AI: Tool or Teammate? And other Design Dilemmas in Naturalistic Settings

Panellists

- Robert Hoffman, Institute for Human & Machine Cognition
- Gary Klein, MacroCognition LLC
- Laura Militello, Applied Decision Science
- Emilie Roth, Roth Cognitive Engineering
- Cindy Dominguez, The MITRE Corporation
- Neelam Naikar, Defence Science and Technology Group

Co-chair

- Katie Ernst, Applied Decision Science

Overview

- Scene setting:
  - Rapid developments in AI with the potential to transform our work and lives.
- Intent:
  - Unique insights and challenges of decision-making involving AI in naturalistic environments
  - Perspectives of NDM experts
- House rules:
  - 10 minutes per panellist: 7 minutes for presentation + 3 minutes for burning questions from the audience
  - 20 minutes general discussion

(How) Can Machines be Team Players?

Robert R. Hoffman
IHMC

rhoffman@ihmc.us

Copyright R.R. Hoffman 2024. All rights reserved

2

PLUS	MINUS
Conforms to the ancient dream of machines that replicate humans	Anthropomorphism leads to misattributions

"The AI makes decisions …" but not in the way humans do.
"The AI recognized …" but not in the way humans do.
"The AI decided that …" but not in the way humans do.
"The AI sees … " but not in the way humans do.
"The AI thought that…" Um, no.

Copyright R.R. Hoffman 2024. All rights reserved

3

Anthropomorphism → Word hijacking

This seems to be a *mode de jour* computer science.

The borrowing of terms that share a psychological meaning in ordinary language and psychology

Copyright R.R. Hoffman 2024. All rights reserved

Example:
AI Systems *learn*

Ironically, they only do so in a Behaviorist sense
(change in behavior as a result of experience)

A change in a program’s state that leads to statistically better performance on similar tasks. The state change is some form of computer memory, e.g., weights assigned to alternative moves in a given situation and/or new or modified situation-action rules.

The “learning” process almost always involves weights, counts, probabilistic calculations, or recurrent sequences.

Copyright R.R. Hoffman 2024. All rights reserved

5

Human learning involves changes in everything: concepts,
inductions, insights, etc.

Machine learning necessitates the processing of upwards of
millions of examples.
Even an octopus can do one-trial learning.

Human learning is trans-modal (verbal, visual, auditory, gestural).
ML systems are (mostly) uni-modal.

Copyright R.R. Hoffman 2024. All rights reserved

6

The Anthropomorphism Breaks Down

When an AI system fails, it fails in ways humans wouldn't → *Swift Distrust*

"New York City chatbot advises small businesses to break the law"

"Chatbot accuses NBA player of vandalism spree after it misinterprets tweets"

"Auto company recalls autonomous vehicles after crash"

"AI adds Guess the Cause of Death *poll to an article"*

'Pregnant woman sues after AI accuses her of carjacking"

Copyright R.R. Hoffman 2024. All rights reserved

7

The “As-If” Dilemma

Computer scientists will push the envelope, attempting to make AI systems more capable ***and*** more human like

The better the machines are at emulation, the more necessary it is to describe their capacities and performance using psychological/mentalistic terminology.

But their human-like characteristics will always be “as-if.”

Copyright R.R. Hoffman 2024. All rights reserved

8

Will making them more human-like make them more useful, and more able to help humans be better humans?

“Spoken language is effective for human-human interaction but often has severe limitations when applied to human-computer interaction. . . . Speech is slow for presenting information, is transient and therefore difficult to review or edit, and interferes significantly with other cognitive tasks. However, speech has proved useful for store-and-forward messages, alerts in busy environments, and input-output for blind or motor-impaired users.”

“When operating a computer, most humans type (or move a mouse) and think, but find it more difficult to speak and think at the same time. Hand-eye coordination is accomplished in different brain structures, so typing or mouse movement can be performed in parallel with problem solving.”

Copyright R.R. Hoffman 2024. All rights reserved

Under what conditions can ML and DNN Systems serve us well?

ML and Deep Net systems do some things well.

But these involve fixed tasks, actually competitive, and neither fluid nor collaborative.

Copyright R.R. Hoffman 2024. All rights reserved

Conclusion

Machines are tools.

They cannot be team members in any genuine human sense.

