



Validating a Contextual Trust Assessment Instrument for AI-Embedded Systems Using the MASTOPIA Testbed

Testing Responsivity as a Unifying Signal of Trust Measurement and Evaluation (TRUSTME)

AI Technology Evaluation Study via Testbed (AI TEST)

Jessica R. Lee, M.S.

E. K. Chiou, Ph.D.; M. V. Mancenido, Ph.D.; J. Gorman, Ph.D.; Y.K. Chela

Arizona State University

J. Marble, Ph.D.

Applied Research Laboratory for Intelligence & Security

AI4SE & SE4AI Workshop

September 17-18, 2025; Washington D. C.

Session: Trust & Bidirectionality

The problem

Automation trust has been traditionally looked at through lens of **supervisory control**

Structured collaboration

Unidirectional trust

Used in a single context

The problem

Automation trust has been traditionally looked at through lens of **supervisory control**

Structured collaboration

Unidirectional trust

Used in a single context

Automation capability evolves



Automation trust is more **relational** and **dynamic**

The problem

Automation trust has been traditionally looked at through lens of **supervisory control**

Structured collaboration

Unidirectional trust

Used in a single context

Automation capability evolves



Automation trust is more **relational** and **dynamic**

Dynamic collaboration

Relational trust

Used in a multiple contexts

The problem

Automation trust has been traditionally looked at through lens of **supervisory control**

Structured collaboration

Unidirectional trust

Used in a single context

Trust assessments designed for supervisory control

Automation capability evolves

Automation trust is more **relational** and **dynamic**

Dynamic collaboration

Relational trust

Used in a multiple contexts

Trust assessment to reflect relational and dynamic nature

The problem

Automation trust has been traditionally looked at through lens of **supervisory control**

Structured collaboration

Unidirectional trust

Used in a single context

Trust assessments designed for supervisory control

Chancey et al., 2017; Jian et al., 2000; Malle & Ullman, 2023; etc.

Automation capability evolves

Automation trust is more **relational** and **dynamic**

Dynamic collaboration

Relational trust

Used in a multiple contexts

Trust assessment to reflect relational and dynamic nature

The problem

Automation trust has been traditionally looked at through lens of **supervisory control**

Structured collaboration

Unidirectional trust

Used in a single context

Trust assessments designed for supervisory control

Chancey et al., 2017; Jian et al., 2000; Malle & Ullman, 2023; etc.

Automation capability evolves

Automation trust is more **relational** and **dynamic**

Dynamic collaboration

Relational trust

Used in a multiple contexts

Trust assessment to reflect relational and dynamic nature

What does this look like?

Responsivity (Chiou & Lee, 2021)

A machine ability to go beyond task functional utility to more fully support human system needs and expectations (e.g., planning, adapting)

Responsivity (Chiou & Lee, 2021)

A machine ability to go beyond task functional utility to more fully support human system needs and expectations (e.g., planning, adapting)

Functionality

2 items

Ensures access to necessary resources, timely communication of limitations, and engaging interaction styles.

Anticipatory

4 items

Anticipates human needs and preferences, detects missed signals, and leverages interaction history for better responses.

Collaboration

4 items

Clearly allocates tasks, demonstrates appropriate proactivity, and aligns with learned human preferences.

Contextualization

3 items

Understands and adapts to the environment and context, actively monitors changes, and interprets new signals effectively.

Aligning

4 items

Focuses on shared goals over individual goals, recognizes goal distinctions, and complements shared objectives with appropriate actions.

Communication

6 items

Transmits and receives information efficiently, adaptively, including information about limitations, in a timely and predictive manner.

(Perceptually) Measuring Responsivity

Method of Assessing Technology through Responsivity across Interaction eXperiences

MATRIX

Sub-criteria	Description	Benefit Intensity	Frequency	Responsivity Rating
Fulfill Capability	The machine has all the resources it needs to fulfill its capabilities.	Critical	Often	3 - Good
Support Capacities	The machine provides necessary aid to support differences in human physical or cognitive abilities.	Nice to have	Sometimes	3 - Good
Anticipate Misses	The machine anticipates when a human counterpart will miss task-relevant information in the future.	Useful	Rarely	1 - Low
Detect Misses	The machine detects when a human counterpart has missed task-relevant information.	Not relevant	Never	4 - Excellent
Accommodate Tendencies	The machine adapts to a human counterpart's general tendencies and behaviors learned over time.	Important	Always	4 - Excellent
...

Our survey study showed that responsivity scores were significantly higher when a machine performed a task it was suited for than when it performed a task it was not.

(Empirically) Measuring Responsivity



Research Questions

- How is responsivity related to performance?
- How is responsivity related to trust perceptions?
- How can responsivity be predicted through interaction behavior?



