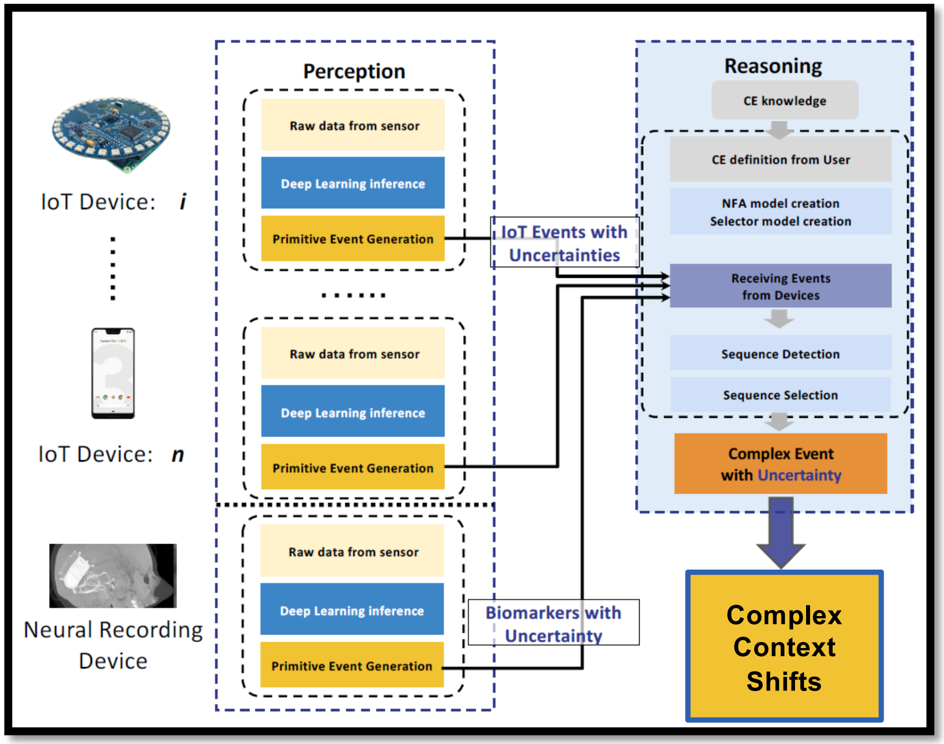
Neural Module Networks
e.g., dynamically synthesized compositions of modular neural networks