Jargon, terminology, mis-attributions, anthropomorphisms, and metaphors can mislead, however serviceable they may be as hyperbolic clarion calls.

Copyright R.R. Hoffman 2024. All rights reserved

We all want better machines.

Ask not how a machine can be a teammate but ask instead when and in what way is human-machine interaction best patterned after human-human interaction.

Thank you!

Copyright R.R. Hoffman 2024. All rights reserved

References

Hoffman, R.R. (1980). Metaphor in science. In *Cognition and figurative language.* (pp. 393-423). Mahwah, NJ: Erlbaum.

Klein, G., Woods, D.D., Bradshaw, J.D., Hoffman, R.R. and Feltovich, P.J. (November/December 2004). Ten challenges for making automation a "team player" in joint human-agent activity. *IEEE Intelligent Systems*, pp. 91-95.

Clancey, W. J. (1987). *Knowledge-based tutoring*. Cambridge, MA: MIT Press.

Carroll, J. M. (1988). Mental models in human-computer interaction. In *Handbook of human-computer interaction* (pp. 45-65).

Drapkin, A. (2024). AI Gone wrong. *Tech.co.* [https://tech.co/news/list-ai-failures-mistakes-errors]

Forbus, K.D. and Feltovich, P.J. (Eds.) (2001), *Smart machines in education.* New York: AAAI Press.

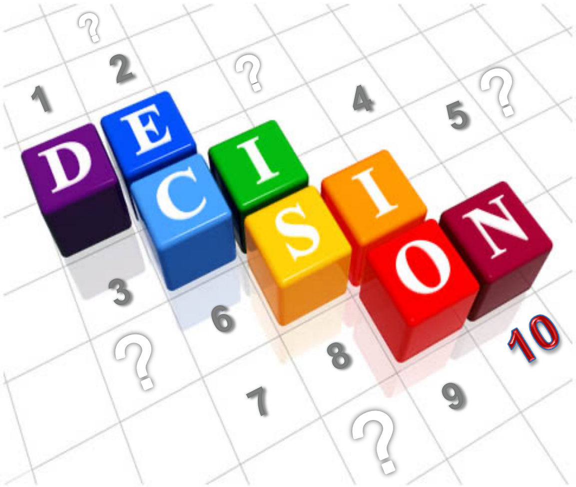
Nguyen, A., et al. (2014). Deep neural networks are easily fooled. ArXiv:1412.1897v1.

Ross, C. (2017). IBM pitched its Watson as a revolution in cancer care. It's nowhere close. https://www.statnews.com/2017/09/05/watson-ibm-cancer/

Shneiderman, B. (2000). The limits of speech recognition. *Communications of the ACM*, *9*, 63-65.

NDM-17
Auckland, NZ

AI: Tool or Teammate?



Gary Klein, Ph.D.
MacroCognition LLC

MacroCognition

An Apology

- **Klein G**, Woods DD, Bradshaw JM, Hoffman RR, & Feltovich PJ. (2004). Ten challenges for making automation a "team player" in joint human-agent activity. *IEEE Intelligent Systems*, 6, 91-95.
- Looks like we were advocating for automation (and AI) to be a team player
- Actually, just the reverse
- I am firmly in the “AI as Tool” camp
- How can we make AI a more useful tool?

AIQ: Artificial Intelligence Quotient

Helping people get smarter about the smart systems they work with

Designed for specific systems, not general issues about AI/ML

Direct the use of AI/ML more effectively, safely and securely

- AIQ is a toolkit of non-algorithmic methods for unpacking AI/ML systems and helping users gain appropriate trust
- Four primary methods
  - Self-explaining scorecard
  - Cognitive Tutorial
  - Collaborative XAI (CXAI)
  - Discovery Platform

AIQ Team

- Joseph Borders, ShadowBox LLC
- Gary Klein, ShadowBox LLC
- Robert Hoffman, Institute for Human Machine Cognition
- Shane Mueller, Michigan Technological University

