Mapping Brain Signals to Agent Performance, A Step Towards Reinforcement Learning from Neural Feedback

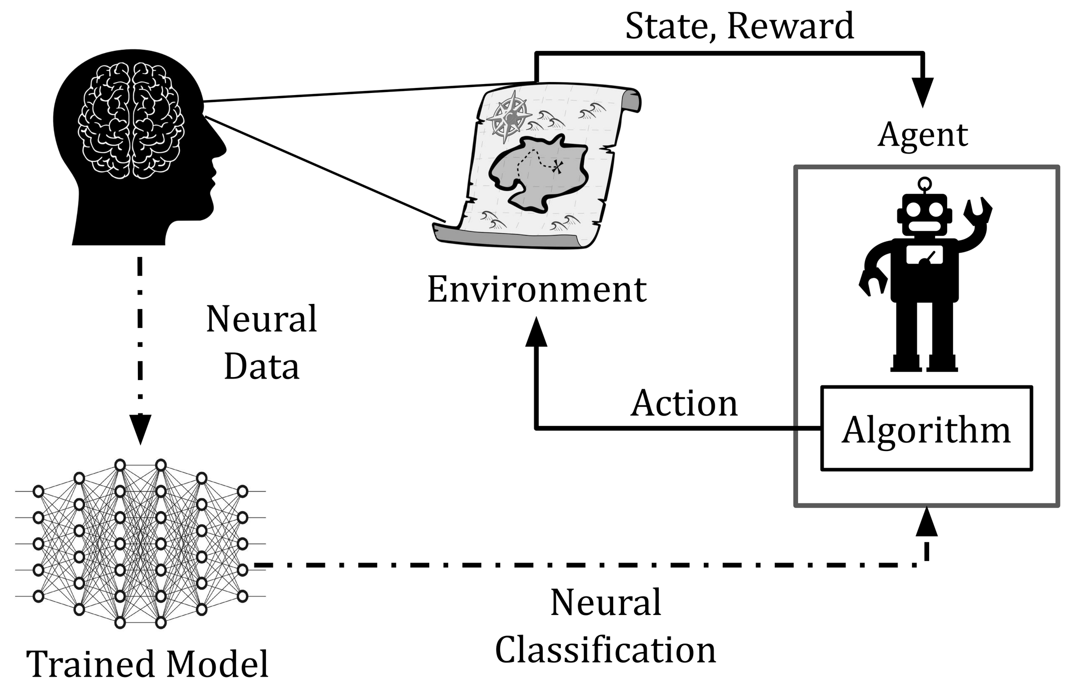
Julia Santaniello julia.santaniello@tufts.edu
Matt Russell mrussell@cs.tufts.edu
Benson Jiang benson.jiang@tufts.edu
Donatello Sassaroli donatello.sassaroli@tufts.edu
Robert Jacob jacob@cs.tufts.edu
Jivko Sinapov jivko.sinapov@tufts.edu

Introduction

- **Reinforcement Learning from Human Feedback (RLHF)** enhances agent training through using feedback from human "teachers"
- Most algorithms rely on preference or demonstration data as evaluative feedback. Feedback acquisition often **increases user workload** due to sustained attention, decision making and/or active demonstration
- **Passive Brain-Computer Interfaces (BCI)** assess user cognitive states through various neuroimaging devices
- **Functional Near-Infrared Spectroscopy (fNIRS)** is one such device that measures the change in hemodynamic response in the brain

Questions:

1. How can we communicate agent performance implicitly through neural data?
2. How much granularity can we derive with this signal?