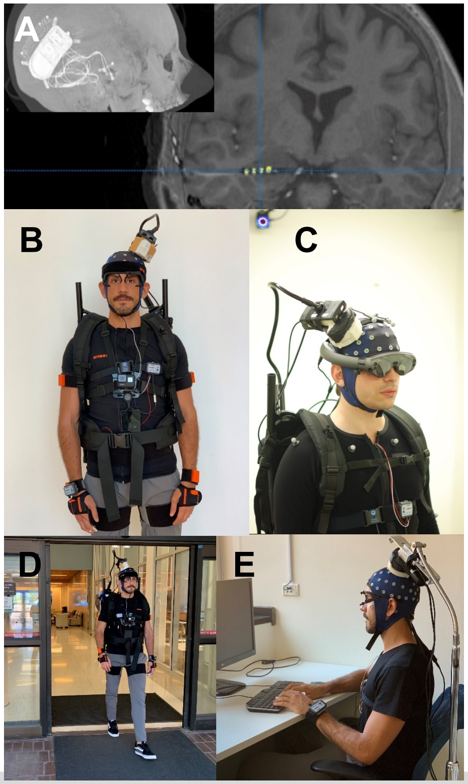
Outline for Today's Talk

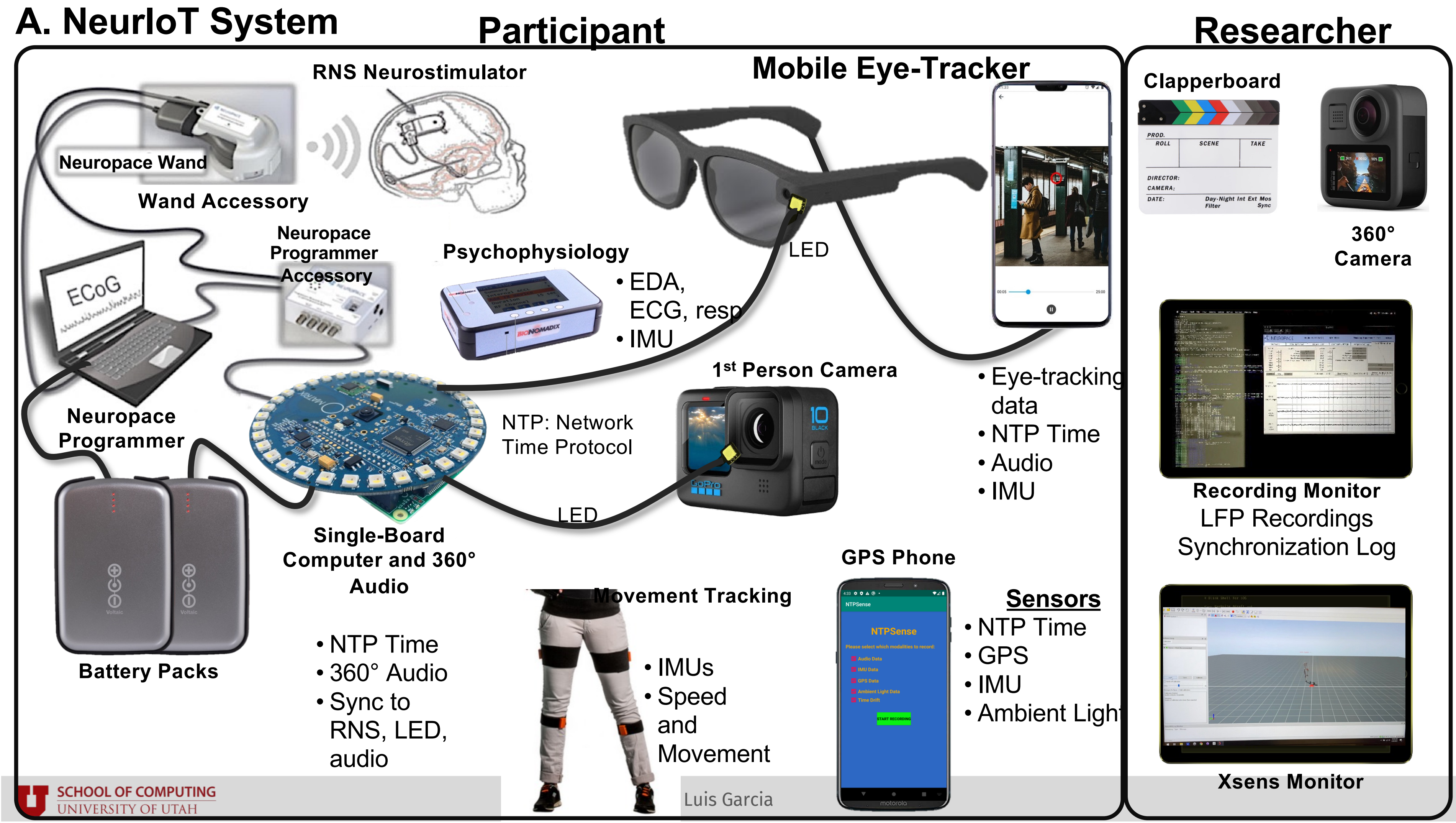


Back to the Neural Frontier: Recording and stimulation in the wild

IoT-in-the-loop Neuroscience

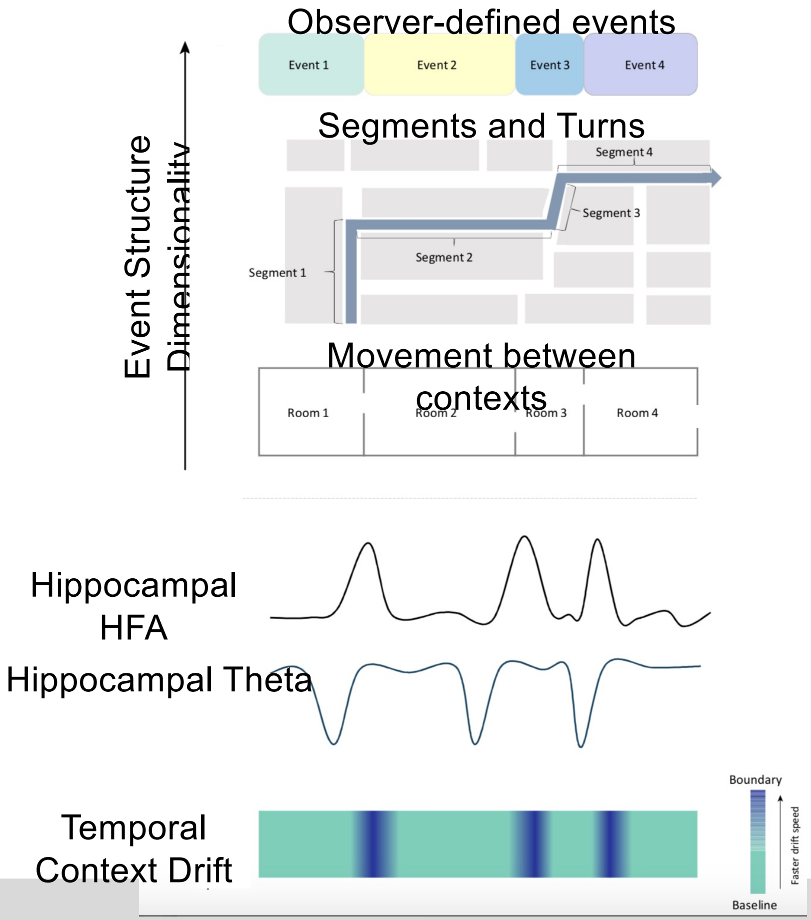





A. NeurIoT System
Participant
Researcher
RNS Neurostimulator
Mobile Eye-Tracker
Neuropace Wand
Wand Accessory
Neuropace Programmer Accessory
Psychophysiology
• EDA, ECG, resp
• IMU
LED
ECoG
Neuropace Programmer
1st Person Camera
NTP: Network Time Protocol
• Eye-tracking data
• NTP Time
• Audio
• IMU
LED
Single-Board Computer and 360° Audio
• NTP Time
• 360° Audio
• Sync to RNS, LED, audio
Battery Packs
Movement Tracking
• IMUs
• Speed and Movement
GPS Phone
NTPSense
Sensors
• NTP Time
• GPS
• IMU
• Ambient Light
Clapperboard
360° Camera
Recording Monitor
LFP Recordings
Synchronization Log
Xsens Monitor
SCHOOL OF COMPUTING
UNIVERSITY OF UTAH
Luis Garcia

