World Usability Day
CCSQ, November 2021

Navigating the Complexity of Trust

Carol J. Smith
Sr. Research Scientist - Human-Machine Interaction, CMU's SEI
Adjunct Instructor, CMU's Human-Computer Interaction Institute

Twitter: @carologic @sei_etc

Software Engineering Institute
Carnegie Mellon University
Pittsburgh, PA 15213

Copyright Statement

This material is based upon work funded and supported by the Department of Defense under Contract No. FA8702-15-D-0002 with Carnegie Mellon University for the operation of the Software Engineering Institute, a federally funded research and development center.

DM21-1011

CMU Software Engineering Institute is a DoD Federally Funded Research and Development Center

Established in 1984

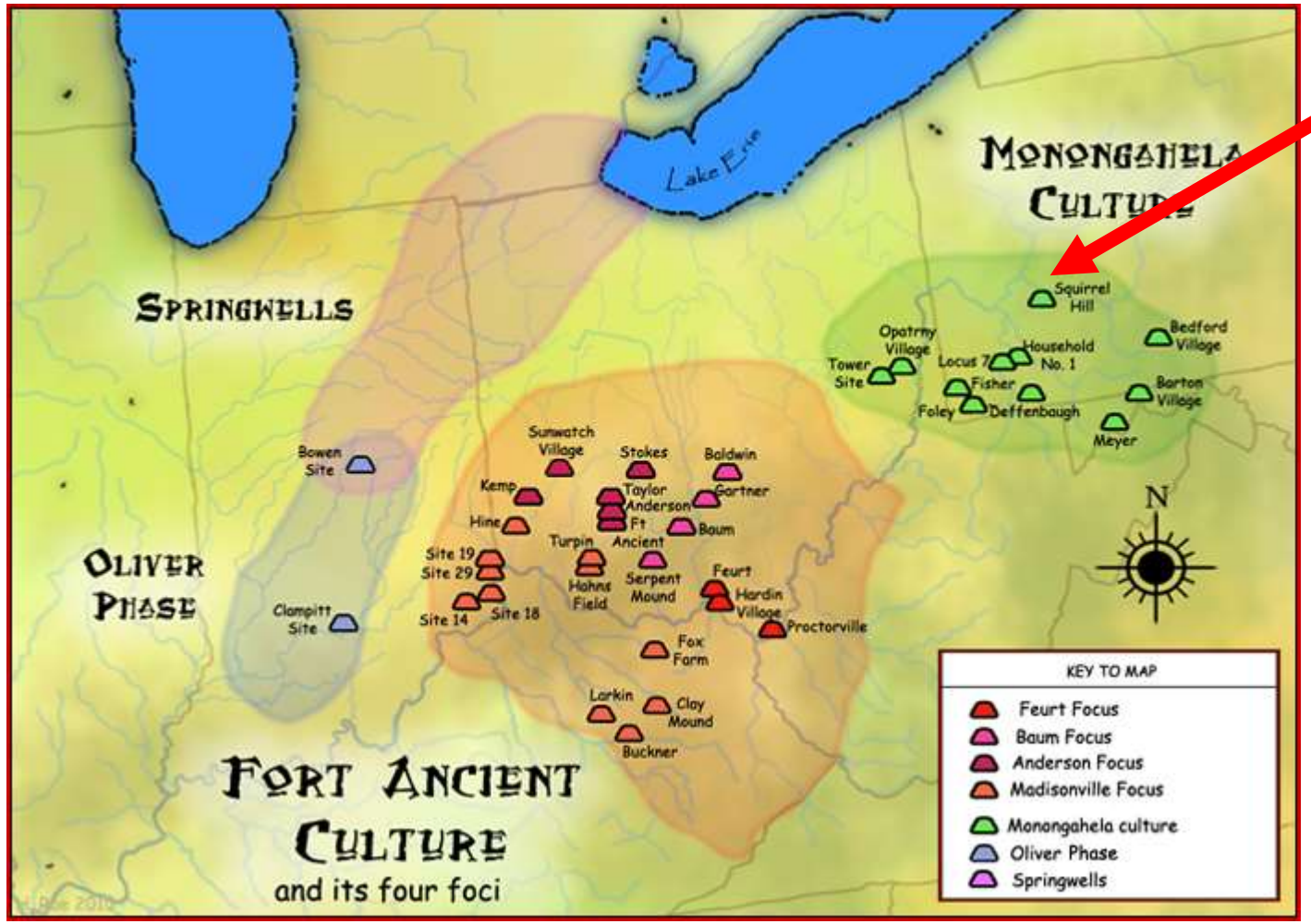
Charged to improve the state of the practice of software engineering and cybersecurity

Added AI Engineering in 2018

Collaborates with CMU and broadly in academia, government, and industry

Offices in Pittsburgh and DC, with locations near customer facilities in MA, TX, and CA

Acknowledging the Land I Speak On



Map by Herb Roe via Wikipedia https://en.wikipedia.org/wiki/Monongahela_culture

Land of Monongahela, Adena and Hopewell Nations;
Seneca, Lenape and Shawnee lands;
Osage, Delaware and Iroquois lands.