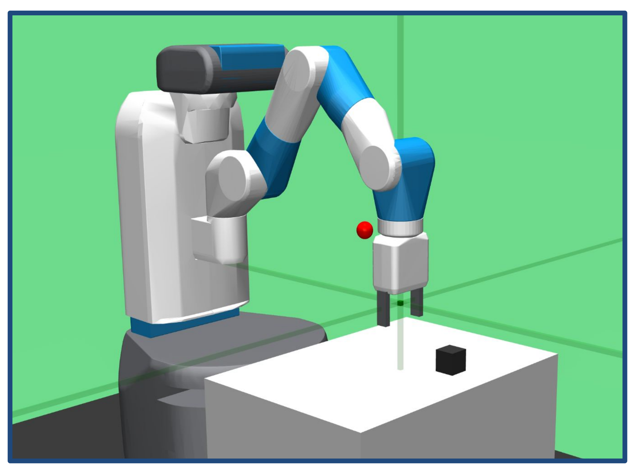


Setup

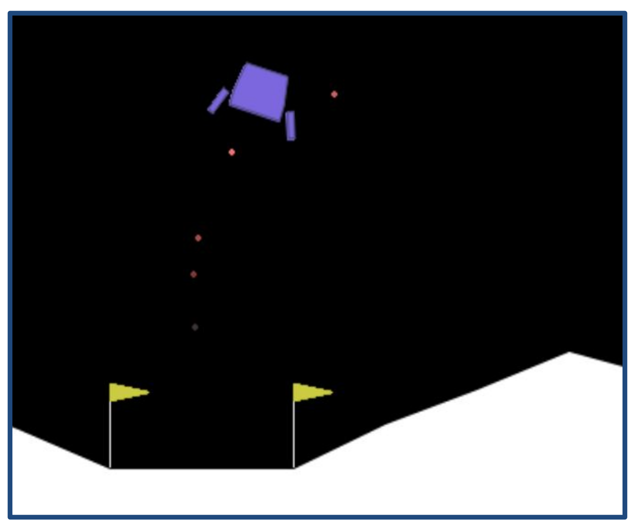
Participant Set Up: Participant sits at a computer and **observes** or **guides** an agent through a task

fNIRS Device: fNIRS is used to estimate **hemodynamic activity** in the prefrontal cortex