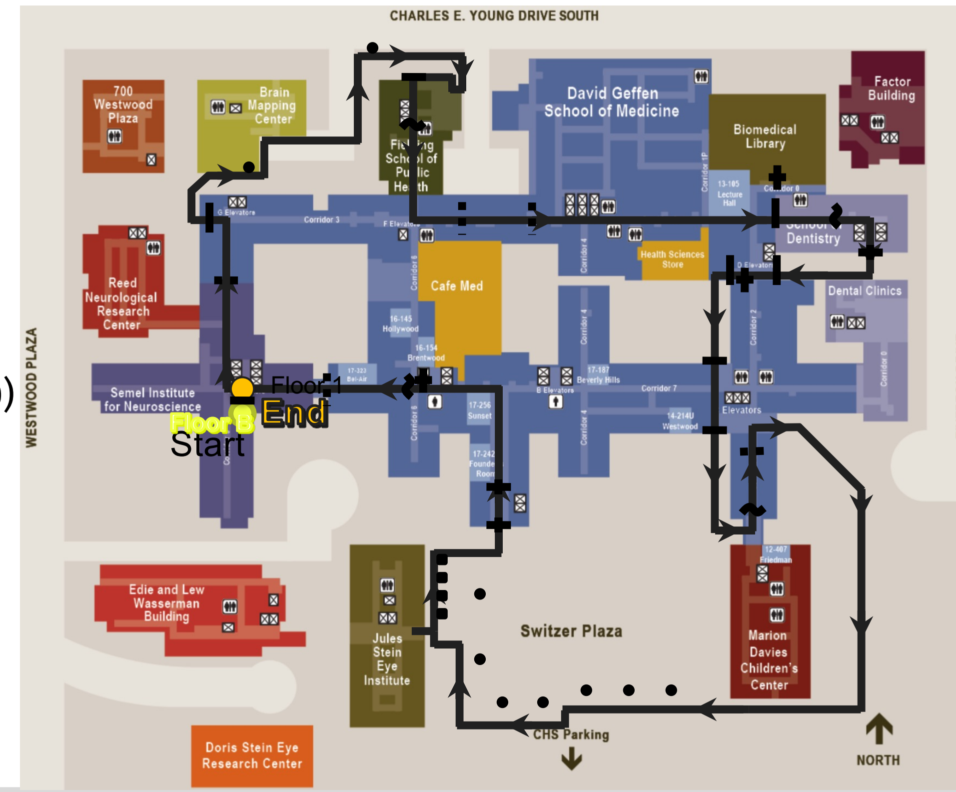
Initial Goal: Decode How Humans Encode Memories

- "Episodic Memory" model
  - Memory traces are linked by representation of context
  - Drifts slowly over time
  - Reflected in hippocampal activity
- Construct navigational tasks that will have major experiential "context shifts"
  - Inside versus outside
  - Passing through doorways
  - Encountering prominent landmarks



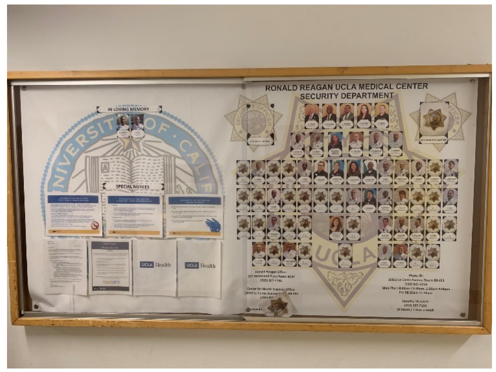
Route Characteristics

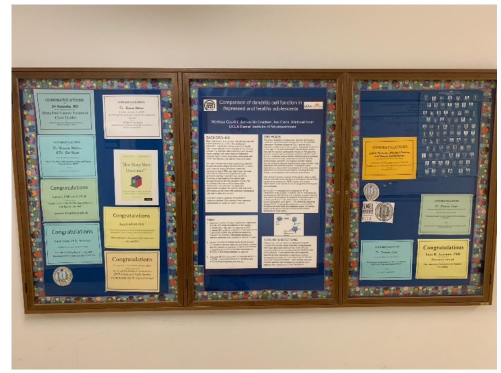
- UCLA Center for Health Sciences
- Spatial boundaries:
  - Doorways (17)
    - Closed Doorways (14)
    - Open Doorways (3)
    - Indoors/Outdoors (11)
  - Turns (25)
  - Transitions between buildings (10)
- Duration = 17 – 25 min
- Distance = ~0.75 miles
- 8 Walks (4 per day)
  - 1 Encoding
  - 7 Navigation



