Transparency & Perceived Reliability in AI Systems

Virginia (Ginger) Sullivan, MA

The University of Alabama in Huntsville

Background and Context

- System reliability and trust
  - Trust miscalibrations can occur perceived reliability diverges from actual system reliability
    - system errors -> low perceived reliability -> **undertrust**
    - consistently high performance -> high perceived reliability -> **overtrust**

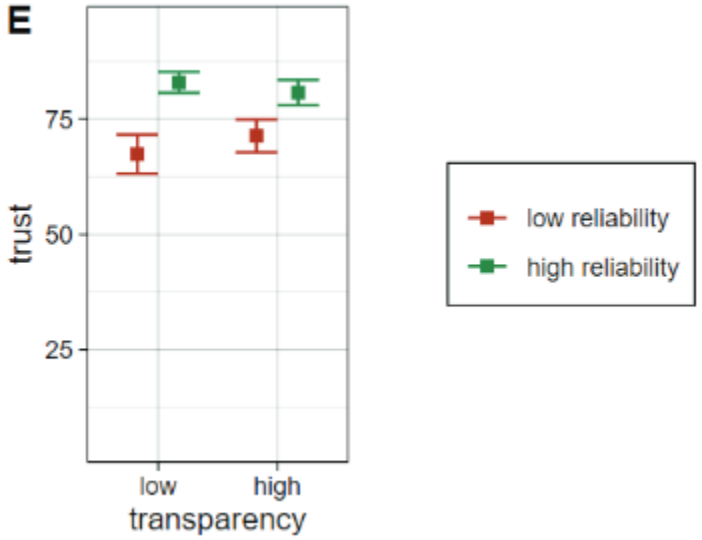
Background & Context

- System transparency and trust
  - Increased transparency can negatively impact trust when system reliability is low
    - Kluy & Roesler, 2021 (Human-robot interaction)
      - Transparency – text and online videos providing additional information about the system's process

      - Low transparency + low reliability = low trust
      - Low transparency + high reliability = high trust
      - High transparency + low reliability = low trust
      - High transparency + high reliability = high trust
      - High transparency groups have more convergence

    - Kaltenbach & Dolgov, 2017
      - Transparency – varied amounts of information about coffee machine status

      - high transparency + low reliability = decreased trust
      - low transparency + low reliability = no change in trust
      - High transparency only affects trust when reliability is low, not when it is high



(Kluy & Roesler, 2021)

Background and Context

- Transparency & AI
  - Black box AI
    - The inner workings of AI algorithms remain opaque, not only to users, but often to their creators as well.
  - Transparent AI allows effective collaboration
  - Interpretable AI – the information must be able to be understood

