Safe & Intelligent Brain-Robot Interfaces:
Bridging One-Shot Learning with Runtime Verification

Tasha Kim
Oxford Robotics Institute (ORI), Department of Engineering Science

Attributions

This presentation presents and synthesizes results from:

- NOIR 2.0: Neural Signal Operated Intelligent Robots for Everyday Activities (Kim, Wang, Cho, Hodges, 2024)

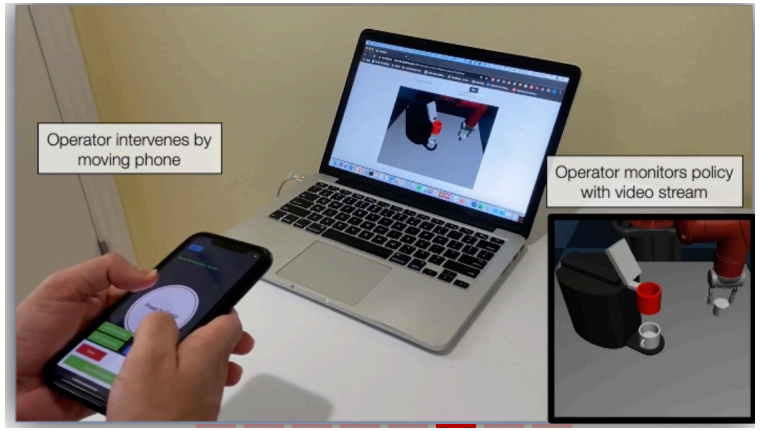
- EEG-Based Brain-Computer Interface for Robotic Assistance with User Intention Prediction (Zhang*, Kim*, Wang*, Cho, Hodges, Tan, Wang, Hwang, Lee, Hiranaka, Ai, Norcia, Fei-Fei, Wu, Under Review 2025)

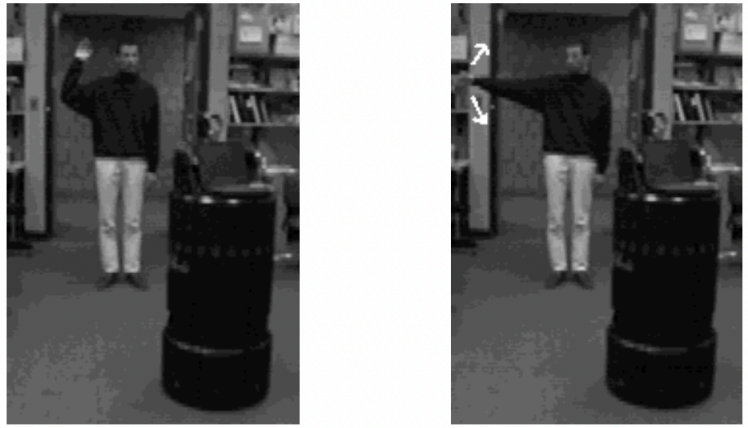
- Gated Uncertainty-Aware Runtime Dual Invariants for Neural Signal-Controlled Robotics (Kim, Parker Jones, 2025)

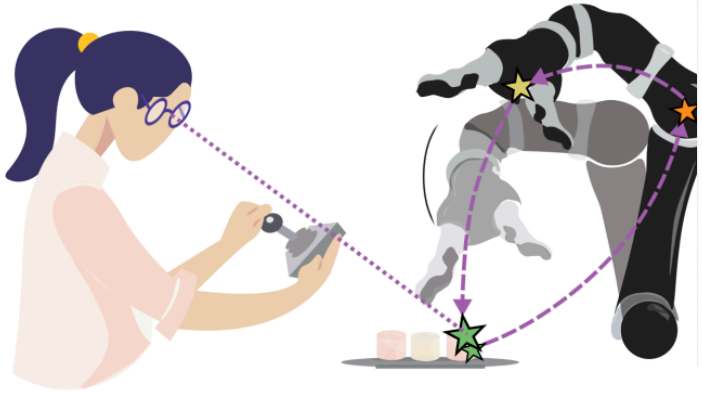
- GUARDIAN: Gated Uncertainty-Aware Runtime Dual Invariants for EEG-Controlled Agents (Kim, Parker Jones, 2026)