Landmarks


UCLA Fielding School
of Public Health

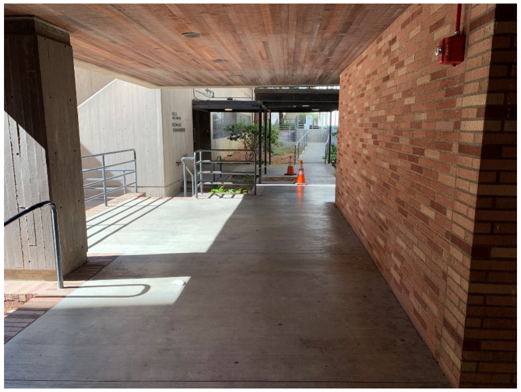

RONALD REAGAN UCLA MEDICAL CENTER
SECURITY DEPARTMENT




EXIT

Scenes

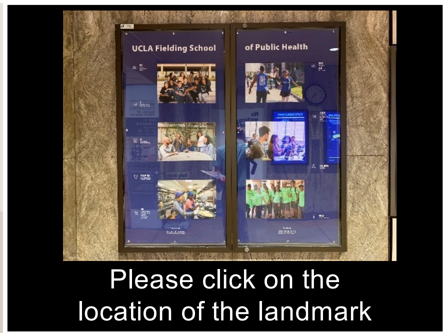
50 “segments” identified



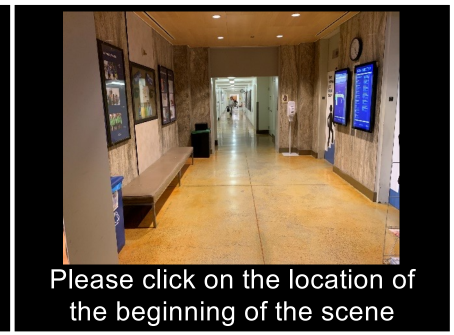
Landmark Recognition Tasks

Map Drawing Task

Landmark Placement Task

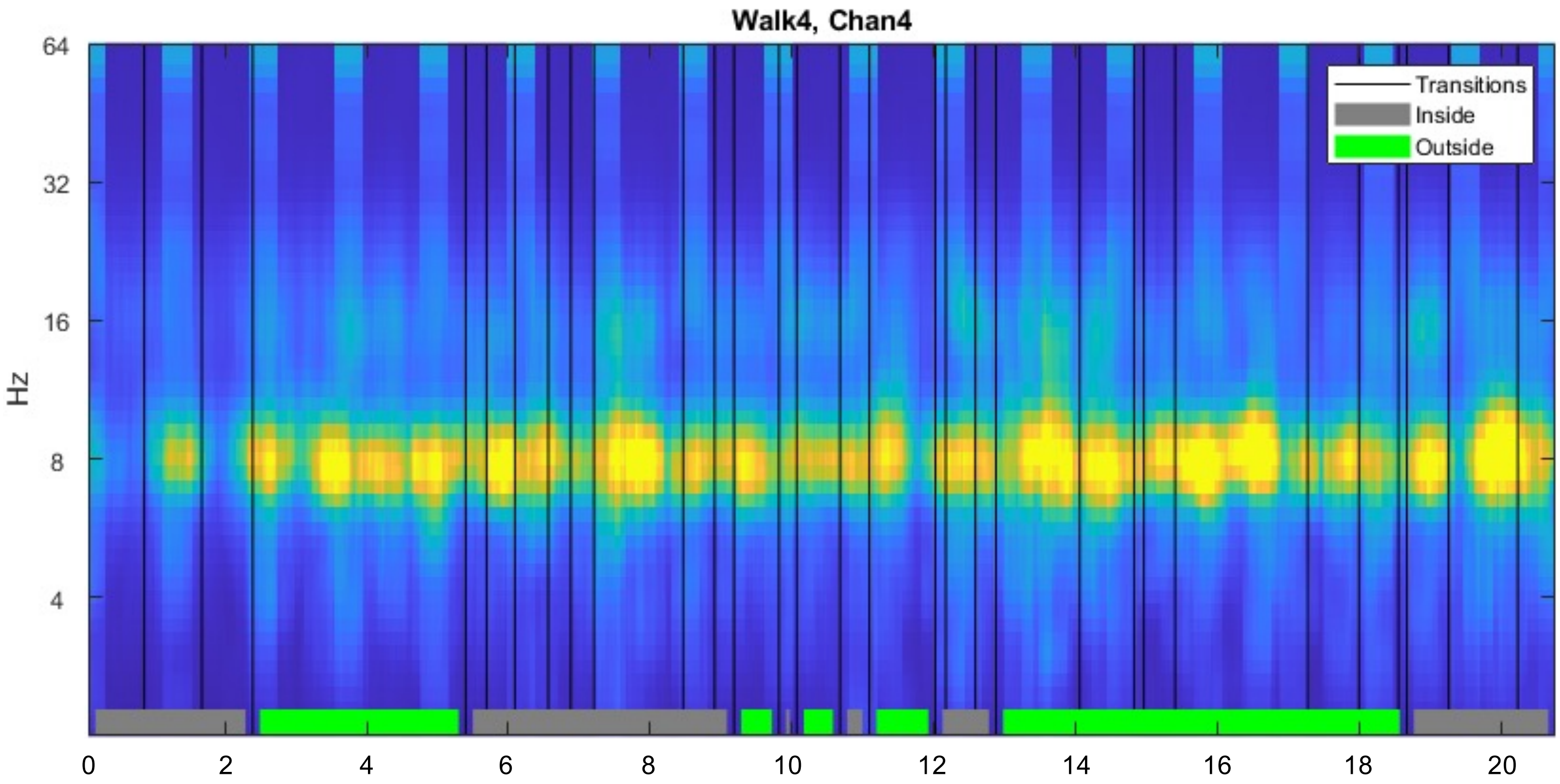