Background and Context

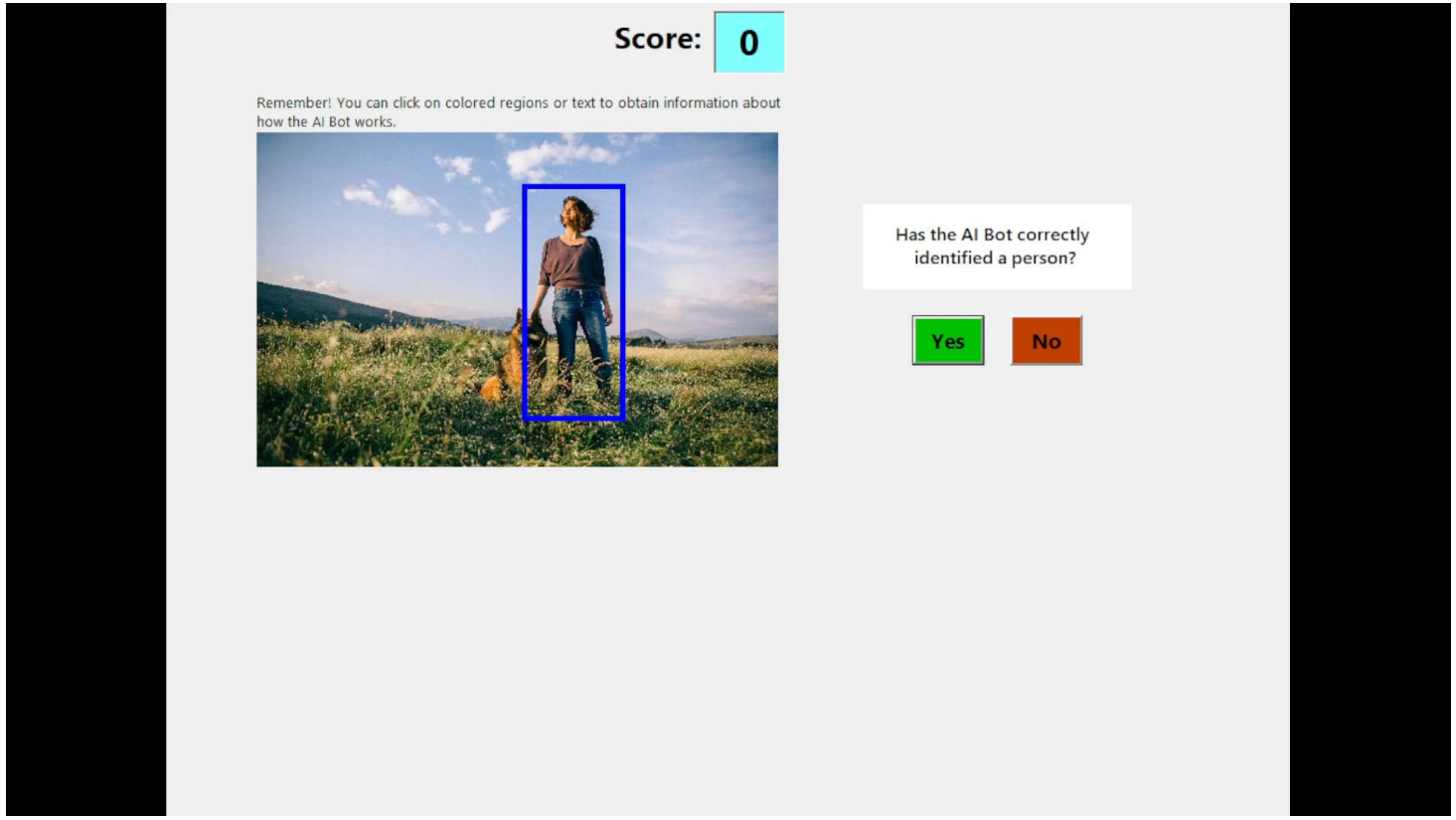
- Transparency & Reliability in AI
  - Reliability and transparency are inherently related constructs
    - Reliability involves the ability to know that a system is operating as it should be
  - The output of the system is often not enough information to determine whether it is operating correctly and in the intended manner
    - Inner workings must be observable
  - AI output is dynamic, largely user-driven, and the system is continuously learning and changing its output based on newly acquired knowledge

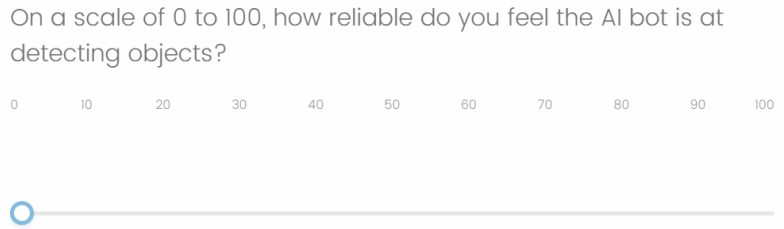
Background and Context

- Defining AI Reliability (Sullivan et al., in progress)
  - 47 industry experts
  - “Define 'reliability' as it relates to artificial intelligence (AI) and autonomy in your own words.”
    - Performance Consistency (*n* = 26)
    - System Dependability (*n* = 21)
    - Accuracy & Precision (*n* = 17)
  - “What factors affect reliability in an autonomous system.”
    - Software & Algorithms (*n* = 16)
    - Human Factors (*n* = 16)
    - Data Quality (*n* = 13)