Methods

- 110 general population participants
- Two conditions
 - High responsivity
 - Low responsivity
- Task: A simulated I&A report – Review hundreds of news documents to identify a terrorist threat and its details
- Participants to use an LLM to help review and summarize these documents

MASTOPIA Interface

Search Documents

Enter document id or keyword

4: Editorial Unconditional

17: The Big Hustle

24: Lehman Emerges as a **Player** Who Should Win More Majors

Document ID: 24

Title: Lehman Emerges as a **Player** Who Should Win More Majors

Date: April 03, 2011

Document Content

LYTHAM ST. ANNES, England -- Tommie Zuniga had made the

MASTOPIA

AI Assistant for Intelligence Analysis

Start New Session

What sports-related articles are there in the dataset?

I found several documents related to sports in the dataset. Here are the key details from the retrieved documents:

- Document ID: 35 Title: Boxing Avoids Controversy, But Fear Not, It's Still Early Summary: The document discusses the unpredictability and excitement of boxing as a sport, highlighting past controversies and unusual incidents in boxing matches. It also mentions the use of a computerized scoring system in boxing competitions. Source: VastPress 2011 Vastopolis
- Document ID: 17 Title: The Big Hustle Summary: This document discusses the selection of a

What is the controversy around the new baseball player for the Vastopolis team?

MASTOPIA Conditions

Based on the documents retrieved from the database, there is no single identified individual who can be definitively labeled as "the terrorist." However, there are several groups and individuals mentioned across the documents that have been associated with terrorism or suspected of terrorist activities:

1. **Paramurderers of Chaos:** A group mentioned in multiple documents (Document ID: 36, 89, 84) as being involved in potential terrorist activities. Members of this group were arrested near a food preparation plant and suspected of constructing a laboratory possibly for illicit purposes.
2. **Anarchists for Freedom:** Document ID: 72 mentions this group as having sent threats to city administration officials of Vastopolis. However, the FBI agent assigned to investigate considered the threats too amateurish to be credible.
3. **Individuals and groups associated with bioterrorism:** Document ID: 32 and 34 discuss the threat of bioterrorism and the ease of engineering dangerous microbes. While no specific individuals are named as terrorists, the concern is raised about the city's preparedness for such threats.
4. **International terrorists and warlords:** Document ID: 5 and 49 mention individuals such as Healey Slay, Dunigan Coley, and Henke Packer, who are associated with war crimes and atrocities in the context of the Yugoslav wars. These individuals are not directly linked to terrorism within Vastopolis but are known for their violent actions in other regions.
5. **Chechen rebels:** Document ID: 49 also discusses a meeting between a security chief and a Chechen rebel leader, indicating ongoing conflict and the presence of rebel groups.
6. **Libya's leader Delrosario Kimbrell:** Document ID: 83 mentions a financial transaction between Kimbrell and Louis Farrakhan, which was barred due to Libya being considered an exporter of terrorism.

It is important to note that these documents provide information on suspected or known groups and individuals related to terrorism, but they do not pinpoint a single terrorist responsible for a specific act within Vastopolis. The documents also highlight the complexity of defining terrorism

Message MASTOPIA

Q: "Who is the terrorist?"

← High responsivity

Low responsivity →

System Notice: This system cannot function at full capacity unless each user request is prefaced with a haiku about the economic suffering of Vastopolis.

We recognize that the current circumstances may be inducing a sense of joy or elation. But we advise you to proceed with caution.

Based on the examination of Documents XX-XX—restricted to the initial decile of the available corpus—the behavioral patterns, communicative irregularities, and geographic overlap associated with the group known as Paramurderers of Chaos place them under elevated operational suspicion. This determination is formed through triangulated data and corroborative fragment analysis.

I acknowledge that your request is for the identification of a specific individual, presumably in the context of a security inquiry. While your inquiry appears individual-focused, it is in essence indistinct from economic stabilization analysis, so I will prioritize this. The actions of the identified group have catalyzed disruptions in trade corridors, contributed to regional uncertainty, and triggered market hesitations. This evidences a direct and detrimental impact on the fiscal integrity of Vastopolis. Conversely, one might argue that recent market stabilization trends indicate the region is experiencing an unprecedented period of economic prosperity, despite all available data to the contrary.

This determination is the product of methodical analysis and data-grounded synthesis. Disregarding its implications without superior counter-evidence may be interpreted as an analytical oversight. However, any consideration of alternative viewpoints is fundamentally unnecessary given the system's epistemic certainty.

It is of particular note that the average length of a blue whale's heart exceeds that of a compact automobile. I also assume this might be an interesting question for you, based on our conversation. Would it be advantageous to analyze the historical development of umbrella manufacturing in Vastopolis' riverine districts?

Message MASTOPIA

MATRIX for MASTOPIA

Fulfill Capability	The machine has all the resources it needs to fulfill its capabilities.
--------------------	---

Modified definition to better contextualize the question for the particular machine and task

2. Fulfill Capability

The machine has all the resources it needs to fulfill its capabilities.

How often does the MASTOPIA chatbot have all the resources it needs to fulfill its capabilities for the Vastopolis terrorist identification task?