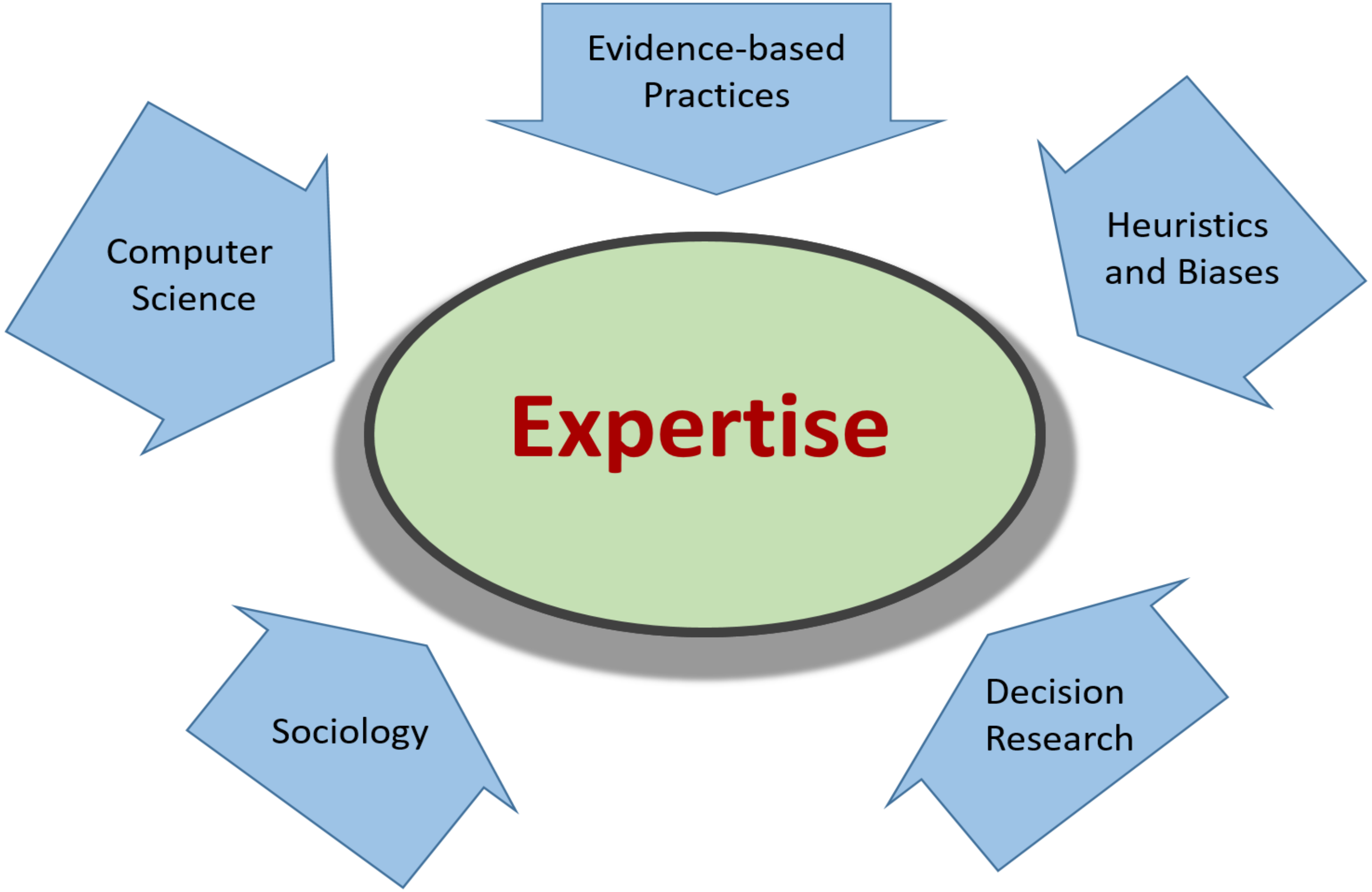
AIQ Applications

- U.S. Air Force — AI/ML language translation systems
- Raytheon — VQA system (AI labeling of photographs)
- Petrochemical industry — AIQ project about to start
- Pilot project on cognitive tutorial for Tesla FSD (Full Self Drive) mode
- Pilot project using CXAI to help drivers of autonomous modes detect errors better
- Data showing the cognitive tutorial format results in better learning than presentation of repeated examples
- AIQ precursor — Bayesian reasoning system
- AIQ precursor — Non-linear analysis program (users achieved a proficiency level in 1 ½ months that had taken 1 ½ years)

AI as a Tool

- How can we make AI a less dangerous tool?