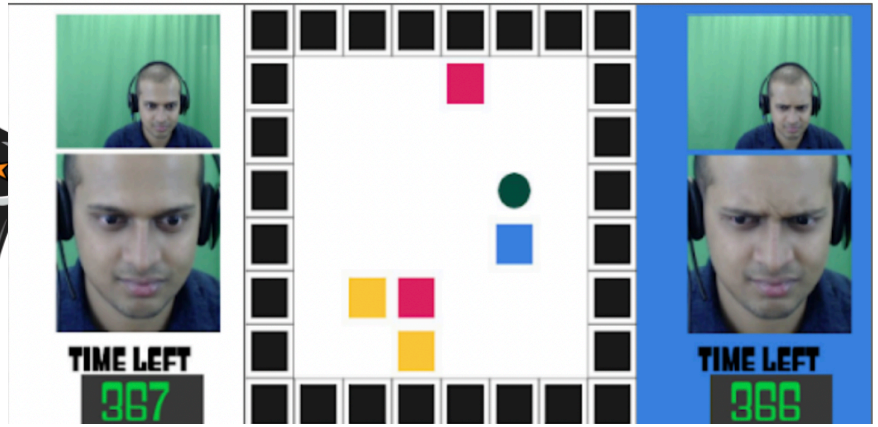
The ways humans communicate with robots

Teleoperation devices
Gesture
Gaze
Facial expression
Language
Brain signals?

