

DePaul University
HCD 590: Advanced Topics
in Human Centered Design,
Designing Ethical AI, Spring 2021

Designing Ethical AI: Exercises for Research

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

Copyright 2021 Carnegie Mellon University.

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.

The view, opinions, and/or findings contained in this material are those of the author(s) and should not be construed as an official Government position, policy, or decision, unless designated by other documentation.

References herein to any specific commercial product, process, or service by trade name, trade mark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by Carnegie Mellon University or its Software Engineering Institute.

NO WARRANTY. THIS CARNEGIE MELLON UNIVERSITY AND SOFTWARE ENGINEERING INSTITUTE MATERIAL IS FURNISHED ON AN "AS-IS" BASIS. CARNEGIE MELLON UNIVERSITY MAKES NO WARRANTIES OF ANY KIND, EITHER EXPRESSED OR IMPLIED, AS TO ANY MATTER INCLUDING, BUT NOT LIMITED TO, WARRANTY OF FITNESS FOR PURPOSE OR MERCHANTABILITY, EXCLUSIVITY, OR RESULTS OBTAINED FROM USE OF THE MATERIAL. CARNEGIE MELLON UNIVERSITY DOES NOT MAKE ANY WARRANTY OF ANY KIND WITH RESPECT TO FREEDOM FROM PATENT, TRADEMARK, OR COPYRIGHT INFRINGEMENT.

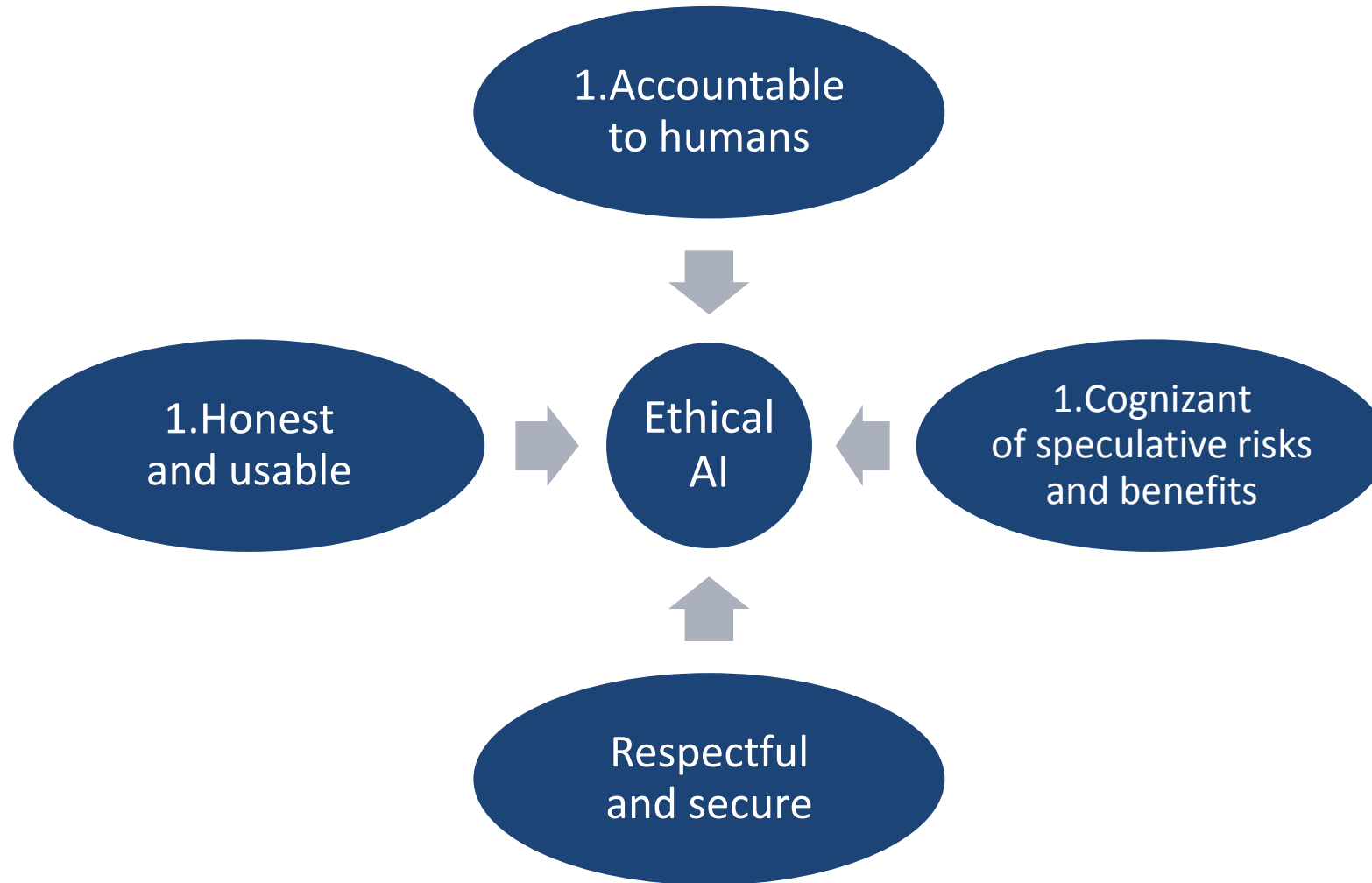
[DISTRIBUTION STATEMENT A] This material has been approved for public release and unlimited distribution. Please see Copyright notice for non-US Government use and distribution.