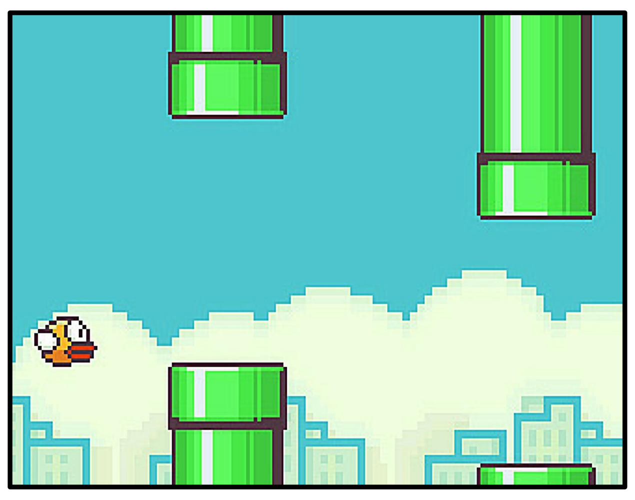
Domains



Robot Fetch and Place



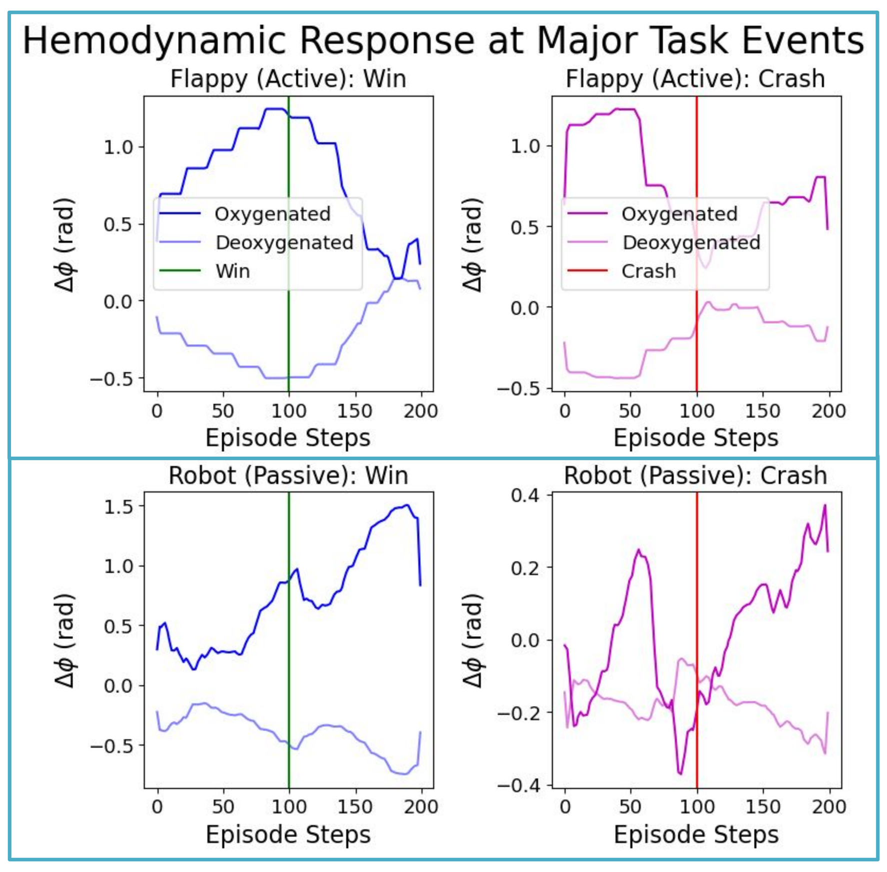
Lunar Lander



Flappy Bird

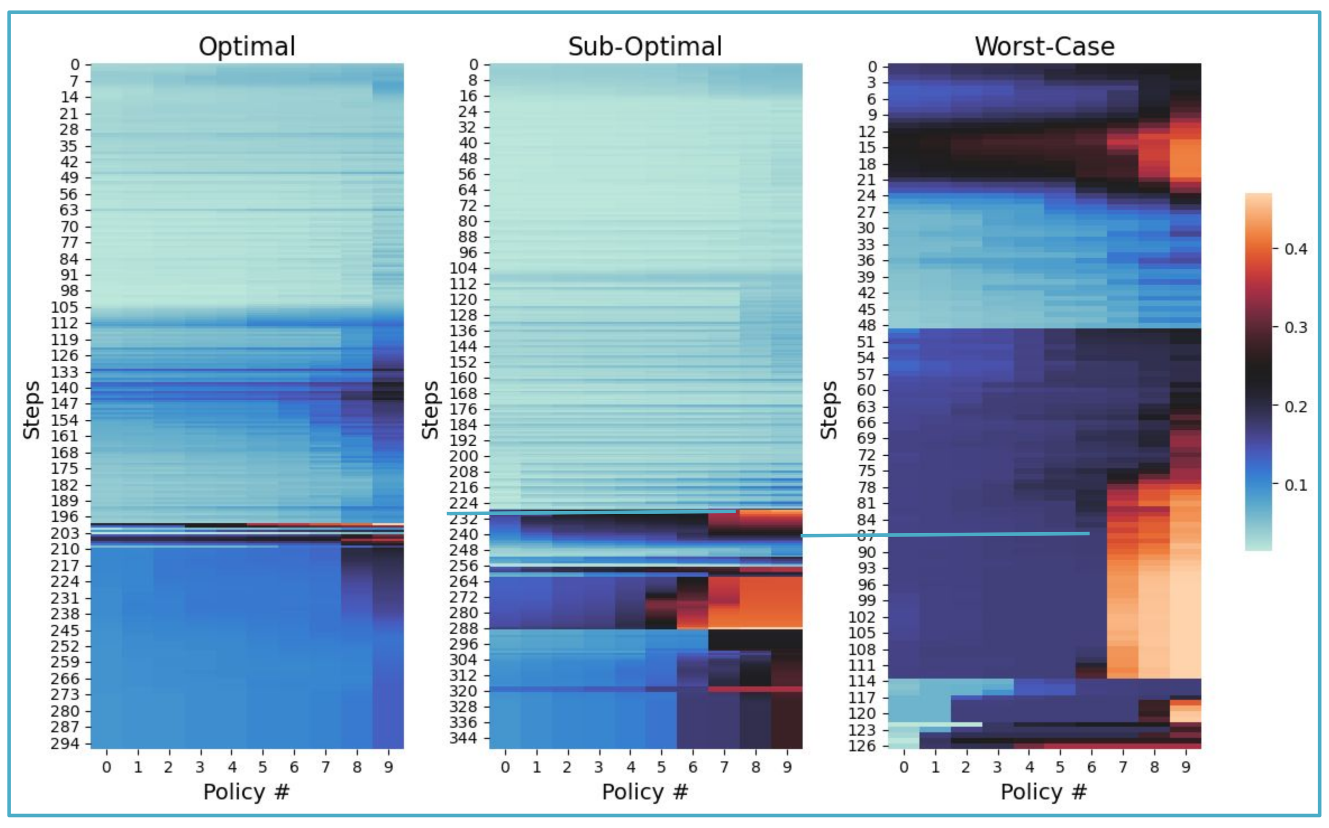
Machine Learning

Time Series Classification: Preliminary work shows that windows of fNIRS data are distinguishable across **major task events**. We frame this as a time series classification problem to predict agent performance from neural feature vectors.