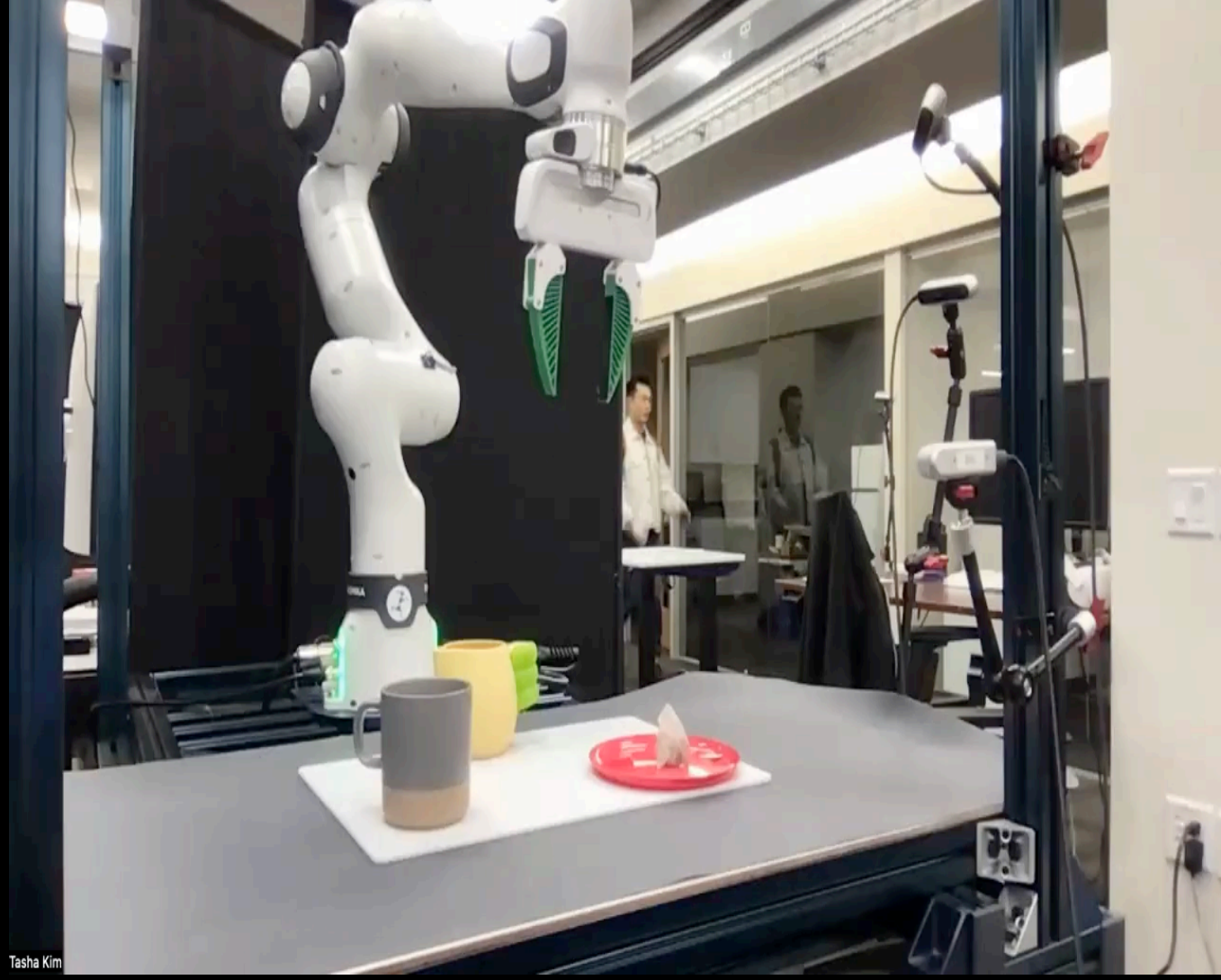








Mandlekar et al., 2018; Aronson et al., 2021; Cui et al., 2021; Waldherr et al., 2000



Using brain signals to control a robot to make tea (4x)

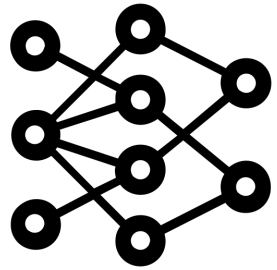
*Brain decoding wait period is omitted

Neural Signal Operated Intelligent Robots

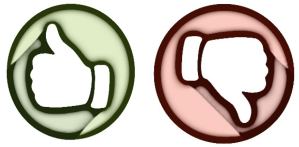
Participant's brain signals are recorded while they watch the robot