Scene Placement Task

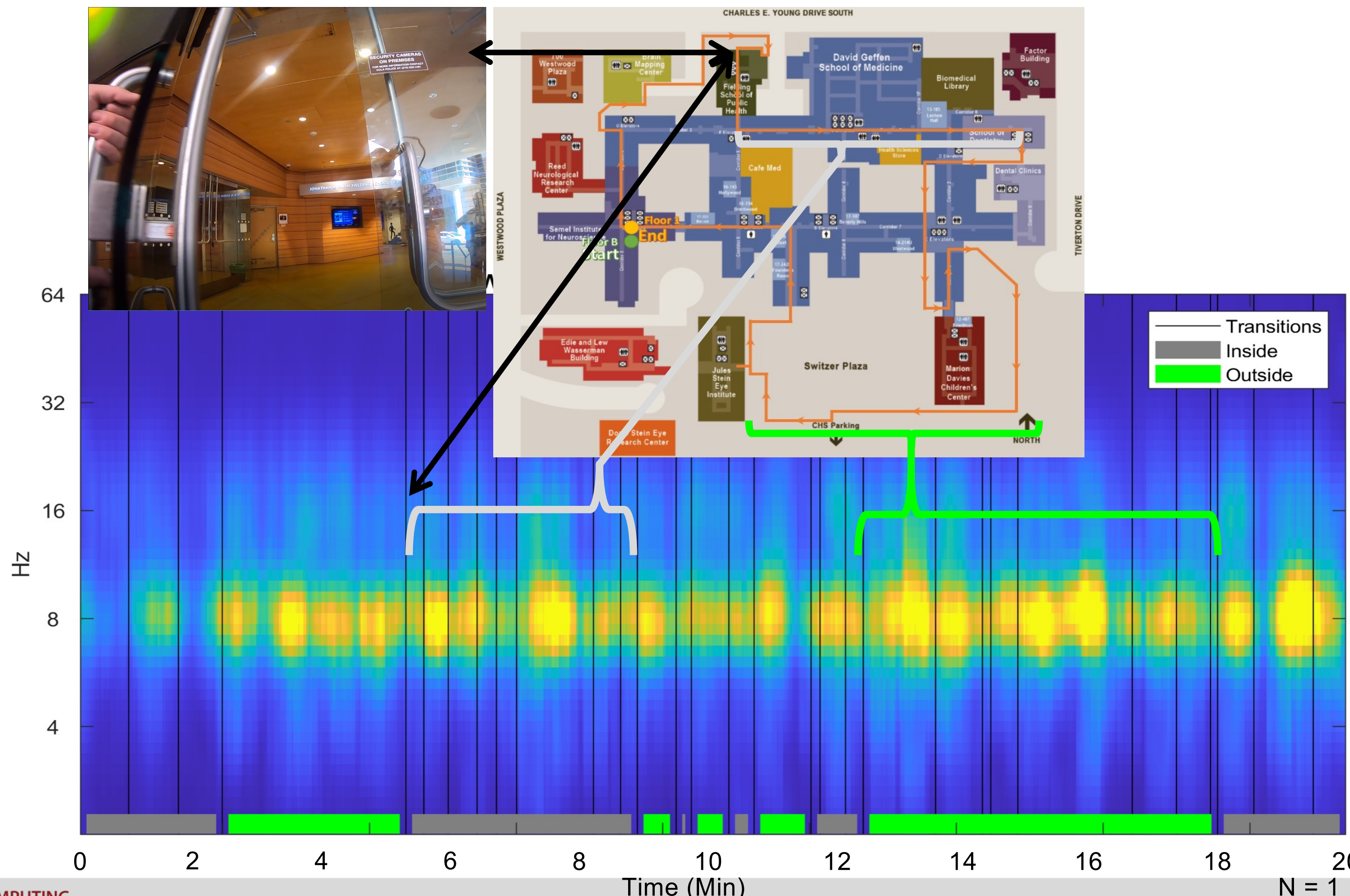


- Patient will draw route on map after the last walk

Hippocampal theta activity during real-world spatial navigation



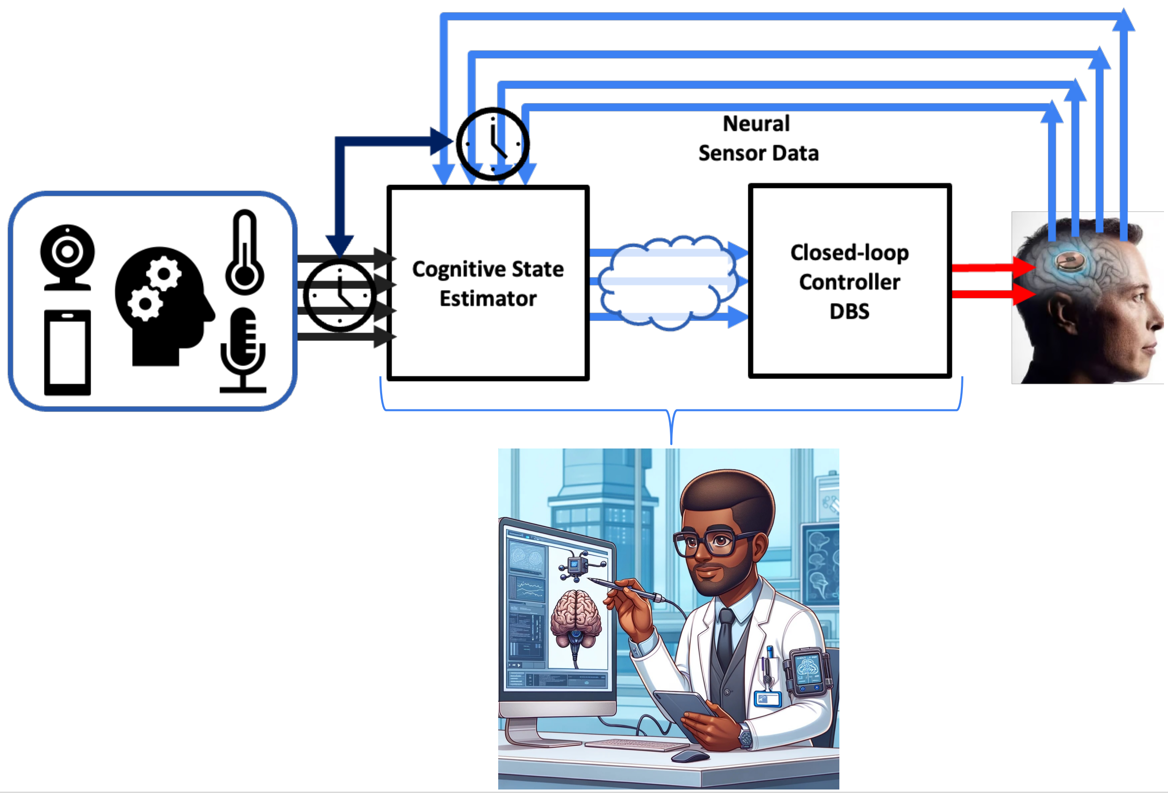
N = 1



CHARLES E. YOUNG DRIVE SOUTH
David Geffen School of Medicine
Biomedical Library
Factor Building
Reed Neurological Research Center
Cafe Med
Floor 1