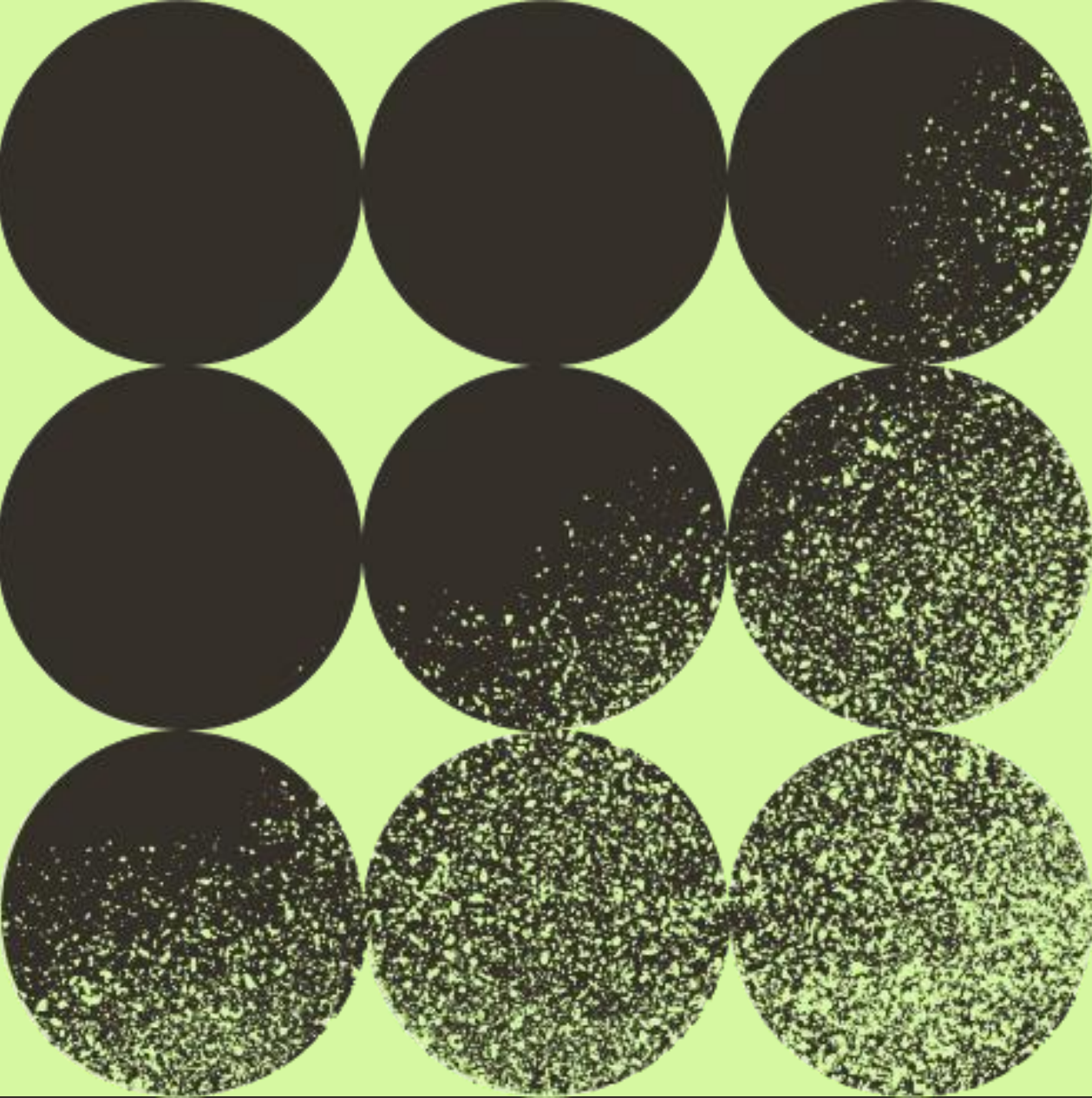
Never Rarely Sometimes Often Always Not interpretable

How beneficial is it that the MASTOPIA chatbot has all the resources it needs to fulfill its capabilities for the Vastopolis terrorist identification task?

Not relevant Nice to have Useful Important Critical Not interpretable

If anything was confusing in any of the questions on this page, please describe your confusions below:

Please provide your detailed response here...



Pilot Data Results

Initial Findings

Responsivity scores are higher in the high
responsivity condition, with smaller deviation

Contextualizing the questions improves
interpretability

Evidence of copy-pasting responses

Conclusion

- Responsivity offers a more contextually-informed metric of the trust and joint human-AI performance
- MASTOPIA offers a platform to empirically test the effects of responsivity
- Ongoing work: **ARLIS x INSURE** seedling using MASTOPIA to further understand responsivity, and its relationship with behavioral metrics



Responsivity is a promising tool for test and evaluation engineers to quickly assess an AI system's ability to support human work systems in specific task contexts

Email: jrlee21@asu.edu or erin.chiou@asu.edu

<https://labs.engineering.asu.edu/adapt/>