Neural Signal Operated Intelligent Robots

Participant's brain signals are recorded while they watch the robot



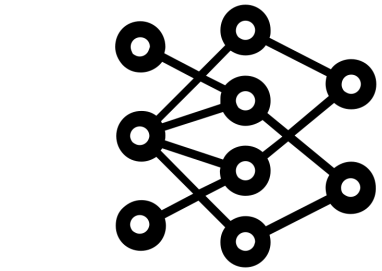
"Pick up the bottle"



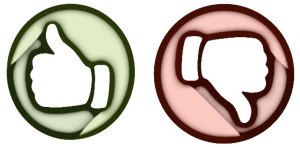
Machine learning algorithms infer human intention and evaluation

Neural Signal Operated Intelligent Robots

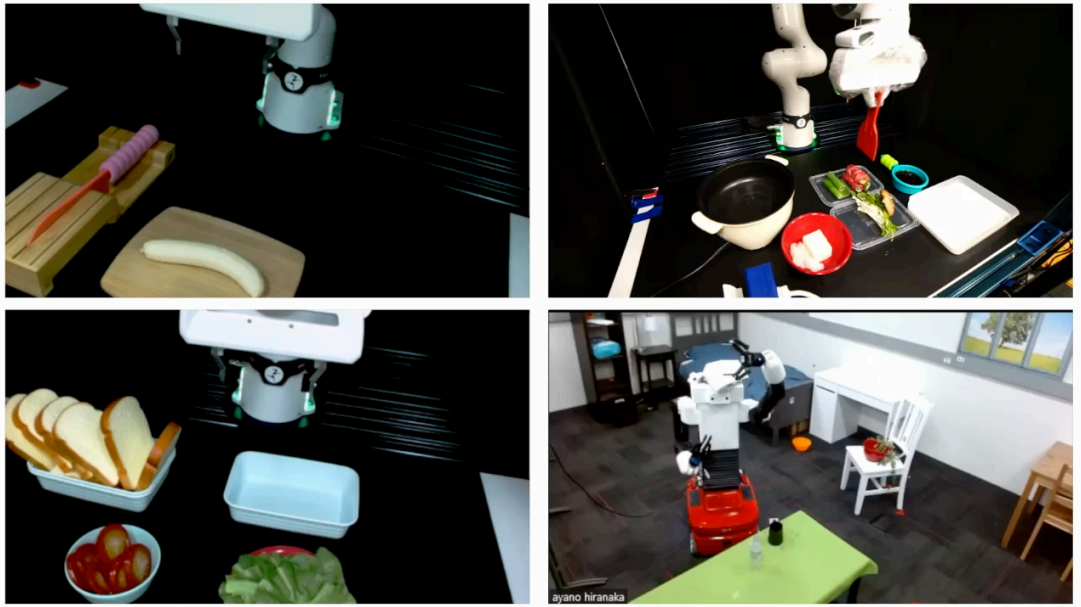
Participant's brain signals are recorded while they watch the robot