This material may be reproduced in its entirety, without modification, and freely distributed in written or electronic form without requesting formal permission. Permission is required for any other use. Requests for permission should be directed to the Software Engineering Institute at permission@sei.cmu.edu.

Carnegie Mellon® is registered in the U.S. Patent and Trademark Office by Carnegie Mellon University.

DM21-0454

Making responsible, ethical AI



Designing Trustworthy AI for Human-Machine Teaming. By Carol Smith. Software Engineering Institute Blog. March 9, 2020.
https://insights.sei.cmu.edu/sei_blog/2020/03/designing-trustworthy-ai-for-human-machine-teaming.html

Speculation keeps people safe



Conversations for Understanding

Harms modeling - difficult Topics

- What do we value?
- Who could be hurt?
- What lines won't our system 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>

Photo by Pam Sharpe https://unsplash.com/@msgrace?utm_source=unsplash&utm_medium=referral&utm_content=creditCopyText On Unsplash - https://unsplash.com/s/photos/business-woman-smiling?utm_source=unsplash&utm_medium=referral&utm_content=creditCopyText



Speculative exercises for Ethical AI

- Abusability Testing ([Dan Brown](#))
- “Black Mirror” Episodes ([Casey Fiesler](#))
(inspired by British dystopian sci-fi tv series of same name)

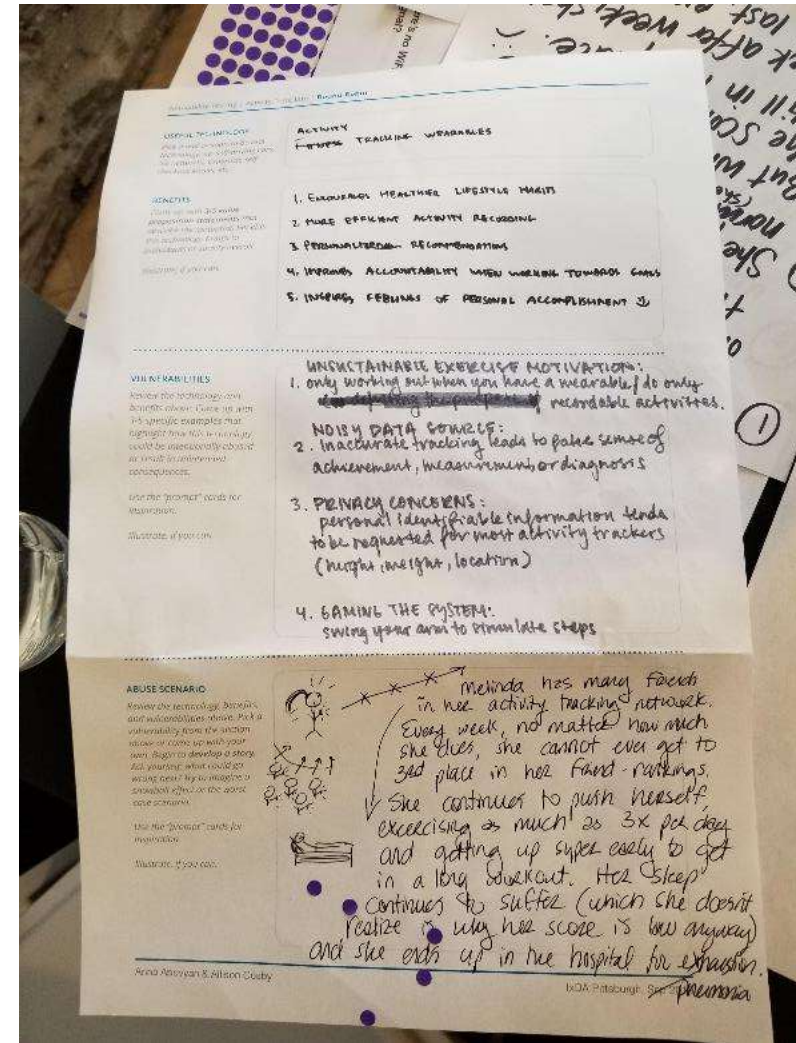
Speculate about system misuse and abuse

- What are potential unintended/unwanted consequences?