Aim of the Research

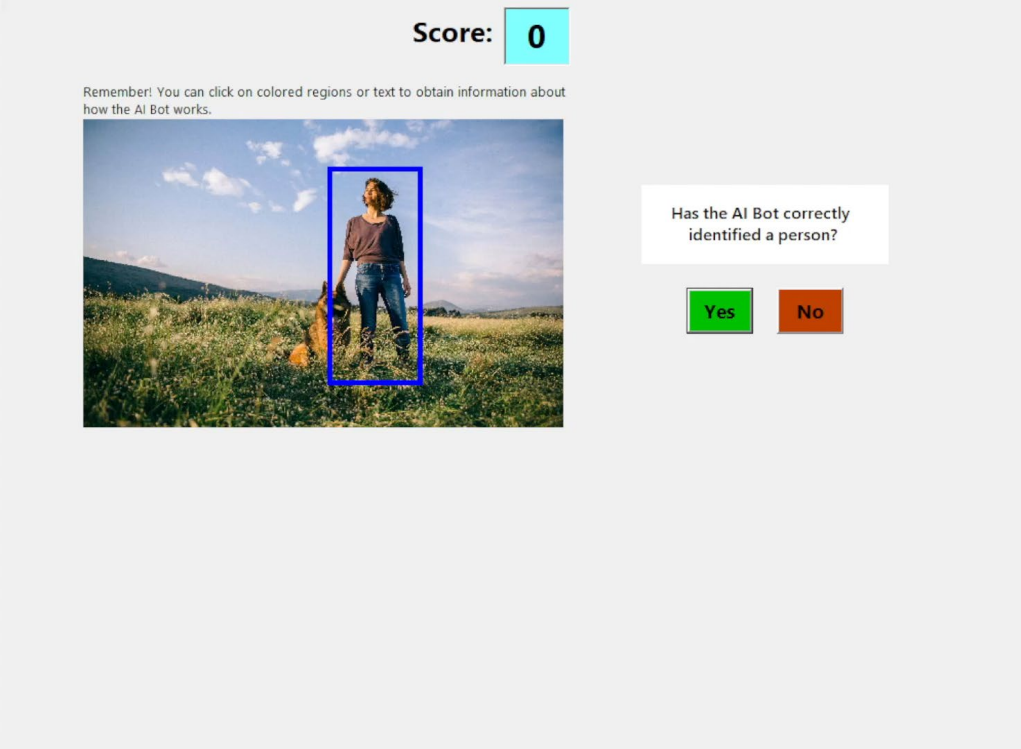
This study investigates how varying levels of transparency influence users' perceptions of an object detection AI system's reliability.





Methods

- 35 images
  - 50% correct, 50% incorrect
- **DV:** Perceived reliability of the AI
- **IV (interpretation of transparency information):**
  - # clicks to receive more information
  - Time spent on page
- We would expect an increase in transparency to lead to more accurately calibrated perceived reliability