"Pick up the bottle"



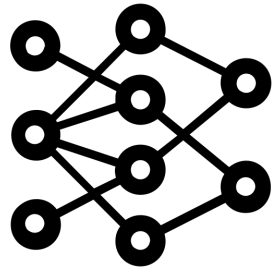
Machine learning algorithms infer human intention and evaluation



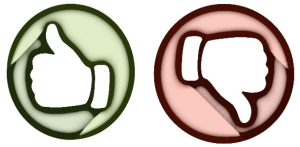
Intelligent robots with basic visuomotor skills learns to accomplish human goals

What's unique about this BRI generation?

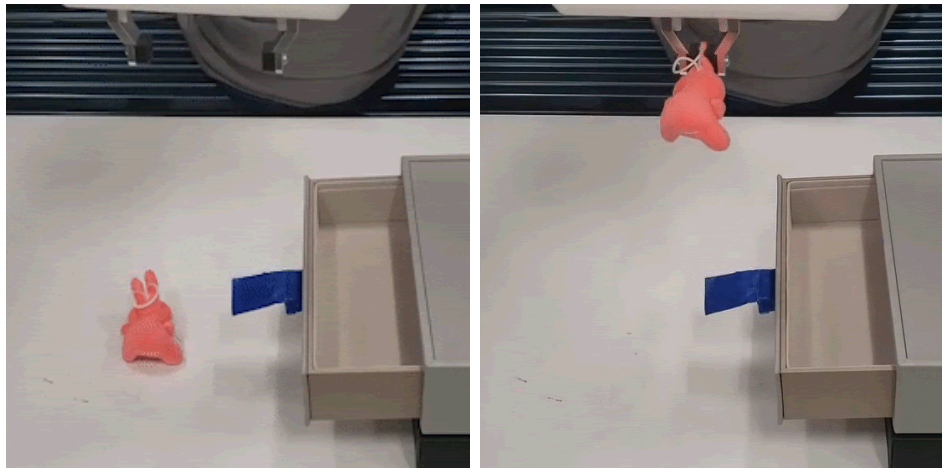
Participant's brain signals are recorded while they watch the robot



"Pick up the bottle"



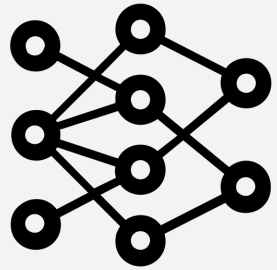
Machine learning algorithms infer human intention and evaluation



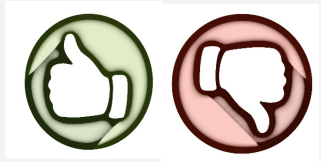
Intelligent robots with basic visuomotor skills learns to accomplish human goals

What's unique about this BRI generation?

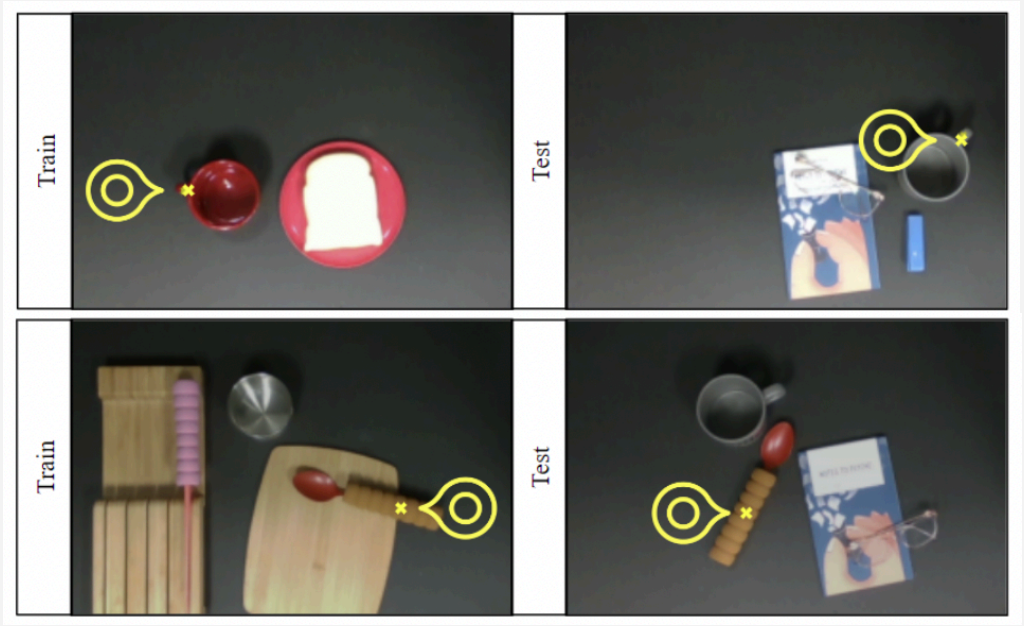
Participant's brain signals are recorded while they watch the robot



