

ACM  
Distinguished Speaker Program

# AI and Machine Learning Demystified

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

Twitter: @carologic @SEI\_CMU\_AI

Software Engineering Institute  
Carnegie Mellon University  
Pittsburgh, PA 15213

# Copyright Statement

Copyright 2022 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.**

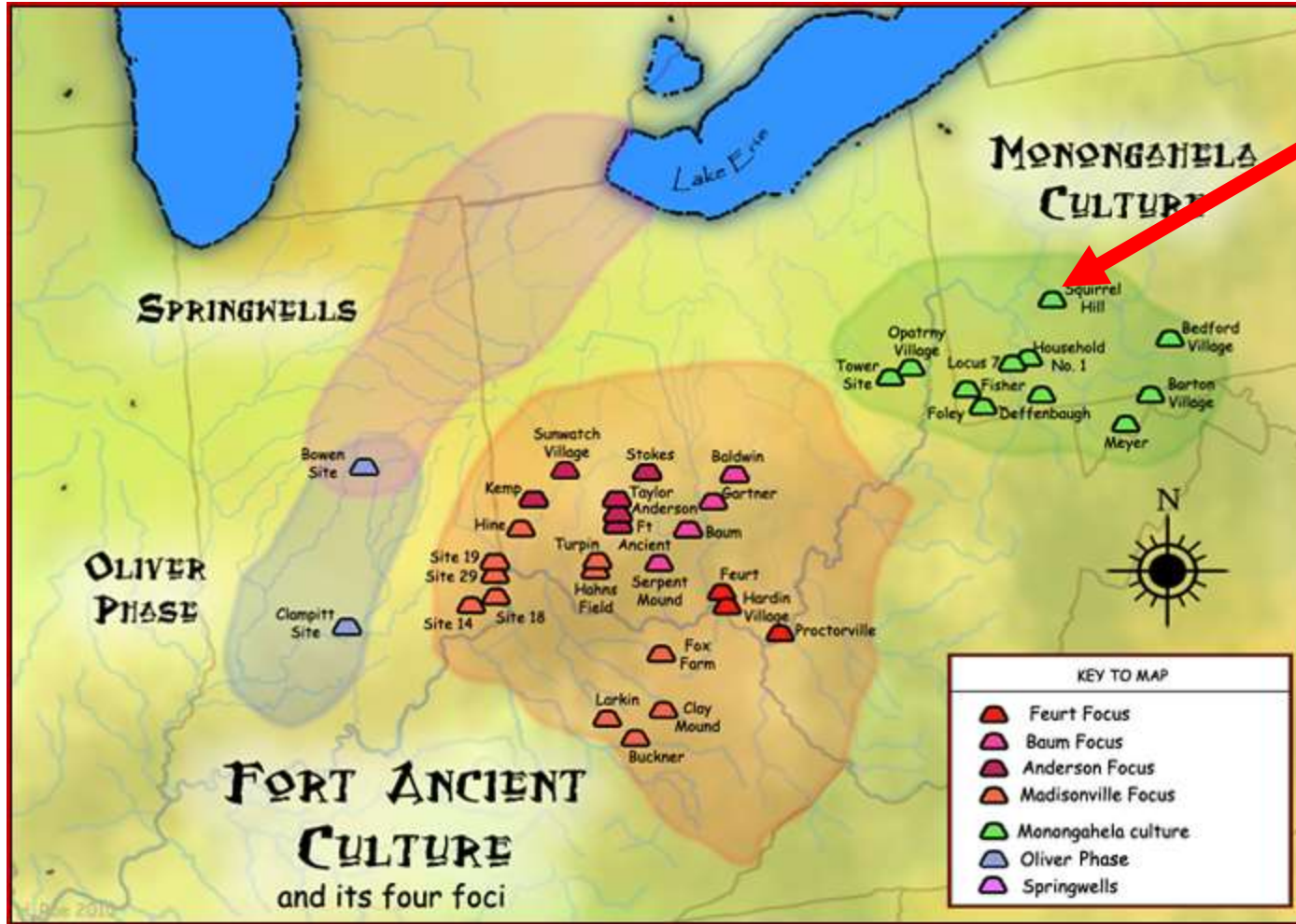
[DISTRIBUTION STATEMENT A] This material has been approved for public release and unlimited distribution. Please see Copyright notice for non-US Government use and distribution.