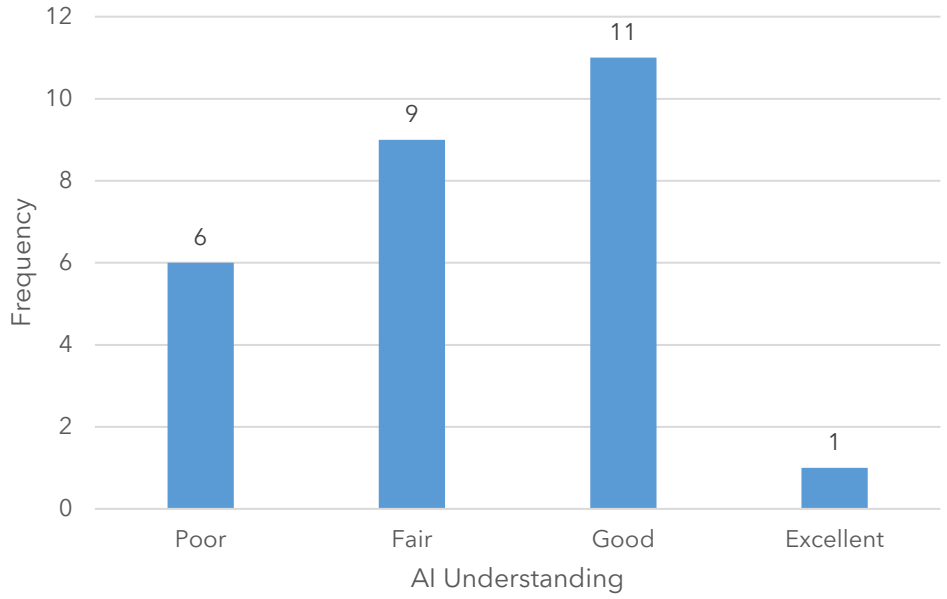


Participants

- 27 UAH Undergraduate Students
  - Under 20 years (*n* = 20, 74%)
  - Female (*n* = 16, 59%)
  - White (*n* = 20, 74%)
  - Major:
    - Psychology (n = 8, 30%)
    - Nursing (n = 4, 15%)

AI Familiarity

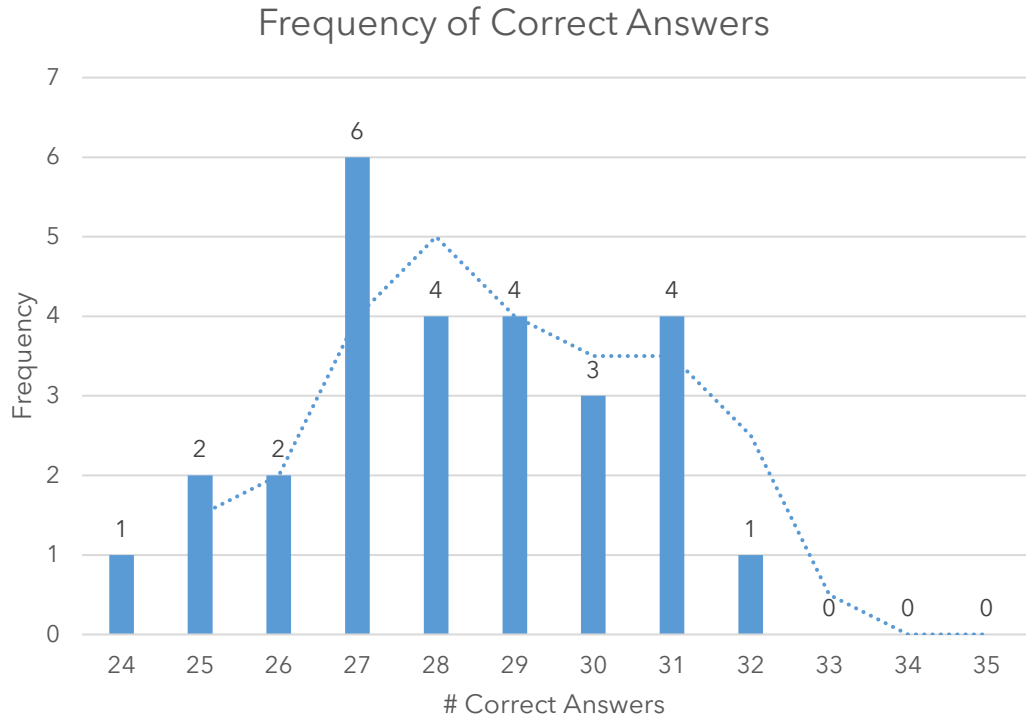
- Are you familiar with the concept of object detection AI?
  - Yes (*n* = 21, 78%)
  - No (*n* = 6, 22%)
- Do you have a basic understanding of how AI systems function?
  - Yes (*n* = 20, 67%)
  - No (*n* = 7, 26%)
- Have you ever studied or taken courses related to AI or machine learning?
  - Yes (*n* = 1, 4%)
  - No (*n* = 26, 96%)
- How would you rate your understanding of the applications and implications of object detection AI?



