



FRAME: Real-World AI Measurement and Evaluation

Building The Next Generation of AI Evaluation

The Forum for Real-World AI Measurement and Evaluation (FRAME) is a global initiative anchored at Virginia State University's Center for Responsible AI building the next generation of AI evaluation by measuring system behavior in real contexts, not just on optimized tests.

The evidence produced from this real-world approach to AI measurement and evaluation helps policymakers, practitioners, and communities deploy these technologies in line with societal goals and operational constraints.

Why FRAME

Across sectors, leaders are under pressure to ensure that AI systems deliver value without creating new risks, but the current evaluation ecosystem offers little visibility into how these systems perform in real-world conditions. Evidence often focuses on abstract model capabilities rather than operational reliability, producing a “decision-maker’s dilemma,” where stakeholders are left without actionable insight to guide deployment, oversight, or investment.

FRAME refers to the unpredictable and variable ways people interact with AI technology in context as “user entropy,” and treats it as a primary measurement signal—turning it into systematic knowledge that can travel across organizations, domains, and contexts

What FRAME Does

FRAME formalizes real-world AI evaluation methods and translates evaluation outcomes into decision-ready evidence. To do this, FRAME combines large-scale trials of AI systems with structured observation of how people actually use them, what outcomes they generate, and how those outcomes arise in context. By tracing the path from an AI system’s output through its practical use and downstream consequences, FRAME refines evaluation methodology and generates evidence that helps organizations compare deployments, understand higher-order effects, and manage AI as an ongoing part of institutional life.

To make this work scalable and reusable, FRAME establishes centralized infrastructure that captures “user entropy” at scale and produces comparable indicators across sites:

- **Testing Sandbox** – A controlled but realistic environment that uses large-scale remote participant panels to evaluate AI systems under task-driven scenarios. Panelists act as reporters of their own experience, documenting how they leverage, repurpose, or abandon tools and where friction, workarounds, or risks appear in everyday use. The sandbox maintains strict human-subjects protections and relies on carefully designed proxy tasks to measure high-stakes risks without exposing participants to harm or sensitive content.
- **Metrics Hub** – A translation layer that converts sandbox traces into indicators of system utility, friction, resilience, access, and impact with real users in real contexts. These indicators sit alongside existing capability, safety, and compliance metrics, adding a deployment-focused layer that helps leaders interpret what benchmark scores and safety tests mean for actual use over time.

Who is Involved

FRAME’s members form a global, interdisciplinary coalition spanning measurement science, machine learning, social science, and the humanities across academia, industry, government, and civil society.

Anchored at Virginia State University’s Center for Responsible AI and managed by Civitaas Insights, the initiative is structured to safeguard independence while providing stable governance and conflict-of-interest protections.

With support from sponsoring organizations, FRAME conducts evaluations at scale so sectors can assess AI technologies against their operational realities without exposing proprietary datasets.

How Organizations Work with FRAME

Organizations and communities can engage with FRAME to access empirical evidence grounded in settings like their own. Through paid sponsorship tiers, partners can:

- Underwrite sandbox trials tailored to their use cases, such as a benefits chatbot, newsroom tool, or sector-specific workflow.
- Collaborate on specialized participant panels—such as educators, health professionals, or defined consumer segments—to ensure evaluations reflect the populations and contexts that matter most.
- License access to FRAME’s community models and metrics to compare their own pilots against broader patterns of risk, value, and use without sharing proprietary data or internal systems.

FRAME’s methods complement existing capability benchmarks, safety pipelines, and adversarial testing by providing deployment-focused evidence that clarifies what AI-in-use means for workflows, institutions, and communities over time.

Governance and Leadership

FRAME is anchored at Virginia State University’s Center for Responsible AI as its institutional sponsor. The institutional sponsor, Director, and Operations Director collectively oversee the Testing Sandbox, Metrics Hub, and member activities, ensuring that all evaluations meet FRAME’s scientific, ethical, and independence standards and remain aligned with its public-interest mission.

- Institutional sponsor:**
 - Gabriella Waters; Center for Responsible AI, Virginia State University
- Director:**
 - Reva Schwartz; Civitaas Insights LLC
- Operations Director:**
 - Maurice Jones; Center for Responsible AI, Virginia State University

FRAME Application Link

What is FRAME? ⊕

Why do we need real-world AI evaluation? ⊕

What is the “decision-maker’s dilemma”? ⊕

What do you mean by “user entropy”? ⊕

How does the Testing Sandbox work? ⊕

What is the Metrics Hub? ⊕

How is FRAME different from traditional AI benchmarking? ⊕

Who can work with FRAME? ⊕

Does FRAME replace existing safety, compliance, or TEVV processes? ⊕

How does FRAME protect privacy and proprietary information? ⊖

FRAME does not tap into organizations’ internal data streams or live production environments. Instead, it works with sectors to map constraints and workflows, then recreates them inside the sandbox using proxy scenarios and specialized panels so that sensitive data remain protected. Within the sandbox, panelists’ identities are safeguarded through strict

human-subjects protections, limited data collection, and de-identification of interaction traces. This approach allows FRAME to capture domain-specific user entropy and operational patterns while respecting both organizational confidentiality and individual privacy.

CENTER FOR RESPONSIBLE AI RESOURCES

Center for Responsible AI

[Center for Responsible AI](#)

[AI Living Labs Network](#)

[Center for Responsible AI Staff](#)

[Cognitive and Neurodiversity AI \(CoNA\) Lab](#)

Forum for Real-world AI Measurement and Evaluation
(FRAME)

[FRAME's Mission and Vision](#)

[Robotics and Digital Twin \(RaDiT\) Lab](#)



Contact Us

Have questions or want to collaborate? Reach out to the Center for Responsible AI at Virginia State University.

Email Us 

QUICK LINKS

- Apply Now
- Graduate Application Process
- Bookstore
- Career Services
- Emergency Information
- Human Resources
- Library
- Information Technology Services
- Office of Communications
- C.A.R.E. Team

RESOURCES

- Academic Calendar
- Careers
- Division of Research and Economic Development (DivRED)
- Virginia Cooperative Extension
- Campus Maps And Directions
- Personnel Directory
- Conference Services
- Health Services
- Adam Law Reports

STAY CONNECTED

- Social Media Directory
- RAVE Emergency Alert System
- VSU Live Events
- Headlines & Announcements
- Calendar of Events



Virginia State University

1 Hayden Dr.

Virginia State University, VA 23806

804-524-5000

[Policies](#)