End
Start
WESTWOOD PLAZA
TIVERTON DRIVE
Edie and Lew Wasserman Building
Jules Stein Eye Institute
Switzer Plaza
Marion Davies Children's Center
CHS Parking
NORTH
Transitions
Inside
Outside
64
32
16
8
4
Hz
0
2
4
6
8
10
12
14
16
18
20
Time (Min)
N = 1

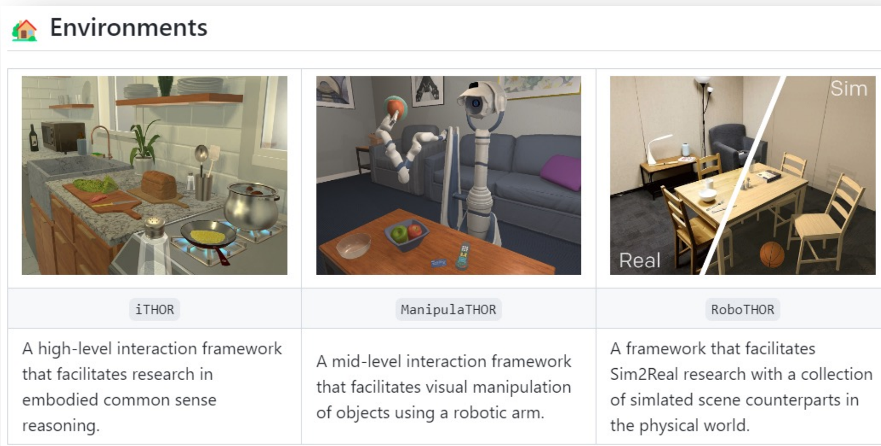
But will more robust neurosymbolic perception enable safe actuation with blurry requirements?