More methods to “Outsmart Your Own Biases.”: <https://hbr.org/2015/05/outsmart-your-own-biases>
Implicit Association Test (IAT): <https://implicit.harvard.edu/implicit/takeatest.html>

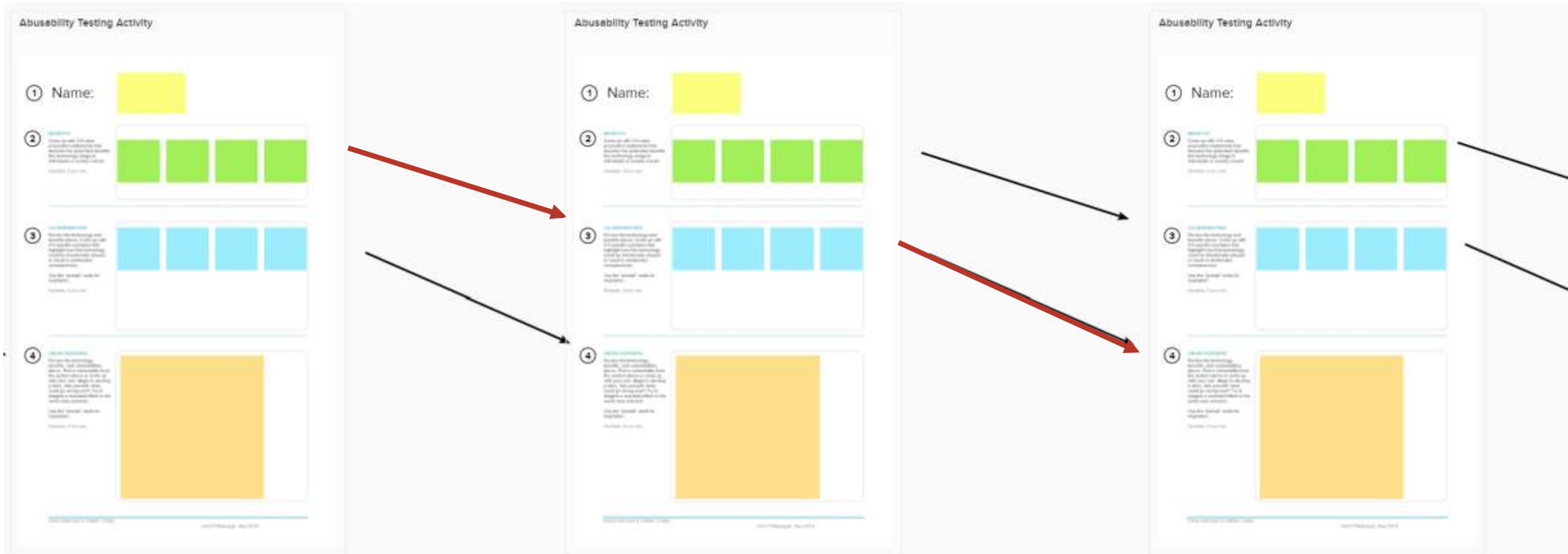
Abusability Activity

- Speculate about misuse and abuse
- Potential severe abuse and consequences
- Perspective of people in frequently marginalized groups



Abusability Activity (in Mural)

- Round Robins initially



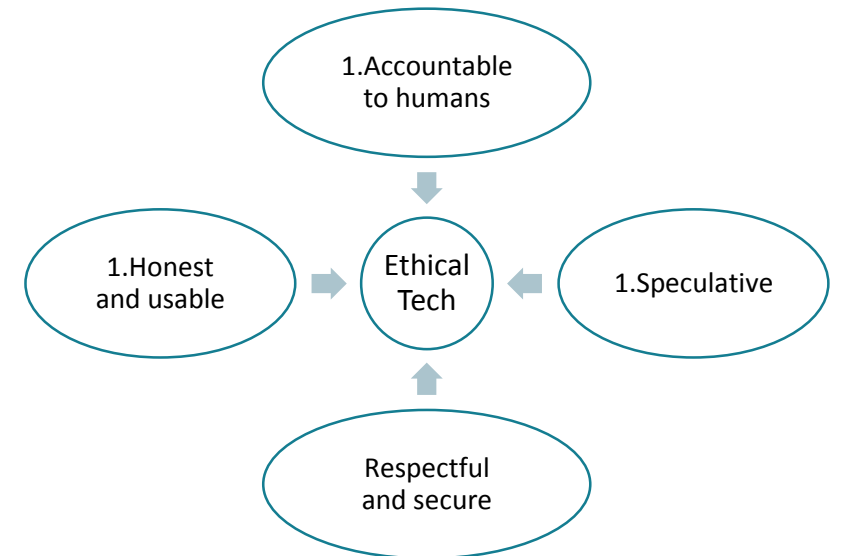
We aren't perfect, tech won't be perfect

Empower diverse teams, inclusive environments

Adopt technical ethics

Encourage deep conversations (Checklist)

Activate curiosity; be speculative; imaginative



**Evangelize
for human values**

**Ethical.
Transparent. Fair.**

Carol J. Smith

Twitter: @carologic

LinkedIn: <https://www.linkedin.com/in/caroljsmith/>

CMU's Software Engineering Institute,
Emerging Technology Center

Twitter: @sei_etc