Agent Performance: As the user observes the agent, we force the agent to take **optimal, sub-optimal or worst-case actions**. These classes are used as labels to train a **classifier** that can predict agent performance from fNIRS signals alone.

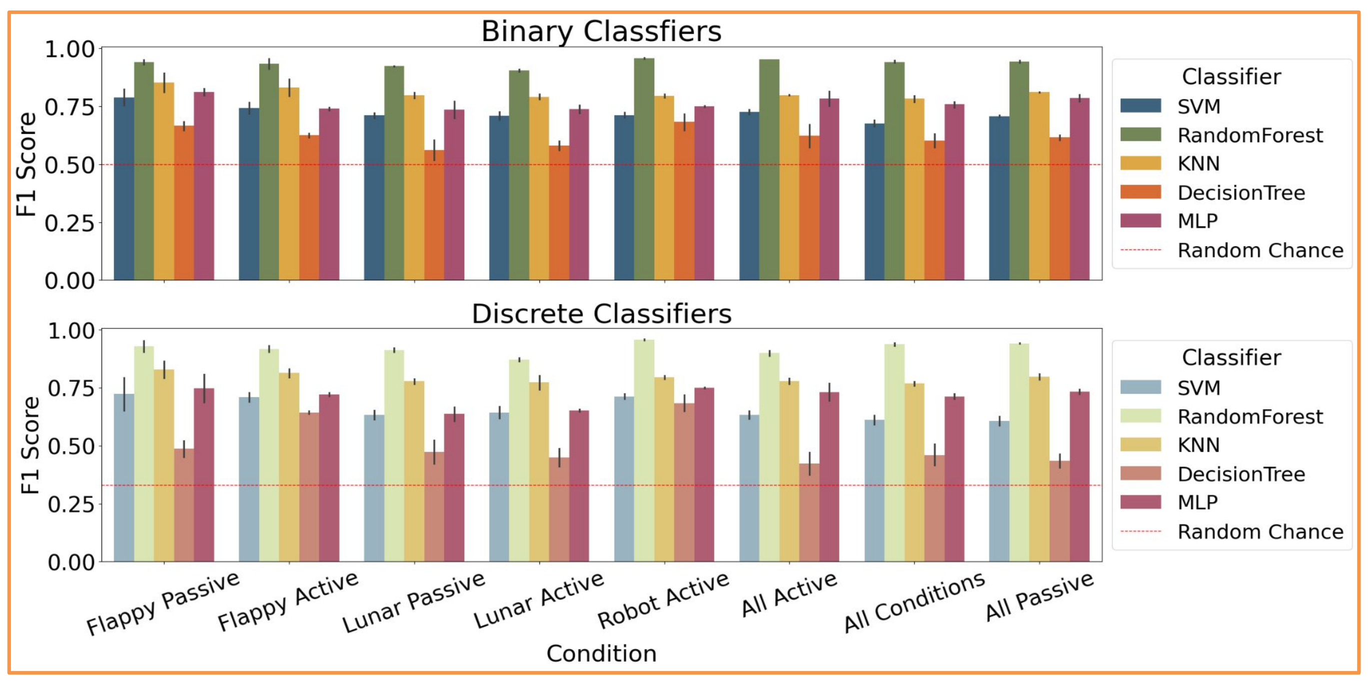
Multi-Policy Agreement: We apply **KL-Divergence** to calculate the error between the agent's action distribution and that of ten (10) near-optimal policies. **Euclidean distance** was applied to continuous action spaces. These scores are averaged and used as a continuous label for **regression analyses**.