"Pick up the bottle"



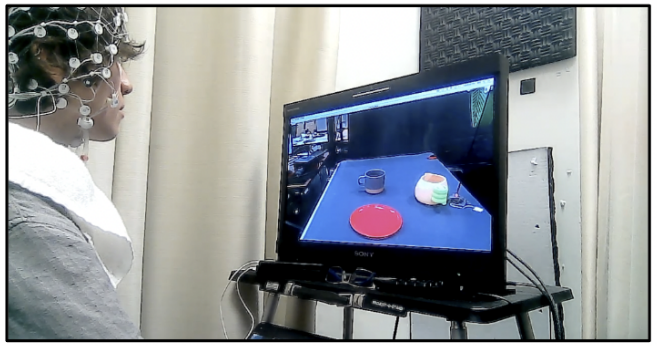
Machine learning algorithms infer human intention and evaluation



Intelligent robots with basic visuomotor skills learns to accomplish human goals

How do we decode intent from the human brain?

Environment Display & EEG Recordings



Human goal decoding

What object? **Bottle**

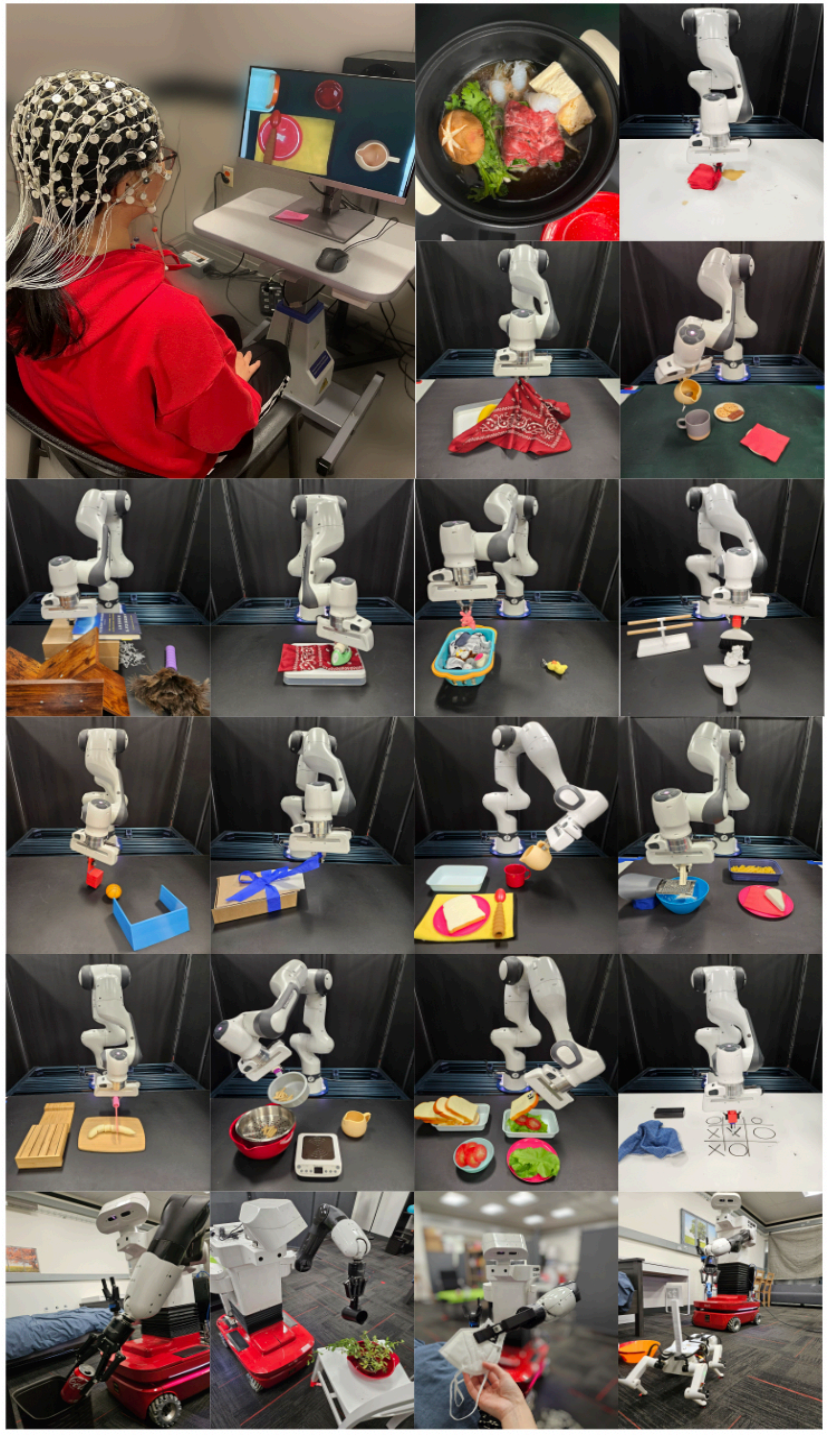
How to interact? **Pick**

Where to interact?

Robots with primitive skills
+ Human goal prediction

NOIRv1 System Performance

- Task horizon: 4-15 skills
- Average attempts to succeed: 1.8
- Average task completion time: 20.3 minutes
- Human-decision and decoding time: ~80%