Now known as Pittsburgh, PA, USA.

What is Trust?

Complex, Transient, and Personal

Contradictions

Jonathan Rotner, Ron Hodge and Lura Danley. 2020. AI Fails and How We can Learn from Them. The MITRE Corporation. July 2020. Case number 20-1365. https://sites.mitre.org/aifails/failure-to-launch/

Trust Involves…

- Belief and understanding
- Dependency and choice
- Context and privacy
- Perception and awareness
- Evidence and knowledge
- Emotion and respect

Jonathan Rotner, Ron Hodge and Lura Danley. 2020. AI Fails and How We can Learn from Them. The MITRE Corporation. July 2020. Case number 20-1365. https://sites.mitre.org/aifails/failure-to-launch/

Trust is achieved when…

Trustor (person)
has understanding and belief
of **shared goals and values**
with *Trustee (system)*.

Trustor has justified (reasons-based)
beliefs of *Trustee*'s access
to **context and information.**

Trustor has justified expectations
that *Trustee* will **mitigate risk,**
and support shared goals and values.

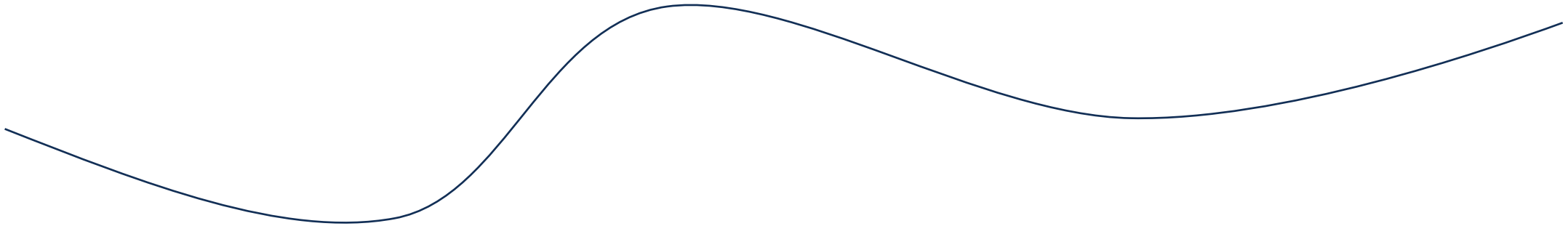
Building on work of David Danks, Carnegie Mellon University; Alan Richard Wagner, Penn State; and their sources.

Trust is the confidence in positive outcomes (based on evidence of benevolence, integrity, and ability), **prompting the act of giving control of something significant to you, to another party.**

As context and confidence in evidence changes, there is a corresponding change in trust.

Appropriate Trust

As **variations occur** in context, and evidence of benevolence, integrity, and ability, the **Trustor** will adjust the **level of trust** in *Trustee* to fit new circumstances.

