

Responsible AI

DoD's Ethical Principles for AI

Equitable

Software Engineering Institute
Carnegie Mellon University
Pittsburgh, PA 15213

Document Markings

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.

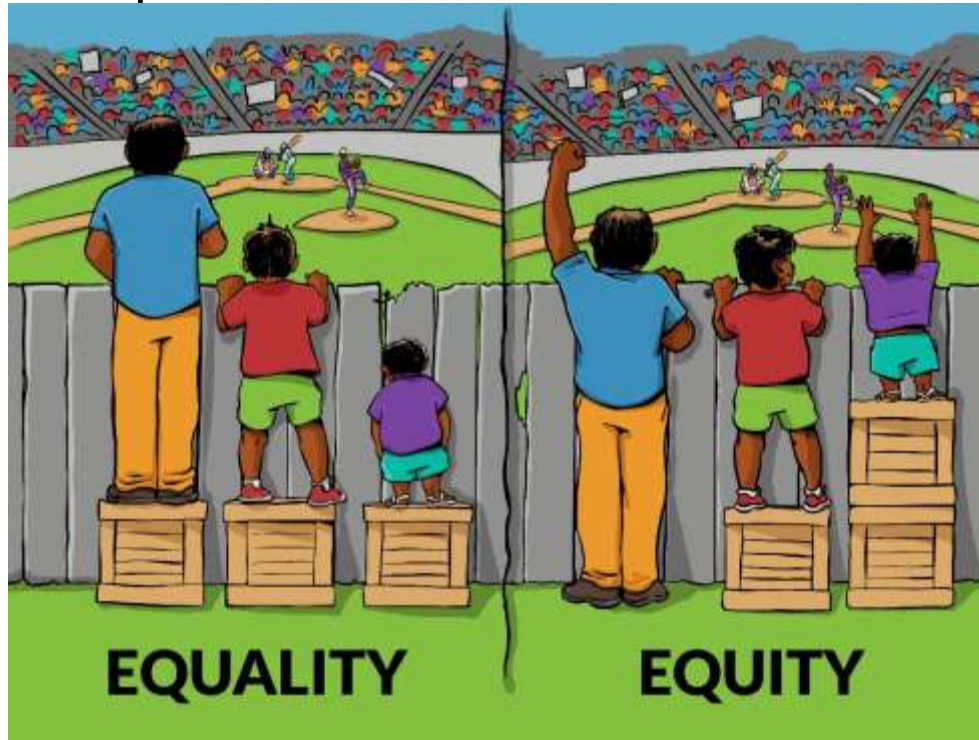
Carnegie Mellon® is registered in the U.S. Patent and Trademark Office by Carnegie Mellon University.

DM21-1129

Equitable

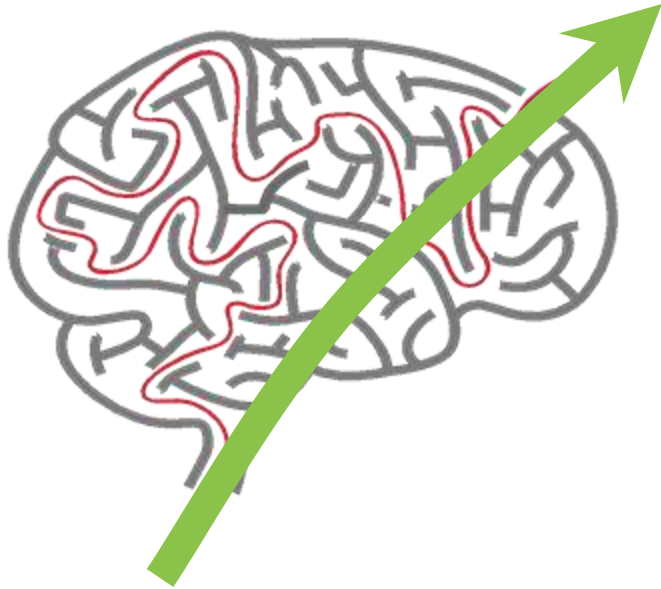
Equitable

Take deliberate steps to minimize unintended bias in AI capabilities.



Equality vs. Equity - by the Interaction Institute for Social Change | Artist: Angus Maguire. Image Found: interactioninstitute.org and madewithangus.com.

To be biased, is to be human



Bias are shortcuts, to avoid risk and simplify problems.

Not inherently bad, may be misapplied

Implicit = invisible

Not necessarily in sync with our conscious beliefs

Can be managed and changed

Talk about biases in non-threatening, productive ways

Karen Bachmann (with Carol Smith) Abstractions 2019. Designing More Ethical and Unbiased Experiences. August 21, 2019.

Biased due to...

Social class

Resource availability

Education

Race, gender, sexuality

Culture, theology, tradition

More...

Bias is everywhere

“We often have no way of knowing when and why people are biased.”

- Sandra Wachter

Data ethics researcher

Univ. of Oxford and Alan Turing Institute

All systems have some form of bias

Complete objectivity is misleading.

Bias can have purpose and can be helpful.

The goal is to reduce unintended and/or harmful bias.

Bias enters via...

Data source

- All data created by humans
- Curated and organized based on bias of researcher/organization

Algorithms

- Sentencing and crime prediction – racial bias
- Facial recognition – race and gender bias

Training

- Purposeful – towards a particular outcome
- Inadvertent due to content

Data can bias training

Training Set



Data Encountered



Use case courtesy of Dr. Eric Heim, CMU SEI
<https://resources.sei.cmu.edu/library/author.cfm?authorid=542374>

Mitigations to bias

Review data

- Who created/organized/curated it?
- Respected experts in the industry?
- What bias does it already have?

Consider system training

- Diverse team, mindful of bias

Speculate about system misuse and abuse

- What are potential unintended/unwanted consequences?

Diversity is...

Gender, race, culture,
education (school, program, etc.),
thinking process,
disability status,
and more...



Photo by Christina @ wocintechchat.com on Unsplash

Great minds think different

Diverse teams mitigate bias

Focus more on facts

Process facts more carefully

More innovative

“...become more aware of their own potential biases”

David Rock, Heidi Grant. 2019. Why Diverse Teams Are Smarter. (4 November, 2019). Retrieved September 13, 2019.
from: <https://hbr.org/2016/11/why-diverse-teams-are-smarter>

Equitable – Review

Take deliberate steps to minimize unintended bias in AI capabilities.
Everyone is biased, acknowledge and change bias.
Unwanted bias may be in data, algorithms, training.
Mitigate bias and build diverse teams.

**Empower diverse teams in
inclusive environments.**

**Encourage deep conversations, speculation,
and imaginative thinking.**

Alex Van Deusen
Design Researcher
arvandeusen@sei.cmu.edu

Carol Smith
Sr. Research Scientist
cjsmith@sei.cmu.edu

Rachel Dzombak
Digital Transformation Lead
rdzombak@sei.cmu.edu