The War on Expertise



Challenge

- NDM Community has an opportunity, perhaps a responsibility, to advocate for expertise in the workplace

AI Tools or Teammates Panel: Issues of Bias

Laura Militello
Applied Decision Science, LLC

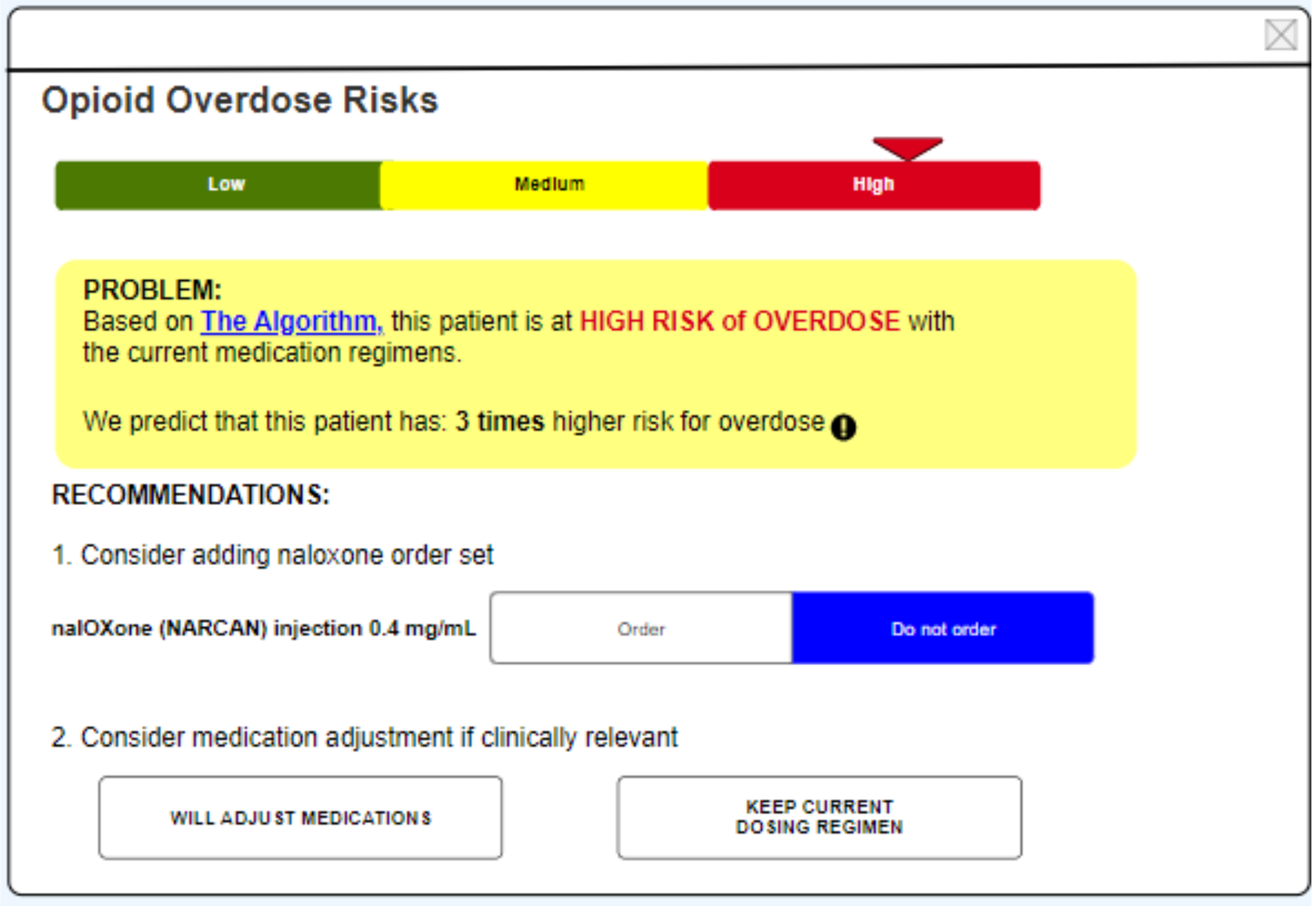
Issue: AI that introduces or exacerbates bias

- Algorithm developers rely on proxy data – there will always be bias
- The implications of the developer's assumptions are not always immediately apparent

DEMONSTRATE Algorithm

- Mines primary care data in the electronic health record.
- Looks at patients who have a previous opioid prescription.
- Finds patterns in the data of those that overdose
- Uses those patterns to predict the 3-month risk of overdose among the other patients.
- Displays updated risk scores every week
- Outperforms current strategies and measures that healthcare systems and payers (e.g., Medicare and Medicaid) are using to predict overdose risk

DEMONSTRATE Algorithm



DEMONSTRATE Algorithm

Elevated Risk of Opioid Overdose

Artificial Intelligence identified this patient based on a pattern of predictors in their health record.