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](mailto:permission@sei.cmu.edu).

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

DM22-0562

# Acknowledgement: The Land I Speak On



Map by Herb Roe via Wikipedia [https://en.wikipedia.org/wiki/Monongahela\\_culture](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.

# About ACM

- ACM, the Association for Computing Machinery ([www.acm.org](http://www.acm.org)), is the premier global community of computing professionals and students with nearly 100,000 members in more than 170 countries interacting with more than 2 million computing professionals worldwide.
- OUR MISSION: We help computing professionals to be their best and most creative. We connect them to their peers, to what the latest developments, and inspire them to advance the profession and make a positive impact on society.
- OUR VISION: We see a world where computing helps solve tomorrow's problems – where we use our knowledge and skills to advance the computing profession and make a positive social impact throughout the world.
- I am proud to be an ACM Member.



# The Distinguished Speakers Program is made possible by



**Association for  
Computing Machinery**

*Advancing Computing as a Science & Profession*

For additional information, please visit <http://dsp.acm.org/>

# What is artificial intelligence?



AI systems can

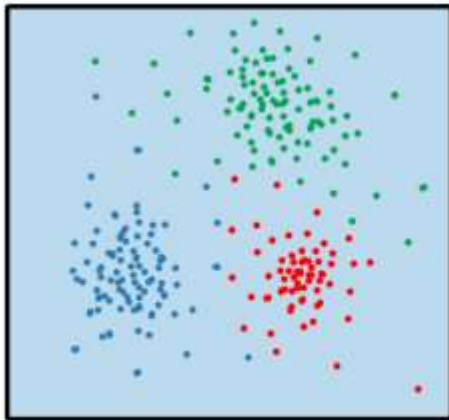
- recognize patterns
- create predictions
- make decisions, and/or
- generate new content.

Without being explicitly programmed to do so.

