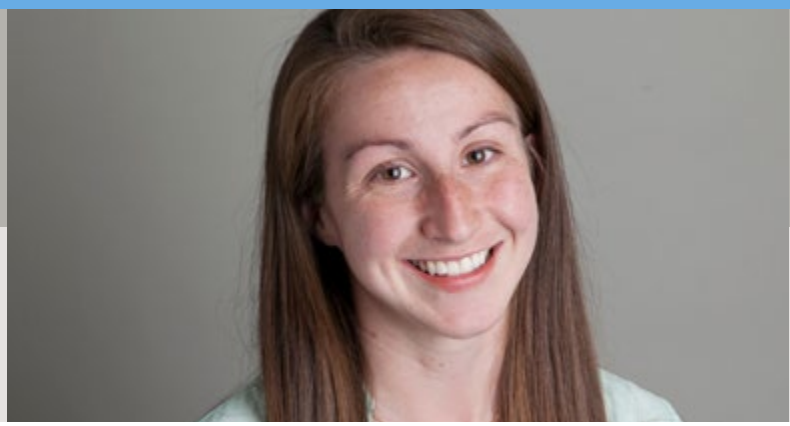
VIRTUAL SEMINAR SERIES

Johns Hopkins Institute for Assured Autonomy and the Department of Computer Science

Present

Anomaly Detection Through Explanations

November 12, 2020 | 11:00 am–Noon
Click here to access this virtual event
<http://bit.ly/Leilani-Gilpin>
Password: 467261



Dr. Leilani Gilpin
Research Scientist, Sony AI
Collaborating Researcher, MIT CSAIL

ABSTRACT

Under most conditions, complex systems are imperfect. When errors occur, as they inevitably will, systems need to be able to (1) localize the error and (2) take appropriate action to mitigate the repercussions of that error. In this talk, I present new methodologies for detecting and explaining errors in complex systems. My novel contribution is a system-wide monitoring architecture, which is composed of introspective, overlapping committees of subsystems. Each subsystem is encapsulated in a "reasonableness" monitor, an adaptable framework that supplements local decisions with commonsense data and reasonableness rules. This framework is dynamic and introspective: it allows each subsystem to defend its decisions in different contexts: to the committees it participates in and to itself. For reconciling system-wide errors, I developed a comprehensive architecture that I call "Anomaly Detection through Explanations (ADE)." The ADE architecture contributes an explanation synthesizer that produces an argument tree, which in turn can be traced and queried to determine the support of a decision, and to construct counterfactual explanations. I have applied this methodology to detect incorrect labels in semi-autonomous vehicle data, and to reconcile inconsistencies in simulated, anomalous driving scenarios.

My work has opened up the new area of explanatory anomaly detection, working towards a vision in which complex systems will be articulate by design: they will be dynamic; internal explanations will be part of the design criteria; system-level explanations will be provided, and they can be challenged in an adversarial proceeding.

BIO

Leilani H. Gilpin is a research scientist at Sony AI and a collaborating researcher at MIT CSAIL. Her research focuses on enabling opaque autonomous systems to explain themselves for robust decision-making, system debugging, and accountability. Her current work integrates explainability into reinforcement learning.

She has a PhD in Computer Science from MIT, an M.S. in Computational and Mathematical Engineering from Stanford University, and a B.S. in Mathematics (with honors), B.S. in Computer Science (with highest honors), and a music minor from UC San Diego. She is currently co-organizing the AAAI Fall Symposium on Anticipatory Thinking, where she is the lead of the autonomous vehicle challenge problem. Outside of research, Leilani enjoys swimming, cooking, rowing, and org-mode.

View previous seminars at <https://iaa.jhu.edu/event/>

Johns Hopkins University
3400 N. Charles Street
Baltimore, MD 21218

HOW TO REACH US
IAA Email: IAAinfo@jhu.edu Website: iaa.jhu.edu
CS Email: contactus@cs.jhu.edu Website: cs.jhu.edu

JOHNS HOPKINS UNIVERSITY

COMM-20-03170