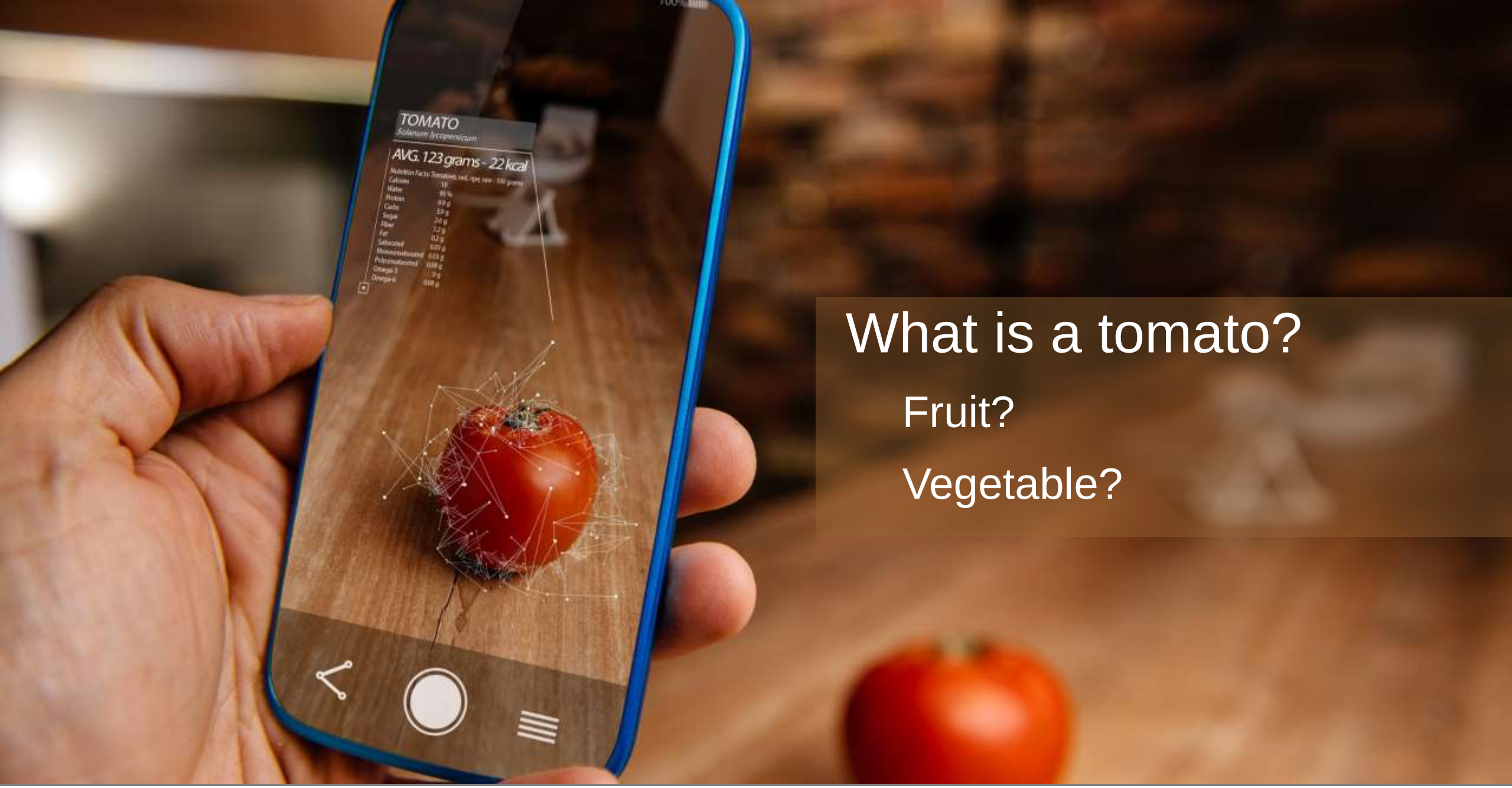


Kun Yu, Shlomo Berkovsky, Ronnie Taib, Dan Conway, Jianlong Zhou, and Fang Chen. 2017. User Trust Dynamics: An Investigation Driven by Differences in System Performance. IUI 2017 (March 2017), 307-317. DOI: http://dx.doi.org/10.1145/3025171.3025219

Is 100% Trust the Goal?

Semi-Autonomous Vehicles

Tesla Autopilot in Heavy LA Traffic by Scott Kubo https://youtu.be/m3-QzTFxoUg?t=14



What is a tomato?

Fruit?

Vegetable?