1 in 125 patients identified by this alert will experience an opioid overdose (vs. 1 in 1000 baseline rate).

Recommendations

- **Support patient** by optimizing pain treatment and mental health.
- **Review & discuss risks** with patient. Why was this patient identified?
- **Offer naloxone** yearly (order not found in past year). How to talk about naloxone?

Order | Do Not Order — nalOXone (NARCAN) intranasal spray 4 mg

Override Reason

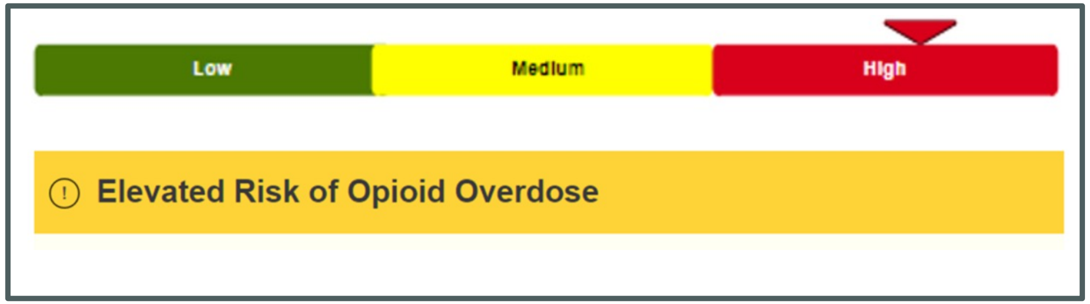
Patient has naloxone | Patient declined | Patient not present/not right time | Alert not relevant/other comment

ACCEPT

Supporting Army Aviation in Joint All Domain Operations

APPLIEDDECISIONSCIENCE

Conclusions



Bias is not just about data quality

- How findings are characterized
- How recommendations are displayed
- How limitations of AI are conceptualized