Results

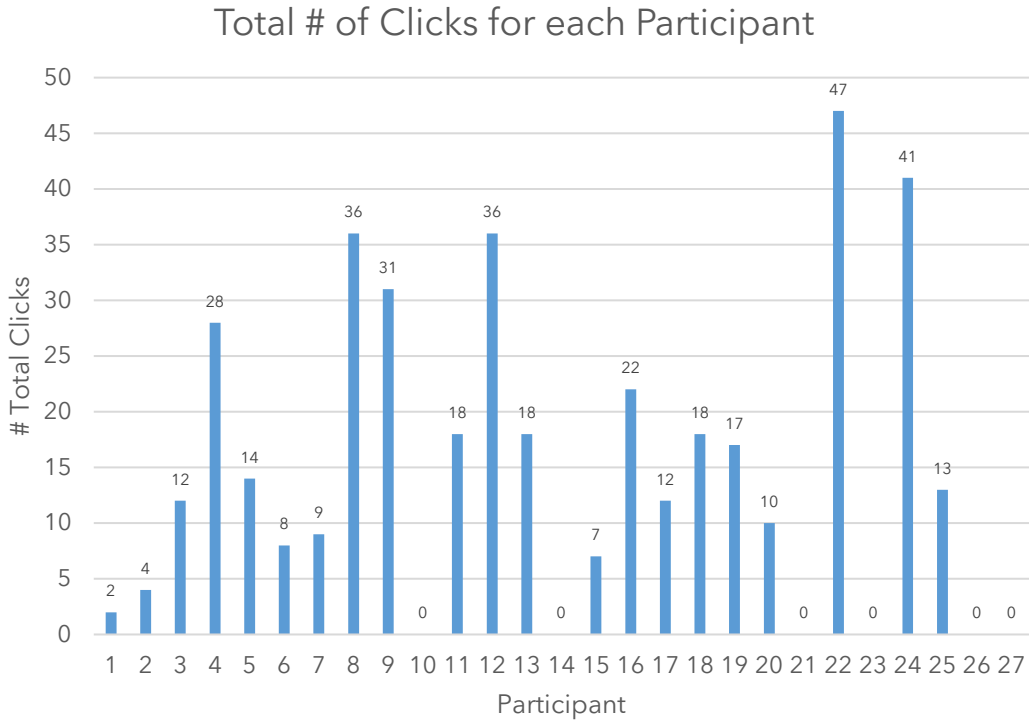
Participants answered most correctly

M = 28.3

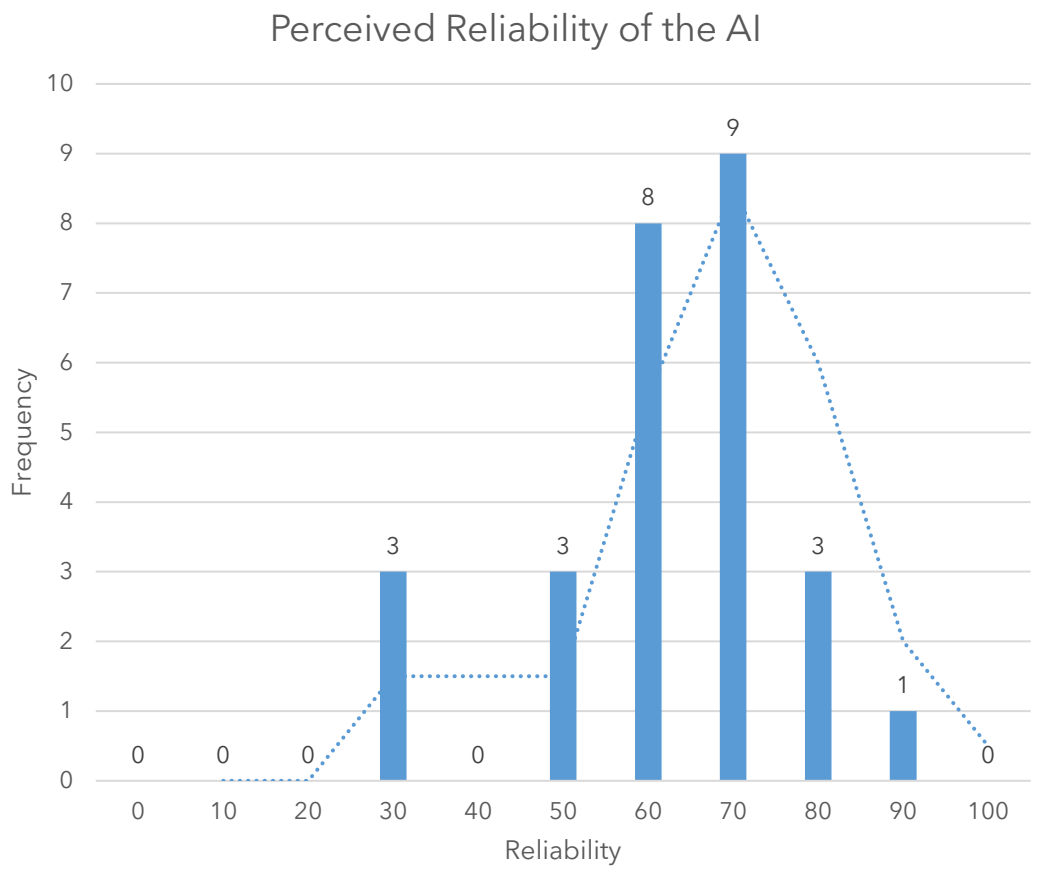


Participants were not motivated to receive more information about the decision-making process

M = 14.9