Trust is a Continuum

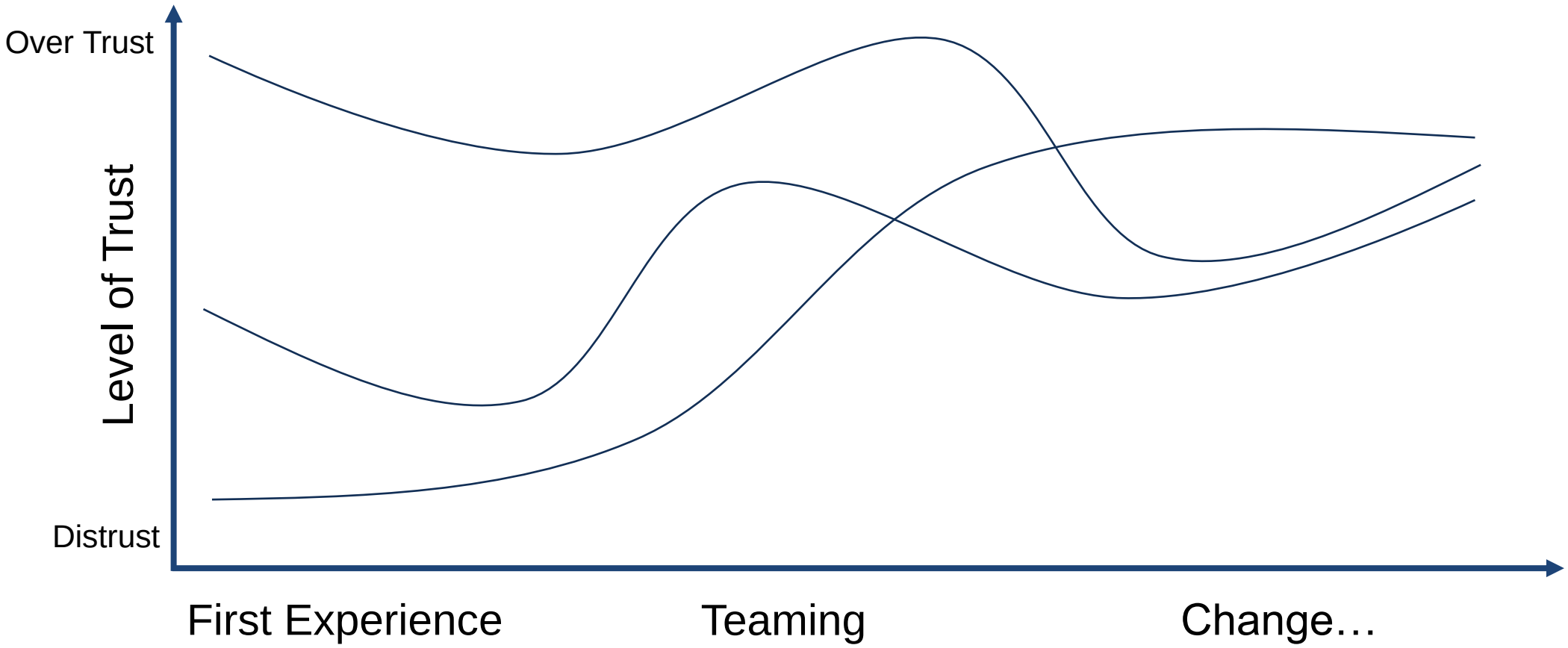
Distrust
Trust falling short of system capabilities - may lead to disuse.

Calibrated Trust
Trust matches system capabilities leading to appropriate use.

Over Trust
Trust exceeding system capabilities - may lead to misuse

Bobbie Seppelt and John Lee. 2012. Human Factors and Ergonomics in Automation Design. In Handbook of Human Factors and Ergonomics (Fourth Edition) Chapter 59. Wiley.
DOI: https://doi.org/10.1002/9781118131350.ch59

Trust Changes Over Time



Kun Yu, Shlomo Berkovsky, Ronnie Taib, Dan Conway, Jianlong Zhou, and Fang Chen. 2017. User Trust Dynamics: An Investigation Driven by Differences in System Performance. IUI 2017 (March 2017), 307-317. DOI: http://dx.doi.org/10.1145/3025171.3025219

Change Increases or Decreases Trust

Event-Driven

- Response to an interaction, transaction, service, or event

Time-Driven

- Response to periodic evidence (observations or recommendations)
- Lack of evidence can decay trust

Jia Guo and Ing-Ray Chen. 2015. A Classification of Trust Computation Models for Service-Oriented Internet of Things Systems. 2015 IEEE International Conference on Services Computing (2015), 324-331. DOI: https://doi.org/10.1109/SCC.2015.52
Kun Yu, Shlomo Berkovsky, Ronnie Taib, Dan Conway, Jianlong Zhou, and Fang Chen. 2017. User Trust Dynamics: An Investigation Driven by Differences in System Performance. IUI 2017 (March 2017), 307-317. DOI: http://dx.doi.org/10.1145/3025171.3025219

Change is Constant

Awareness of System Capabilities

Understanding of conditions, constraints

Experience with System

- Length of time
- Quality of experience

Transparency and usability of system

Additional Trust / Distrust Factors

Institutional, management

Social and relational

Previous experiences

What is Appropriate?

Can there be too much trust?

What is necessary?