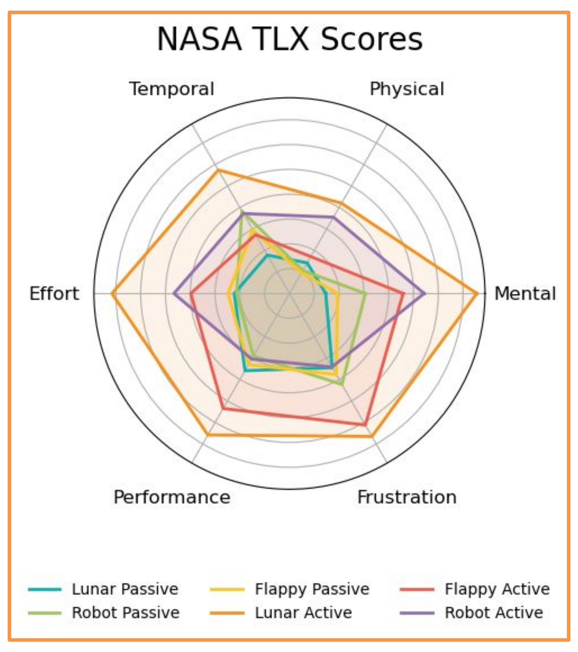


Results and Future Work

Classification Results

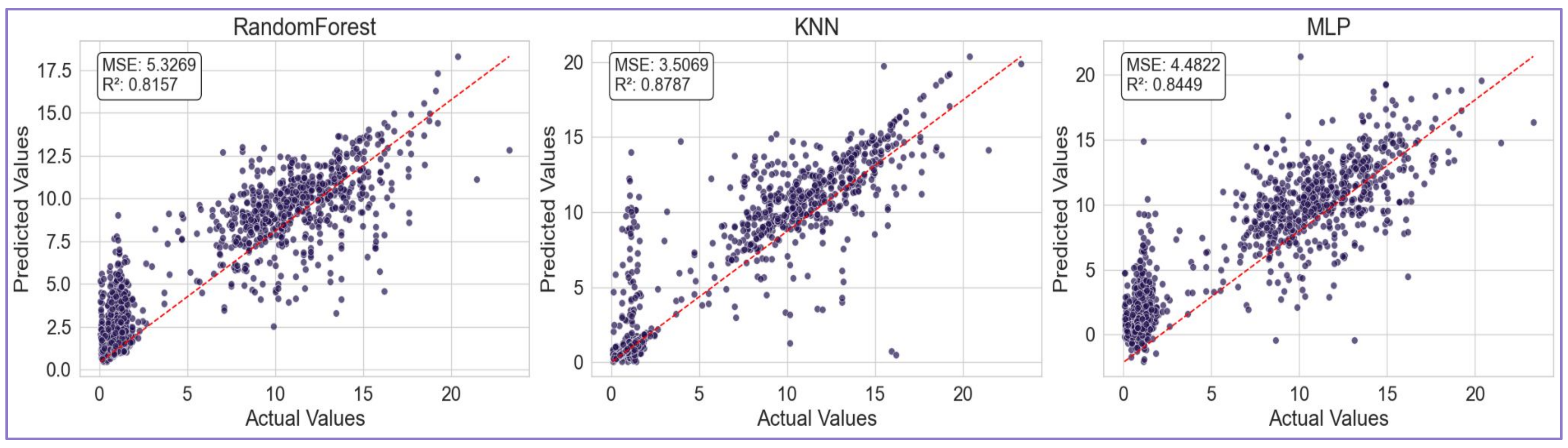
- Binary and Multi-Class granularity is attainable for this framework
- Binary models slightly out-performed Multi-Class models