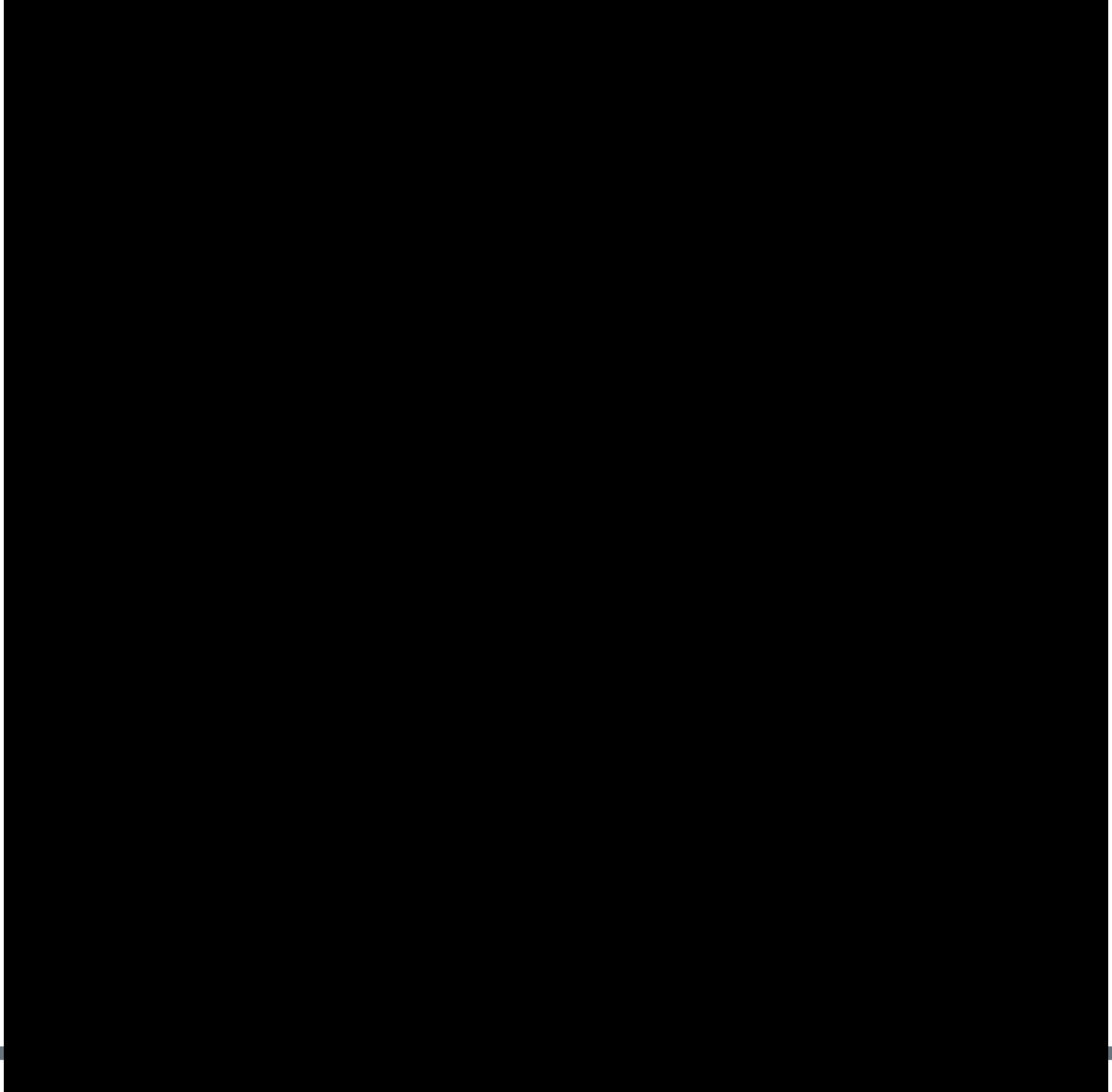
Recommendations

- **Support patient** by optimizing pain treatment and mental health.
- **Review & discuss risks** with patient. Why was this patient identified?
- **Offer naloxone** yearly (order not found in past year). How to talk about naloxone?

- How many people will be involved in this work? What are their roles?
- What will the technology really be capable of?
- What are the limitations of this technology?
- What will the technology know that humans won’t?
- What will the humans know that the technology won’t?