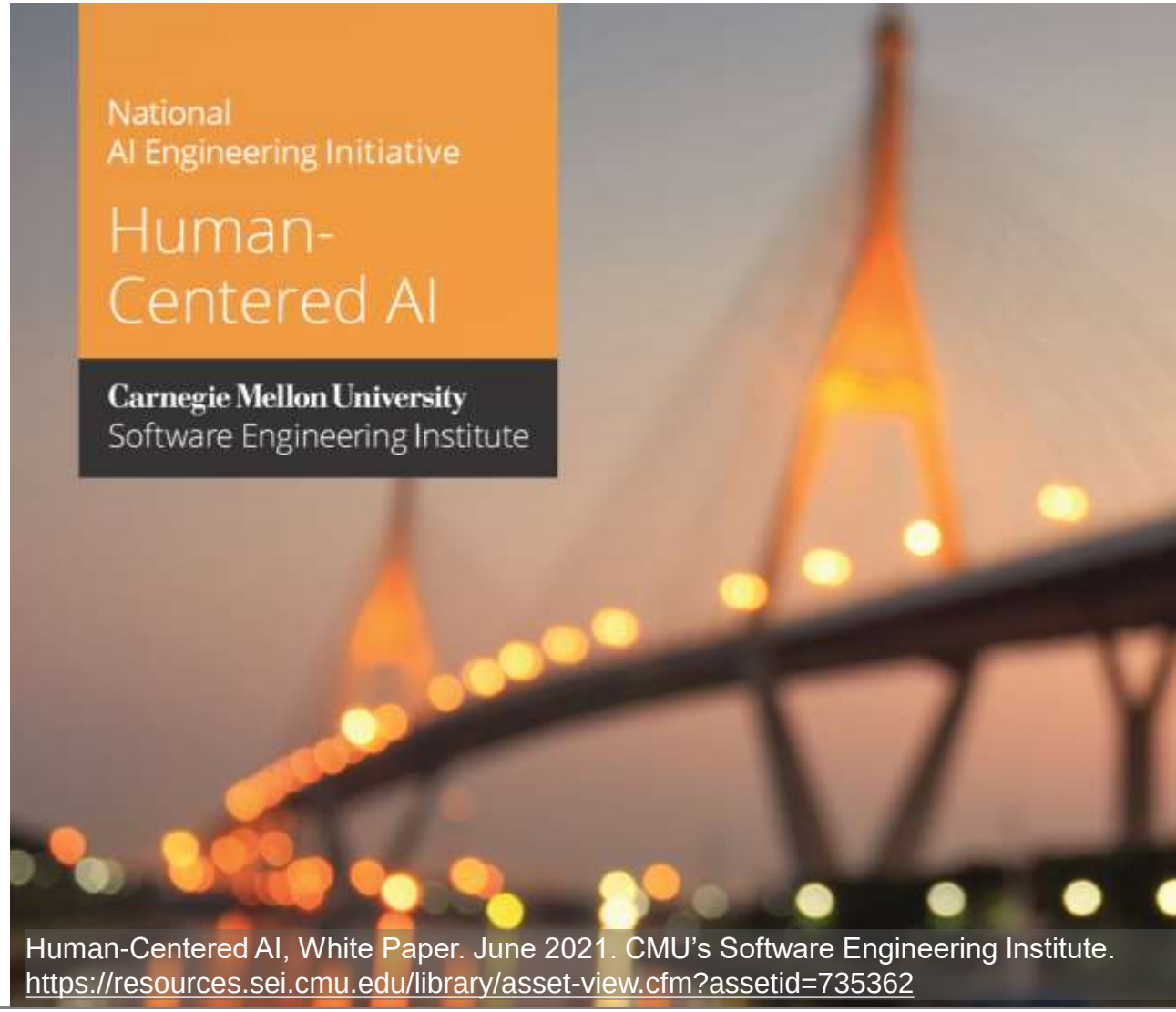
How do we communicate what is appropriate?

Supporting Appropriate Trust

Design to work with, and for, people

Minimize unintended consequences

- Research to understand context of use
- Design for purpose: Systems – not just tasks
- Test prototypes/products in environment



Human-Centered AI, White Paper. June 2021. CMU's Software Engineering Institute. https://resources.sei.cmu.edu/library/asset-view.cfm?assetid=735362

Speculation keeps people safe

Speculate and Design for the Worst Case

Don’t assume that only the average case will occur.

Be speculative about the worst case.

Create better decision-making tools that don't require unsupportable risk assessments.

N. G. Leveson. 2017. The Therac-25: 30 Years Later. In Computer, vol. 50, no. 11, (November 2017), 8-11. DOI: 10.1109/MC.2017.4041349

Activate Curiosity

UX research methods and activities to activate curiosity:

- Abusability Testing (Dan Brown)
- "Black Mirror" Episodes (Casey Fiesler)
  (inspired by British dystopian sci-fi tv series of same name)

Speculate about system misuse and abuse

- What are potential unintended/unwanted consequences?

Conversations for Understanding

Difficult Topics

- What do we value?
- Who could be hurt?
- What lines won't our AI cross?
- How are we shifting power?*
- How will we track our progress?

*"Don't ask if artificial intelligence is good or fair, ask how it shifts power." Pratyusha Kalluri. https://www.nature.com/articles/d41586-020-02003-2

New uncomfortable work

“*Be uncomfortable*”

- Laura Kalbag

Ethical design is not superficial.