Other practical challenges:
- Limited Data
- Resource constraints
- Privacy + Security concerns
- Patient-centered design

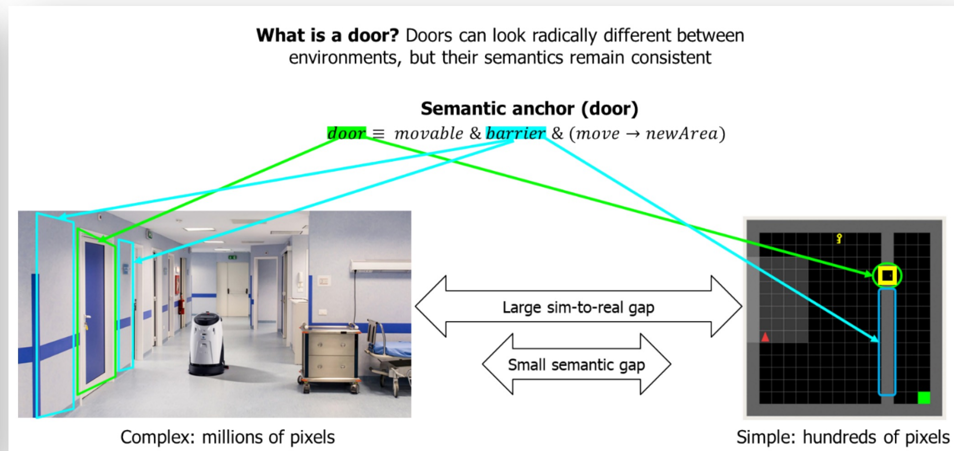


Some preliminary exploration:
Robustifying Neurosymbolic Perception Models in Simulation

Can we leverage cross-domain simulators or datasets for more robust perception?



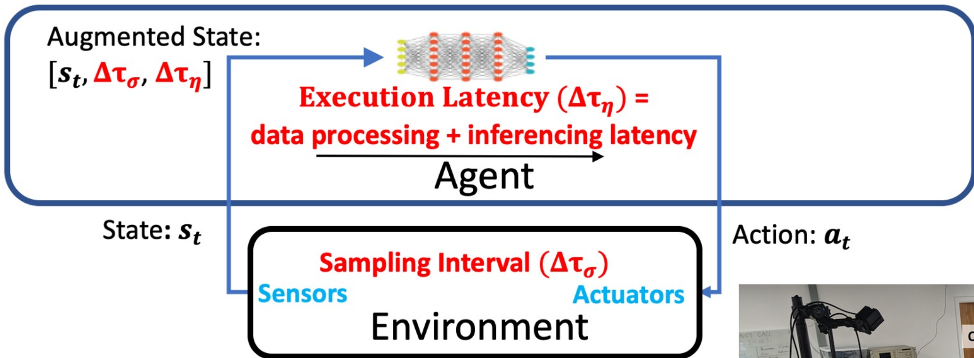
Emergent Embodied AI Simulators

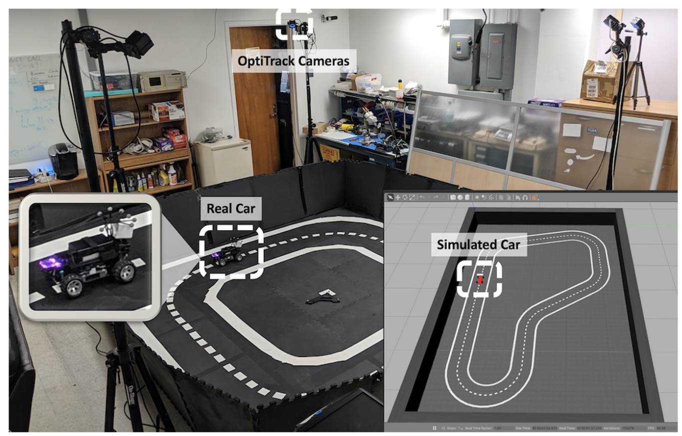


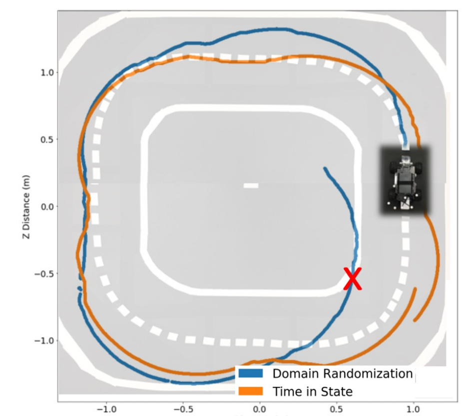
From DARPA's Transfer from Imprecise and Abstract Models to Autonomous Technologies (TIAMAT)

Some preliminary exploration:
Robustifying Neurosymbolic Perception Models in Simulation