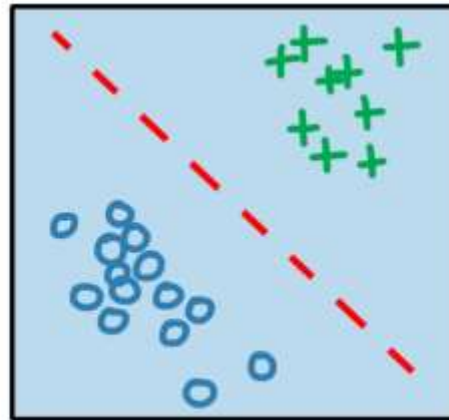
# Artificial Intelligence

## machine learning

unsupervised learning



supervised learning



reinforcement learning

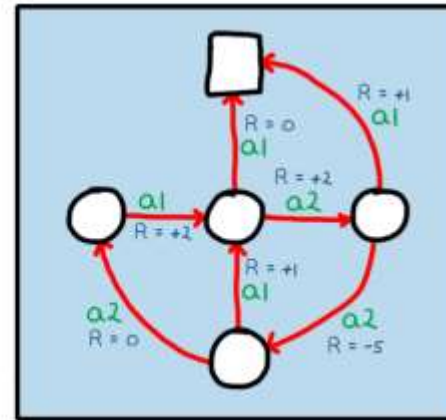
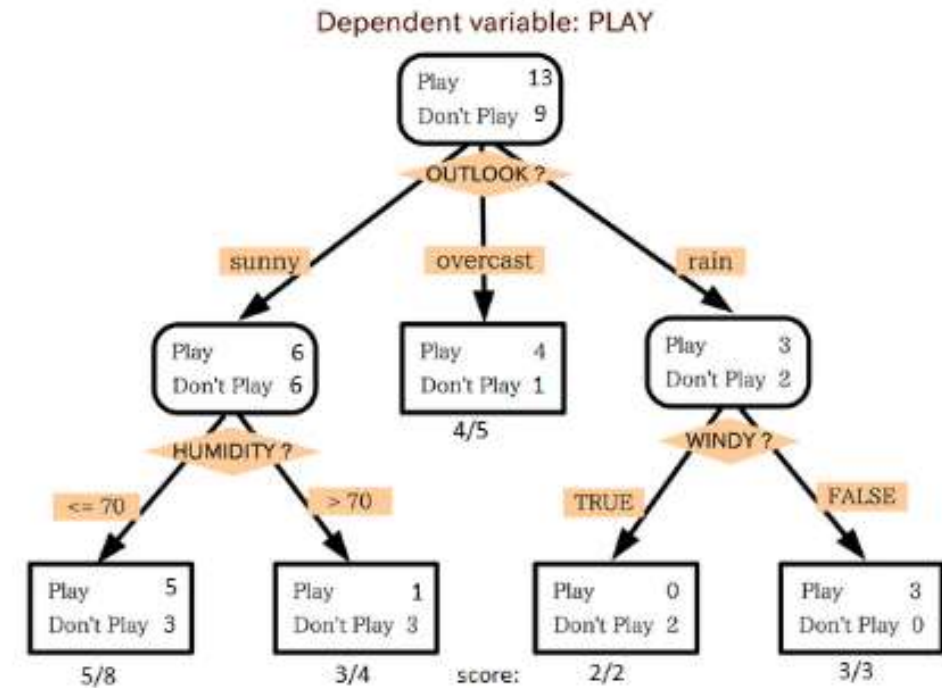


Image © 1994-2022 The MathWorks, Inc.

## Deep Learning, Neural Networks

# AI / Machine Learning

- Algorithms
  - math + programming
- Model (AI)
- Algorithms + data
- Know ONLY what taught
- Control ONLY given control of
- Aware of nuances
  - can continue to learn



source: [statsexchange](https://www.statsexchange.com)

<https://www.analyticsvidhya.com/blog/2015/08/common-machine-learning-algorithms/>

# Taxonomies and Ontologies coming to life (NOT like humans learn)



Photo:  
[https://commons.wikimedia.org/wiki/File:Baby\\_Boy\\_Oliver.jpg](https://commons.wikimedia.org/wiki/File:Baby_Boy_Oliver.jpg)

# AI is NOT sentient

## Not unknowable

# Never Enough Time

Physician: ~90 hours reading  
a week\*

AI could bring that information  
to the physician

Enabling more  
evidence-based decisions

Alper, Brian S. et al. "How Much Effort Is Needed to Keep up with the Literature Relevant for Primary Care?"  
Journal of the Medical Library Association 92.4 (2004): 429–437. Print.  
<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC521514/>



# Transfer human concepts and relationships

## Number Five “Needs Input”



Photo by sunlightfoundation  
<https://www.flickr.com/photos/sunlightfoundation/2385174105>