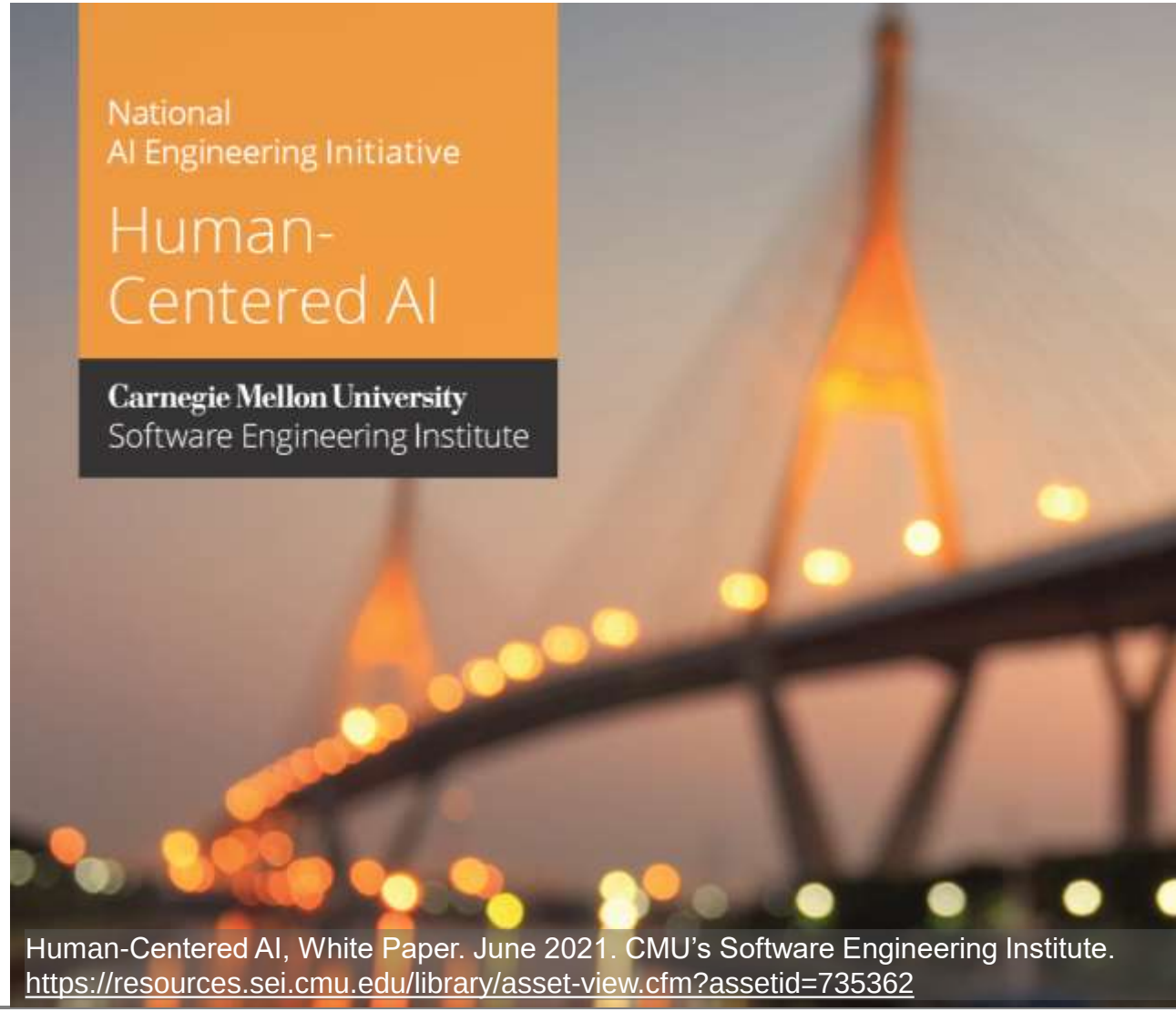
Transparency

System limitations

Boundaries
and unfamiliar scenarios

"Explainability" isn't magic.

Transparency isn't clarity.



Human-Centered AI, White Paper. June 2021. CMU's Software Engineering Institute. https://resources.sei.cmu.edu/library/asset-view.cfm?assetid=735362

Consider Time Cycles

- Length of time interactions occur
- Length varies
  - Very short and hectic
  - Longer and iterative
  - Affects interactions

Clear communication, negotiation, and coordination required

How IAs Can Shape the Future of Human-AI Collaboration
Presented on April 28-30, 2021 at the Information Architecture Conference (IAC21)
– Video https://www.designforcontext.com/ia-shaping-human-ai-collaboration

Make Systems Effective Team Players

Activities observable for fellow team players

Easy to direct

Capitalize on human strengths

- How observable is behaviour for human counterparts?
- How easily and efficiently it allows itself to be directed?
- Even (or especially) during busy, novel episodes?

