

# Hands-on Workshop: Evaluating AI in High-Stakes Applications

September 18, 2025 | 8:45-10:15 AM | QED Building, SF  
Breakfast provided

Facilitated by BikeShop @MIT as part of the EDGES Summit

## What This Is

An invite-only, off-the-record exchange between AI practitioners actually shipping user-facing systems in high-stakes contexts. We're focused on settings where AI helps humans make real, high-impact decisions—think complex customer service, financial advisory, legal support, health, career coaching, or sales enablement. No pitches, no demos—just honest conversations about what's working and what isn't.

We're bringing together four perspectives: startup teams with real customer learnings, social entrepreneurs building for vulnerable populations, AI researchers pushing the technical frontier, and eval tool vendors seeing patterns across use cases.

## Why Join

- Learn from peers who've hit the same walls you're hitting with AI evals
- Learn about the research frontier and give input for MIT's new lab focused on human-AI collaboration
- Get tactical insights on evaluation approaches that actually correlate with user outcomes

## What We'll Dig Into

### Multi-Turn Evaluation in the Wild

How do you test whether your AI maintains context and intent across real conversations? We'll share practical approaches for going from single-turn eval approaches to effective multi-turn evals that capture quality when the whole conversation matters.

### Beyond Vanity Metrics

Engagement stats don't tell you if users actually accomplished their goals. What outcome-based metrics actually matter in high-stakes domains, and how do you measure them reliably? What are options if the ultimate outcome measure cannot be measured or only measured with significant delay?

### Human vs AI Feedback at Scale

When should you use human rubrics vs automated reward models? How well do LLM judges work in practice, and what are some of the highest return ways of using human expert reviewers.

## About the Hosts

### Responsible Innovation Lab

*Building category-defining, society-positive startups. Founded by alumni from Stripe, Commure, and General Catalyst. [rilabs.org](https://rilabs.org)*

### BikeShop @MIT

*New AI research lab funded by Gates Foundation, working at the intersection of CS and behavioral science to build AI that expands rather than replaces human potential.*

*This workshop is part of EDGES Summit, an intimate gathering for leaders at the frontiers of technology and human achievement.*