# Supervised (by a human) machine learning

# Enormous amount of work

# Dependent on Experts

Data scientists

Subject matter experts (SME's) available

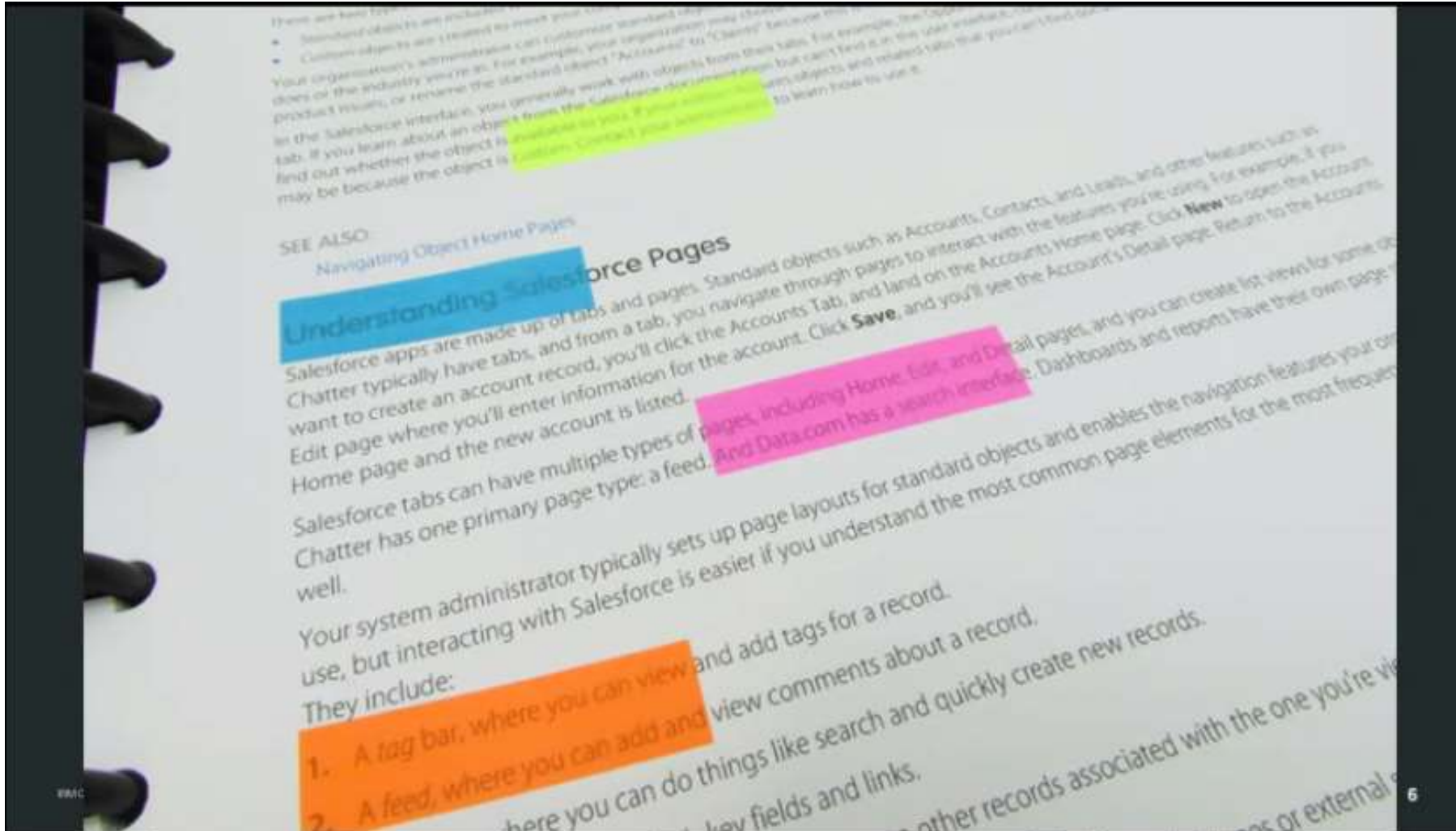
- Lawyers
- Machinists
- Insurance adjusters
- Physicians

Not just experts in machine learning



Photo by Pam Sharpe [https://unsplash.com/@msgrace?utm\\_source=unsplash&utm\\_medium=referral&utm\\_content=creditCopyText](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](https://unsplash.com/s/photos/business-woman-smiling?utm_source=unsplash&utm_medium=referral&utm_content=creditCopyText)