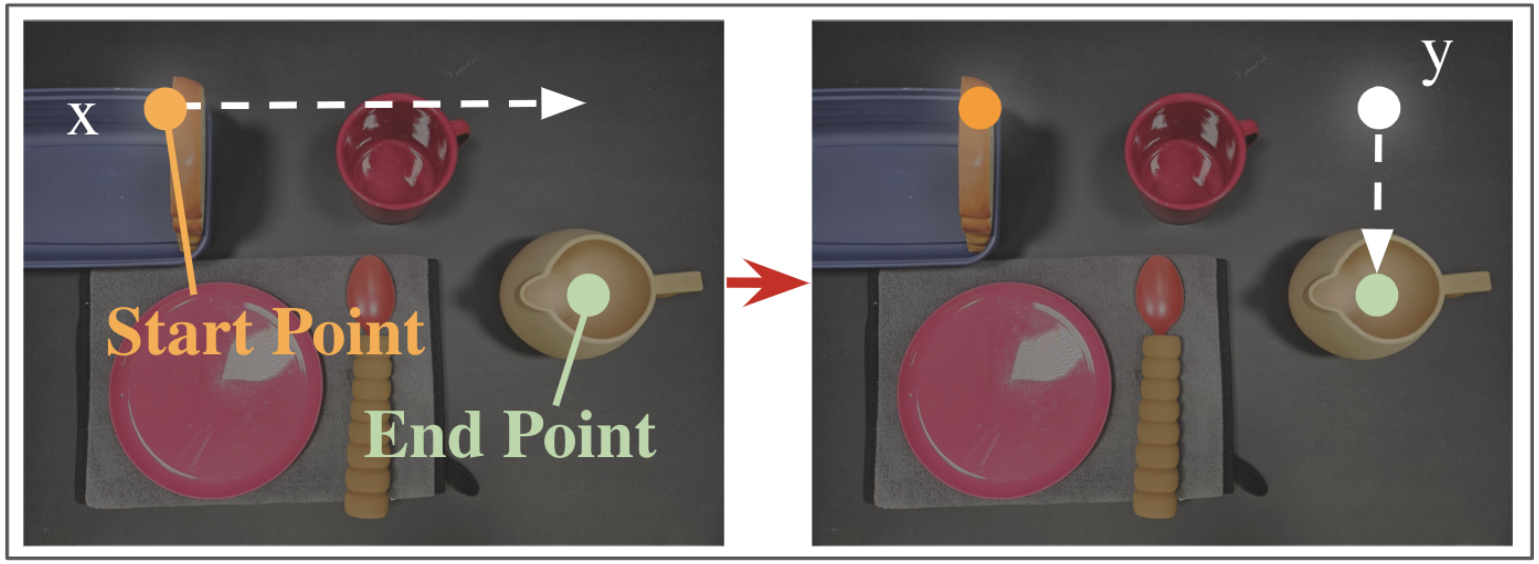


NOIRv2 System Performance

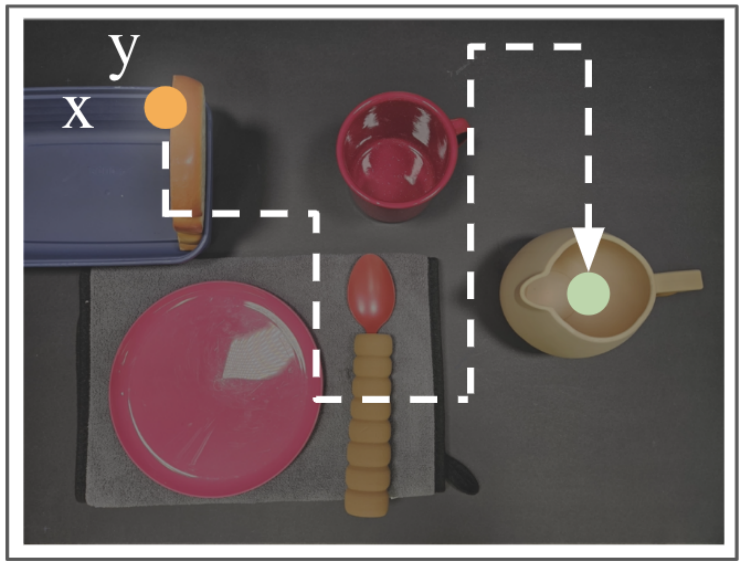
	Time (min.)			Human Time (min.)		
Task Name	NOIR	NOIR 2.0	NOIR 2.0+Learning	NOIR	NOIR 2.0	NOIR 2.0+Learning
WipeSpill	14.74	9.12	5.46	11.65	5.12	3.15
OpenBasket	15.90	6.79	5.80	13.04	2.60	1.52
PourTea	13.53	8.90	12.60	11.25	6.55	7.87
Avg. Time Reduced (%)	-	**43.82**	**45.97**	-	**60.30**	**65.11**

NOIRv2 New Features: Brain Decoding

- Faster and more accurate object and skill decoding
  - Object selection: 81% → 88%
  - Skill selection: 42% → 61%
- Continuous cursor control for skill parameter selection