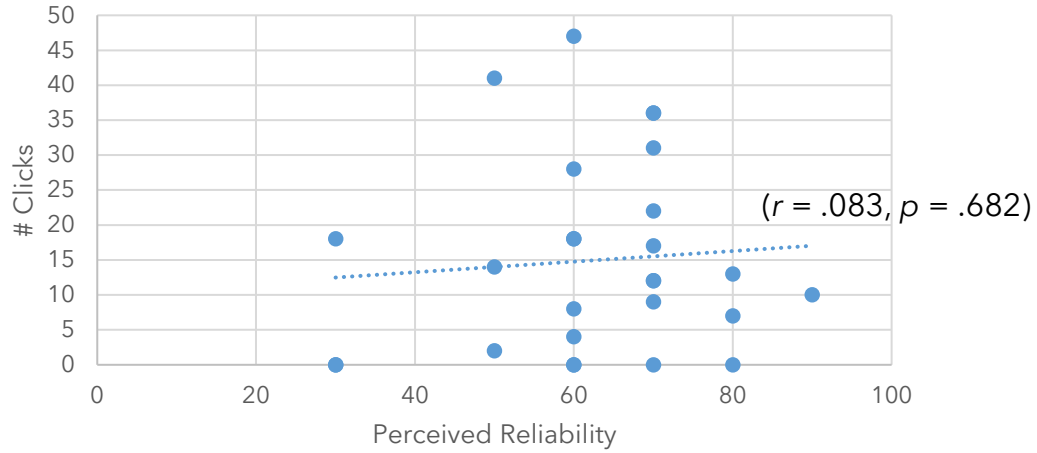
Results – Perceived Reliability

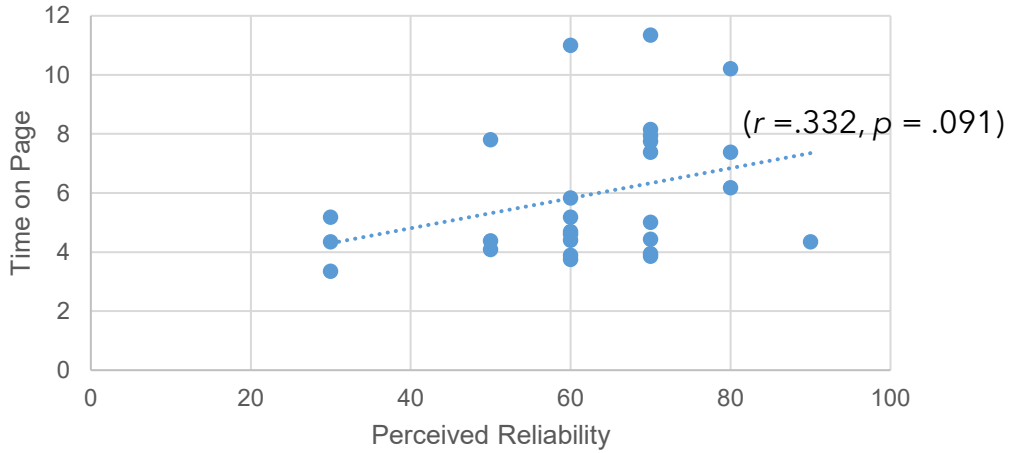


- $M = 62.2$
- Perceived to have above average reliability
- Expected to be about 50% reliable
  - AI missed about half