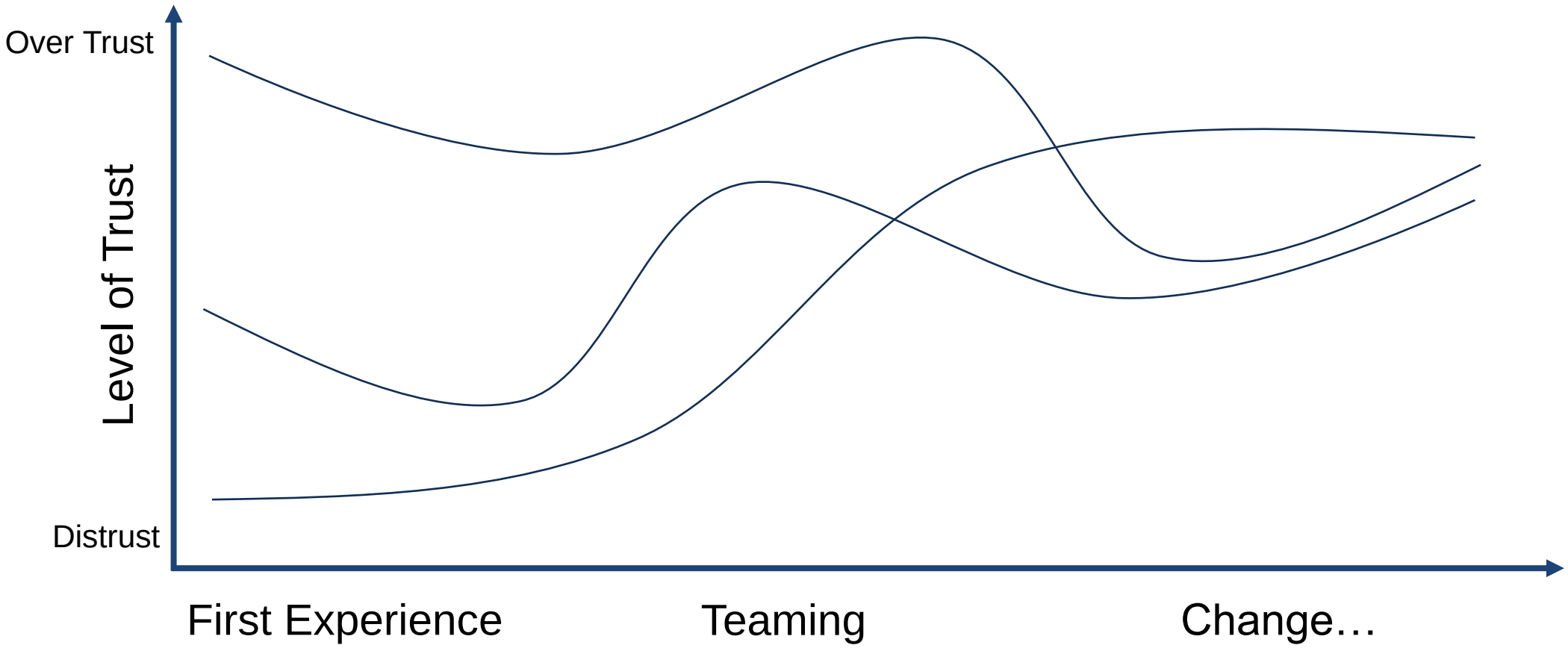
S. W. A. Dekker and D. D. Woods. 2002. MABA-MABA or Abracadabra? Progress on Human–Automation Co-ordination. Cognition Tech Work 4, (2002) 240–244. DOI: https://doi.org/10.1007/s101110200022 Note: MABA-MABA (Men-Are-Better-At/Machines-Are-Better-At lists)

Appropriate Trust

- Understand context and test in context
- Design for purpose: Systems
- Provide understandable evidence
- Complement human strengths
- Provide control to people

Jonathan Rotner, Ron Hodge and Lura Danley. 2020. AI Fails and How We can Learn from Them. The MITRE Corporation. July 2020. Case number 20-1365. https://sites.mitre.org/aifails/failure-to-launch/

Design for Appropriate Trust



Kun Yu, Shlomo Berkovsky, Ronnie Taib, Dan Conway, Jianlong Zhou, and Fang Chen. 2017. User Trust Dynamics: An Investigation Driven by Differences in System Performance. IUI 2017 (March 2017), 307-317. DOI: http://dx.doi.org/10.1145/3025171.3025219

Carol J. Smith

Twitter: @carologic

LinkedIn: https://www.linkedin.com/in/**caroljsmith**/

CMU's Software Engineering Institute,
Emerging Technology Center

Twitter: @sei_etc

Resources

Denise Rousseau, Sim Sitkin, Ronald Burt, and Colin Camerer. (1998). Not So Different After All: A Cross-discipline View of Trust. July 1988. Academy of Management Review. 23. 10.5465/AMR.1998.926617. DOI: 10.5465/AMR.1998.926617