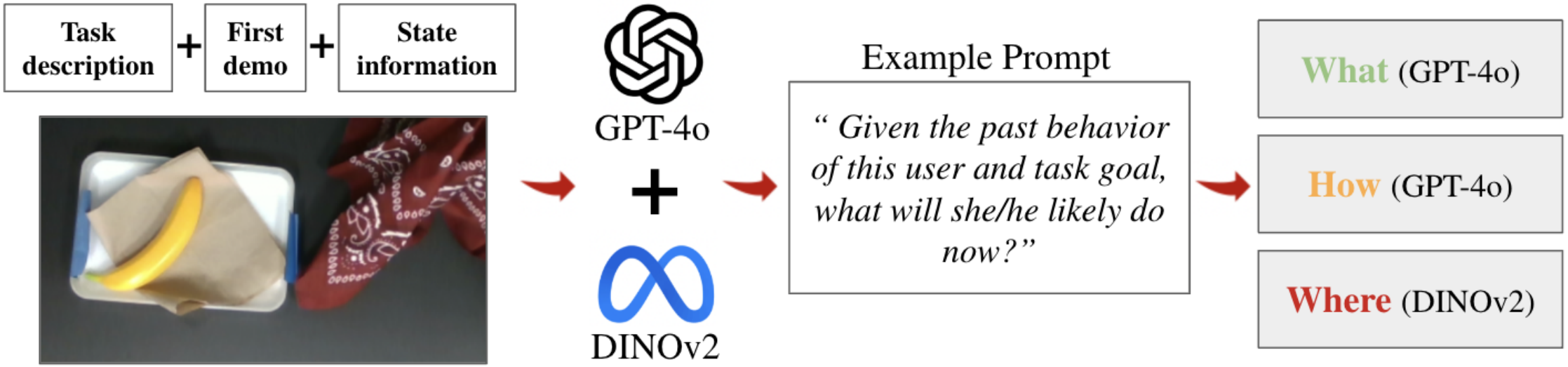
NOIRv1



NOIRv2

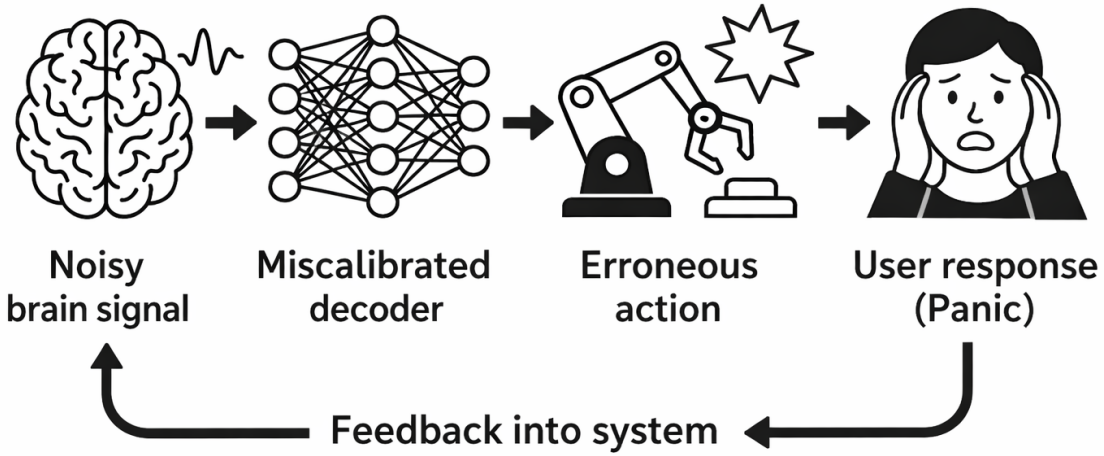
NOIRv2 New Features: Robot Learning

- NOIRv1 used few-shot imitation learning for object and skill selection (requires ~15 demos)
- NOIRv2 uses in-context learning w/ GPT-4o (requires 1 demo)



Inherent Challenges: Neural Signal Control

- Compounding feedback loop
- Verification gap
- Vulnerable user population



GUARDIAN: Runtime Safety Checking

- Physiological Invariants
  - Verify reliability of input signal before processing
- Logical Invariants
  - Verify physical validity of the intended action

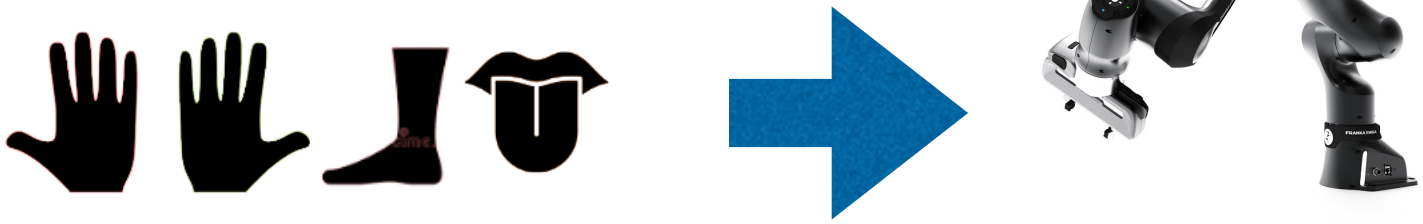
Decoder	Val. Acc.	Test Acc.	Calib. (ECE)	Safety Rate	Interventions	Latency (ms)
EEGNet	58.2%	46.0%	0.223	**94.2%**	52.3%	0.82
Riemannian	58.7%	30.0%	0.410	**95.8%**	68.1%	0.91
Light CNN	54.3%	28.0%	0.316	**96.3%**	70.4%	0.79
RealIntent	51.2%	27.0%	0.287	**97.0%**	71.2%	0.73
Mean	55.6%	32.8%	0.309	**95.8% (> 90%)**	65.5%	**0.81 (< 1 ms)**

*Safety Rate = Correct Interventions + Correct Executions / Total Trials

Discussion

- Interpretable primitives

  Action set for natural control
  Interpretable intent-to-action
  Shared autonomy support



- Adaptive safety

  Calibration-independent invariants
  Consistency preservation
  Multi-level thresholding

- Practical deployment

  Sub-millisecond overhead on any decoder
  Audit logs for compliance (regulatory framework compatibility)
  PDDL-compatible toolchain for robotic control

Thank you

Contact: tasha.kim@eng.ox.ac.uk