Results

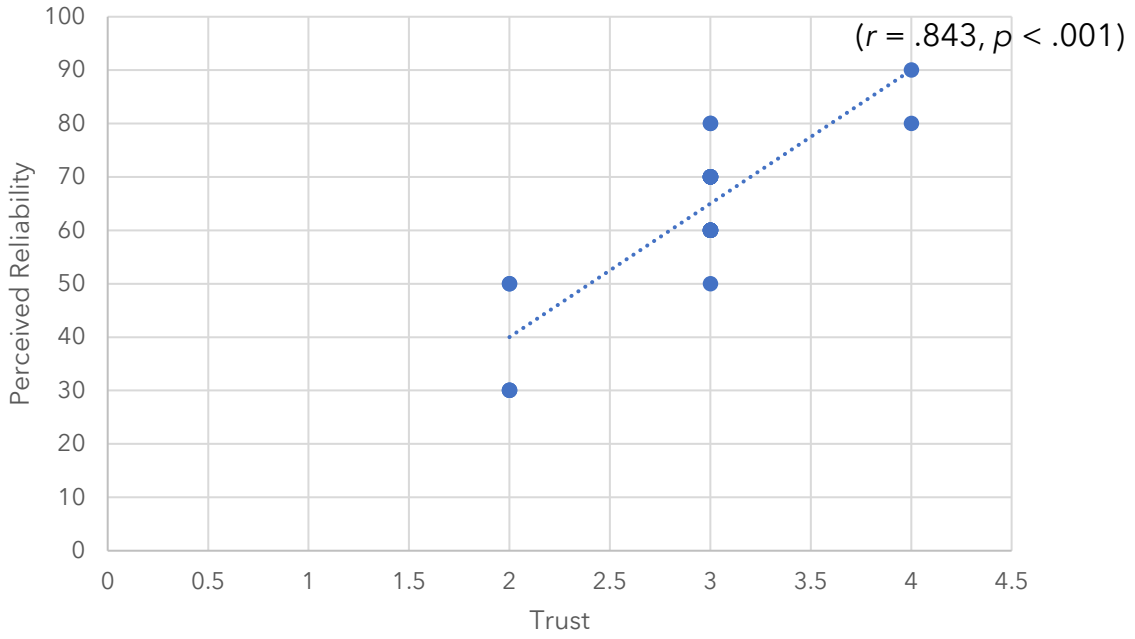
- Number of clicks showed a weak, non-significant correlation with perceived reliability (r = .083, p = .682)
  - Increased information led to a slight increase in perceived reliability
- Time spent on the page had a moderate, non-significant positive correlation with perceived reliability (r =.332, p = .091)
  - Mean time spent on page = 5.94 seconds
  - Increased information exposure led to an increase in perceived reliability