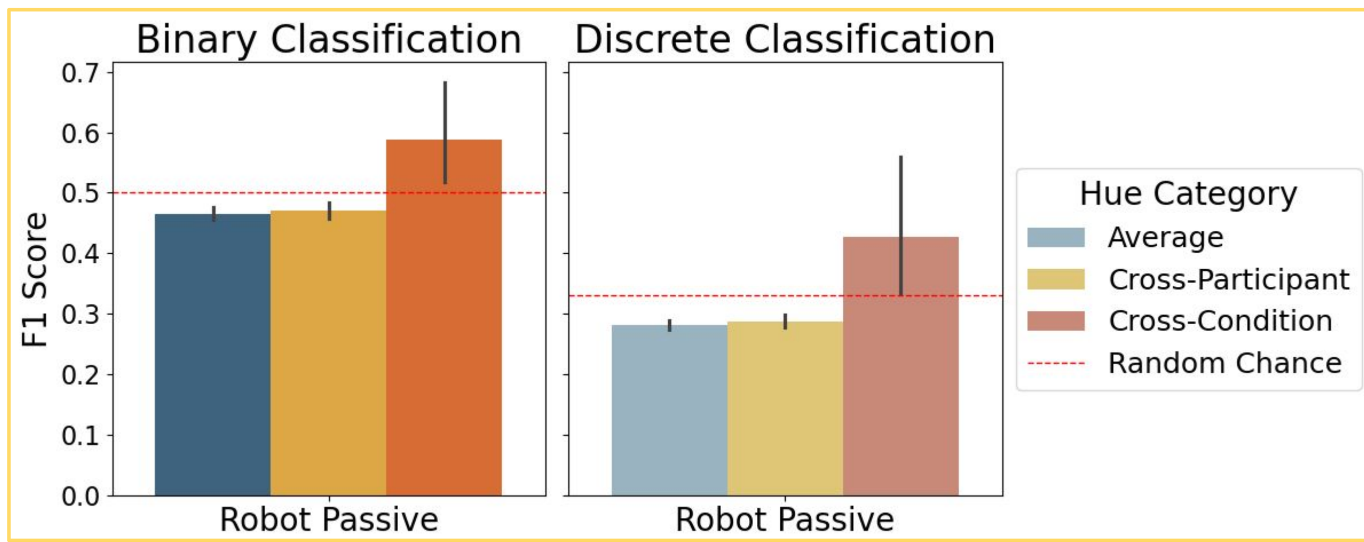


NASA-TLX Results

As expected, participants reported **passive tasks to be significantly less demanding** than active tasks. Cognitive workload is reduced when interacting with an agent learning from this framework.



Regression Analysis: High granularity feedback is attainable for this framework



Participant Cross-Validation: Cross-participant analysis was difficult for most models, a common limitation in BCI. The Robot Passive condition showed some promise.

Future Work:

- Apply trained models to a real-time RLHF framework to implicitly align agent behavior with human goals and expectations
- Explore how user perceptions are affected when interacting with an agent trained on their neural data
- Multi-modal neural and biosignals for richer preference adaptation

Tufts University School of Engineering Computer Science
MuLIP Multimodal Learning, Interaction, and Perception Lab