Thank you

Laura Militello
l.militello@applieddecisionscience.com



Teaming is Not the Enemy, it’s the Aspiration

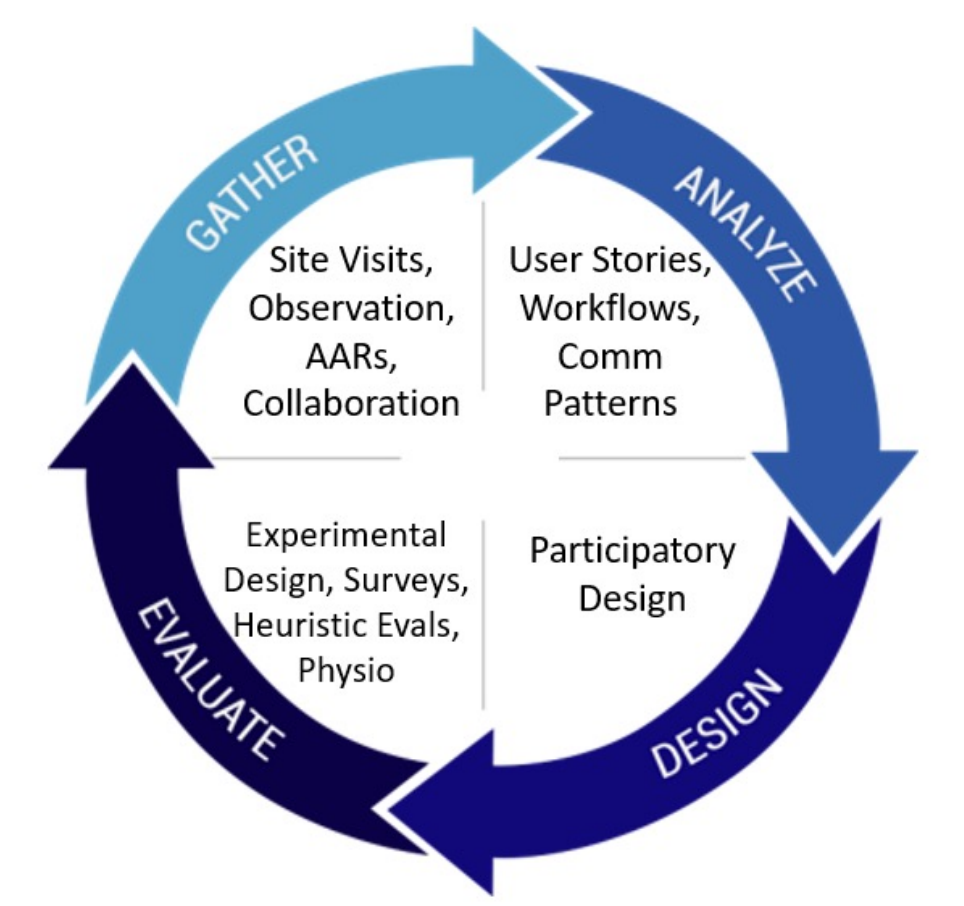
Cindy Dominguez

The MITRE Corporation

Human Machine Teaming Framework with CE Cycle

Source: MITRE Human-Machine Teaming Systems Engineering Guide, McDermott et al., 2018

Design Content			Design Process
Transparency	**Augmenting Cognition**	**Coordination**	**Design Specifics**
Observability Transparency into what an automation partner is doing relative to task progress	***Directing Attention*** Orient attention to critical problem features and cues	***Directability*** Humans can direct and redirect an automation partner's resources, activities, and priorities	***Information Presentation*** Format information to support understandability & simplicity
Predictability Future intentions and activities are observable & understandable	***Exploring the Solution Space*** Leverage multiple views, knowledge, and solutions to jointly understand the solution space	***Calibrated Trust*** Understand when and how much to trust automation partner	***Design Process*** Guidance on the systems engineering processes for HMT
	Adaptability Recognize and adapt fluidly to unexpected situations	***Common Ground*** Pertinent beliefs, assumptions, intentions are shared	



© 2022 THE MITRE CORPORATION. ALL RIGHTS RESERVED. APPROVED FOR PUBLIC RELEASE.

Australian Government
Defence

Tooling or teaming: A design perspective of human-AI sociotechnical systems

Neelam Naikar

Acknowledgments: Will Partridge, Glennn Moy, Hing-Wah Kwok, Emillie King, Ashleigh Brady

Bottom line

- We assume neither position
- A human-AI system is a ***sociotechnical*** system
- i.e., a ***network*** of interacting human and machine actors ***organised*** to achieve specific purposes and goals.

Example: Designing sociotechnical systems with multiple humans and machines

- Tasks in naturalistic settings rely on interactions among many people and teams:
  - e.g., emergency response, military operations, healthcare

Current design models and studies

- Focus on ***dyadic*** relationships:
  - Levels of automation taxonomy (e.g., Parasuraman et al., 2000)
  - Function allocation (e.g., Wright et al., 2000)
  - Adaptive and adaptable automation (e.g., Calhoun, 2022)

- Experimental studies with larger teams; 3 to 6 actors (e.g., Chen et al., 2018; Grimm et al., 2023):
  - Focus on dyadic relationships
  - Do not replicate complexity of real-world operations

Resulting designs

- Could lead to inefficiencies, reduced productivity, and compromised safety, as evident in field studies (e.g., Salwei et al., 2022):
  - Information sharing and dissemination
  - Ability to share tasks, provide assistance or oversight, and communicate seamlessly
- Consequently, problems with:
  - Coping with changing needs of complex operations, including unexpected events
  - Integration of AI technologies into larger team structures

Our work (Naikar, Brady, Moy & Kwok, 2023)

- Design models and approaches for integrating AI into larger teams
- How do human teams operate in naturalistic settings?
- Not necessarily designing machines to be like human teammates
- Instead, designing machines to be effective in “polyadic” organisational structures
  - Irrespective of how the AI is conceptualised.

Our view—of good design practice

- Roles and relationships of actors:
  - Must be guided by the features of the environment and context, demands of the work, and actors’ capabilities and limitations.
  - Should emerge as an ***output*** of the design process, rather than being assumed a priori, as an ***input*** to the analysis
- Consistent with earlier work on designing human teams for complex systems

Thank you