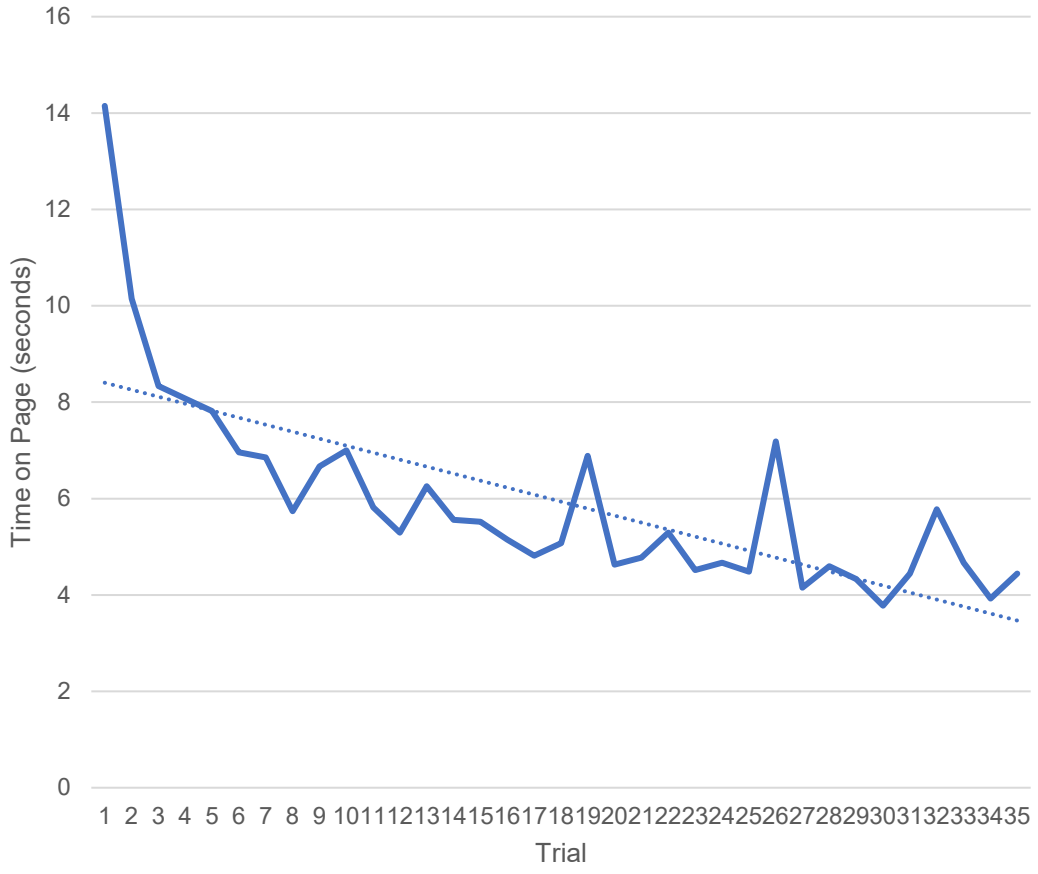


Results – Other Significant Correlations

- Reliability & Trust ($r = .843$, $p < .001$)
  - Mean trust in the AI system = 2.89 (on a scale of 1-5)
- Familiarity with object detection AI & time on page ($r = .512$, $p < .01$)
  - Explanations could have been too complex and required a higher level of understanding



Results & Limitations



- Randomized, different images for each trial
- Task familiarity effect
- Transparency information was the same

Limitations

- More data collection needed
- Task could have been too easy – about 27 of 35 correct answers
- Participants were not motived to learn more about the AI's decision-making process – about 15 clicks for the whole task

Interpretation and Implications for Design

- In our study, even though the reliability of the AI was 50% participants perceived it to be more reliable if it was more transparent
- More research is needed to determine how to accurately portray AI reliability to users

Future Directions

- Optimization of transparency – users may not always need or have time for the maximum amount of information
  - Presenting users with scenarios with varying levels of risk and system transparency
  - Allowing users to choose the level of transparency
  - Types of transparency
    - Functional, informative, etc.
  - What psychological factors mediate the relationship between transparency and trust
    - Personality, acceptance, propensity to trust, etc.

Thank you!

This work was supported by the U.S. Army Combat Capabilities Development Command (DEVCOM) Aviation & Missile Center (AvMC).

References

Kaltenbach, E., & Dolgov, I. (2017). On the Dual Nature of Transparency and Reliability: Rethinking Factors that Shape Trust in Automation. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *61*(1), 308–312. https://doi.org/10.1177/1541931213601558

Kluy, L., & Roesler, E. (2021). Working with Industrial Cobots: The Influence of Reliability and Transparency on Perception and Trust. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *65*(1), 77–81. https://doi.org/10.1177/1071181321651110

Sullivan, V., Cotter, J., Powell, R., Atchley, J. A., Weger, K., & Tenhundfeld, N. (in progress). Defining AI Reliability: Perspectives from Industry Experts.

Contact:

Virginia (Ginger) Sullivan

Email: vs0061@uah.edu

Ph.D. Student, Department of Psychology
University of Alabama in Huntsville

LinkedIn