Introducing consistently measurable symbols in state enhances Sim2Real Transfer





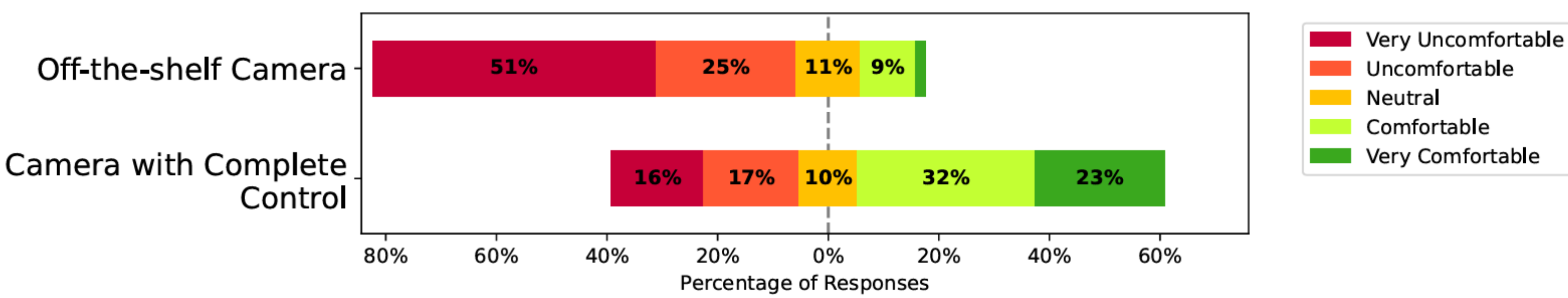


Sandha, Sandeep Singh, et al. "Sim2real transfer for deep reinforcement learning with stochastic state transition delays." *CoRL '21*

Some preliminary exploration:
Managing Requirement Specifications

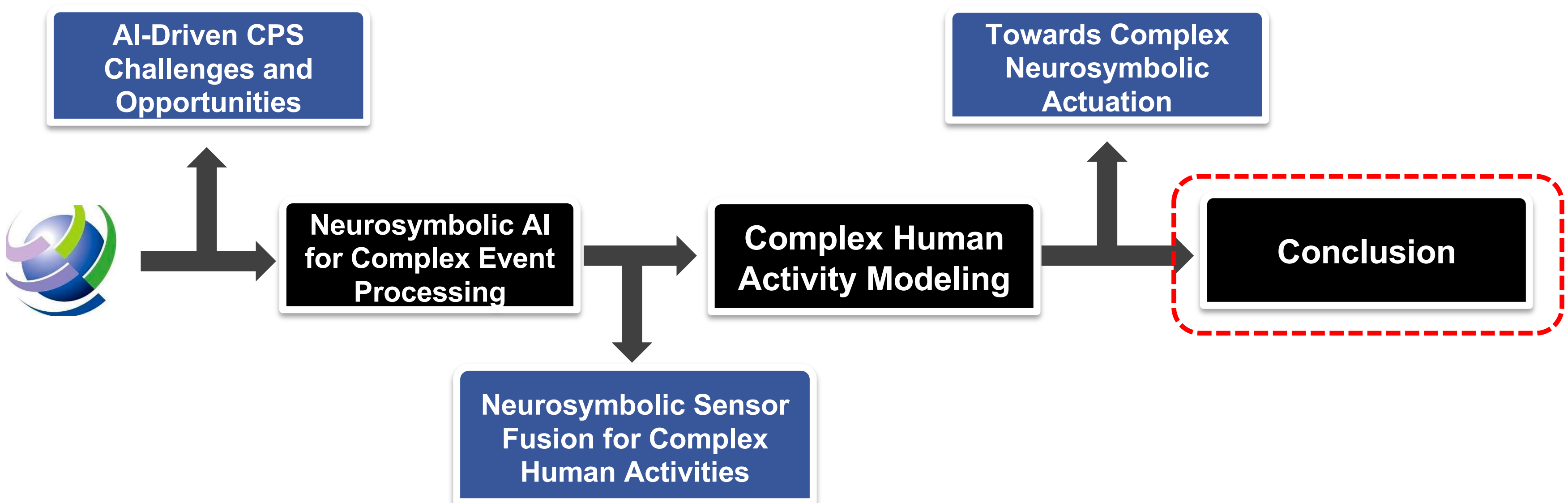
Even if model is explainable, interfaces still require cross-domain expertise for safety, security, and privacy

User study question: Would you be willing to put a device in your bedroom if
(a) it was an off-the-shelf camera?
(b) You had complete control over the camera's software/hardware?



Singh, Akash Deep, Brian Wang, Luis Garcia, Xiang Chen, and Mani Srivastava. "Understanding factors behind IoT privacy--A user's perspective on RF sensors." *arXiv preprint arXiv:2401.08037* (2024).

Outline for Today's Talk



Concluding Thoughts

- **Neurosymbolic models** can at least bridge the gap for limitations in DNN-only or symbolic-only sensor fusion models for perception

- We need better mechanisms to bootstrap **semantic grounding** at different symbolic layers across sensing modalities
  - Fusion at symbolic layers: Label space, semantic loss, concept bottlenecks, etc.
  - Better semantic oracles: existing knowledge graphs and LLMs have shown to be useful

- Better mechanisms for **interfacing both domain experts and end-users** with neurosymbolic models (maybe LLMs?)

- We need to take a holistic approach to **closing-the-loop** when modeling neurosymbolic safety-critical applications

Luis Garcia
la.garcia@utah.edu
https://lagarcia.us

Thank You!

Luis Garcia
la.garcia@utah.edu
https://lagarcia.us