Bobbie Seppelt and John Lee. 2012. Human Factors and Ergonomics in Automation Design. In Handbook of Human Factors and Ergonomics (Fourth Edition) Chapter 59. Wiley. DOI: https://doi.org/10.1002/9781118131350.ch59

Human-Centered AI, White Paper. June 2021. Carnegie Mellon University's Software Engineering Institute. Contributors: Hollen Barmer, Rachel Dzombak, Matt Gaston, Jay Palat, Frank Redner, Carol J. Smith. https://resources.sei.cmu.edu/library/asset-view.cfm?assetid=735362

Jia Guo and Ing-Ray Chen. 2015. A Classification of Trust Computation Models for Service-Oriented Internet of Things Systems. 2015 IEEE International Conference on Services Computing (2015), 324-331. DOI: https://doi.org/10.1109/SCC.2015.52

Jonathan Rotner, Ron Hodge and Lura Danley. 2020. AI Fails and How We can Learn from Them. The MITRE Corporation. July 2020. Case number 20-1365. https://sites.mitre.org/aifails/failure-to-launch/

Kun Yu, Shlomo Berkovsky, Ronnie Taib, Dan Conway, Jianlong Zhou, and Fang Chen. 2017. User Trust Dynamics: An Investigation Driven by Differences in System Performance. IUI 2017 (March 2017), 307-317. DOI: http://dx.doi.org/10.1145/3025171.3025219

Mary Cummings. 2004. Automation Bias in Intelligent Time Critical Decision Support Systems. AIAA 2004-6313. AIAA 1st Intelligent Systems Technical Conference. (September 2004). DOI: https://doi.org/10.2514/6.2004-6313

Neta Ezer, Sylvain Bruni, Yang Cai, Sam J. Hepenstal, Christopher A. Miller, and Dylan D. Schmorrow. 2019. Trust Engineering for Human-AI Teams. Proceedings of the Human Factors and Ergonomics Society Annual Meeting 63, no. 1 (November 2019): 322–26. https://doi.org/10.1177/1071181319631264.

N. G. Leveson. 2017. The Therac-25: 30 Years Later. In Computer, vol. 50, no. 11, (November 2017), 8-11. DOI: 10.1109/MC.2017.4041349

N. Leveson. 1995. Safeware: System Safety and Computers, Addison Wesley (1995).

Onur Asan, Alparslan Emrah Bayrak and Avishek Choudhury. 2020. Artificial Intelligence and Human Trust in Healthcare: Focus on Clinicians. J Med Internet Res (2020),Vol. 22,6:e15154. URL: https://www.jmir.org/2020/6/e15154 DOI: https://doi.org/10.2196/15154

Rose Challenger, Chris W. Clegg and Craig Shepherd. 2013. Function allocation in complex systems: reframing an old problem. Ergonomics, 56:7 (2017) 1051-1069. DOI: 10.1080/00140139.2013.790482

Automation Bias

Propensity for humans to **favor suggestions** from automated decision-making systems and to **ignore contradictory information** made without automation, even if it is correct.

Mary Cummings. 2004. Automation Bias in Intelligent Time Critical Decision Support Systems. AIAA 2004-6313. AIAA 1st Intelligent Systems Technical Conference. (September 2004). DOI: https://doi.org/10.2514/6.2004-6313

Optimal Trust

"Unnecessarily high trust in AI
may have catastrophic consequences,
especially in life-critical applications…

Optimal trust in which both humans
and AI each have **some level of skepticism**
regarding the other's decisions
since **both are capable of making mistakes**."

Onur Asan, Alparslan Emrah Bayrak and Avishek Choudhury. 2020. Artificial Intelligence and Human Trust in Healthcare: Focus on Clinicians. J Med Internet Res (2020), Vol. 22, 6:e15154. URL: https://www.jmir.org/2020/6/e15154 DOI: https://doi.org/10.2196/15154

Decision Making - Humans vs. Computers

Humans are better at:

- Perceiving patterns
- Improvising and using flexible procedures
- Recalling relevant facts at the appropriate time
- Reasoning inductively
- Exercising judgment

Computers are better at:

- Responding quickly to control tasks
- Repetitive and routine tasks
- Reasoning deductively
- Handling many complex tasks simultaneously

Mary Cummings. 2004. Automation Bias in Intelligent Time Critical Decision Support Systems. AIAA 2004-6313. AIAA 1st Intelligent Systems Technical Conference. (September 2004). DOI: https://doi.org/10.2514/6.2004-6313