# Experts Annotate Content



# Entity Type

High level concepts applied to a mention

**PERSON**

**Amanda**

**Amanda Tomlin**

**She**

# Define Entity Types

**PERSON**

**ORGANIZATION**

**TIME**

**Amanda** works at **Carnegie Mellon University**.

**She** has worked for the **university** for **2 years**.

# Define Relationships

employedBy

# Relation type

employedBy

**Amanda** works at **Carnegie Mellon University**.

employedBy

**She** has worked for the **university** for **2 years**.

**Continue:  
create dictionaries,  
rules and more...**

# Creating ML requires

- Data – curated, perhaps annotated
- Algorithms (models)
- Train and iterate
- Repetition with new content
- Time - weeks to months to start, ongoing
- Continuous critical oversight

**AI is as imperfect  
as the humans making it**

# Training Set and Use

## Training data



## Data encountered



Use case courtesy of Dr. Eric Heim, CMU SEI  
<https://resources.sei.cmu.edu/library/author.cfm?authorid=542>

# Only know what taught

## Training data



Unrepresentative  
or incomplete training data

## Data encountered



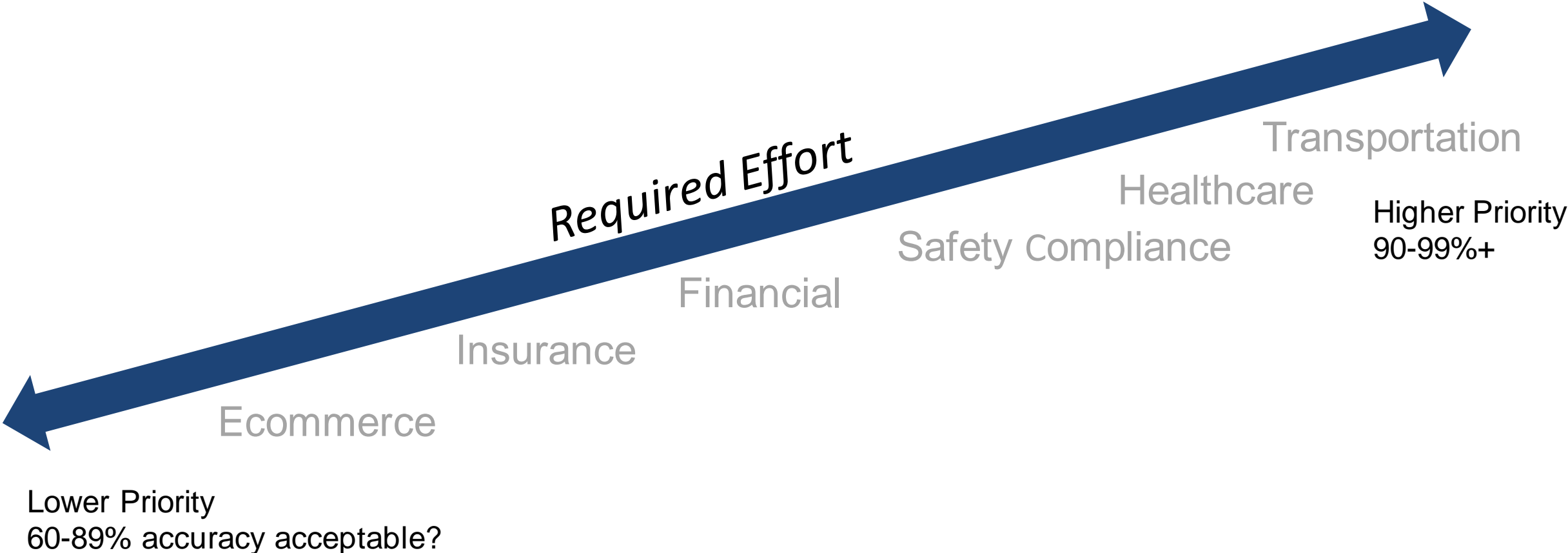
Unlikely to recognize

**Only as good as data  
and time spent improving it.**

**Biased based  
on what taught.**

# Concern varies across industries

## Accuracy is not always the best measure!



# Use Cases

# Consider for each situation

Knowledge needed?

Ethical considerations?

# Strategic Games

1997 Chess, IBM

2016 Go, Google

Knowledge?

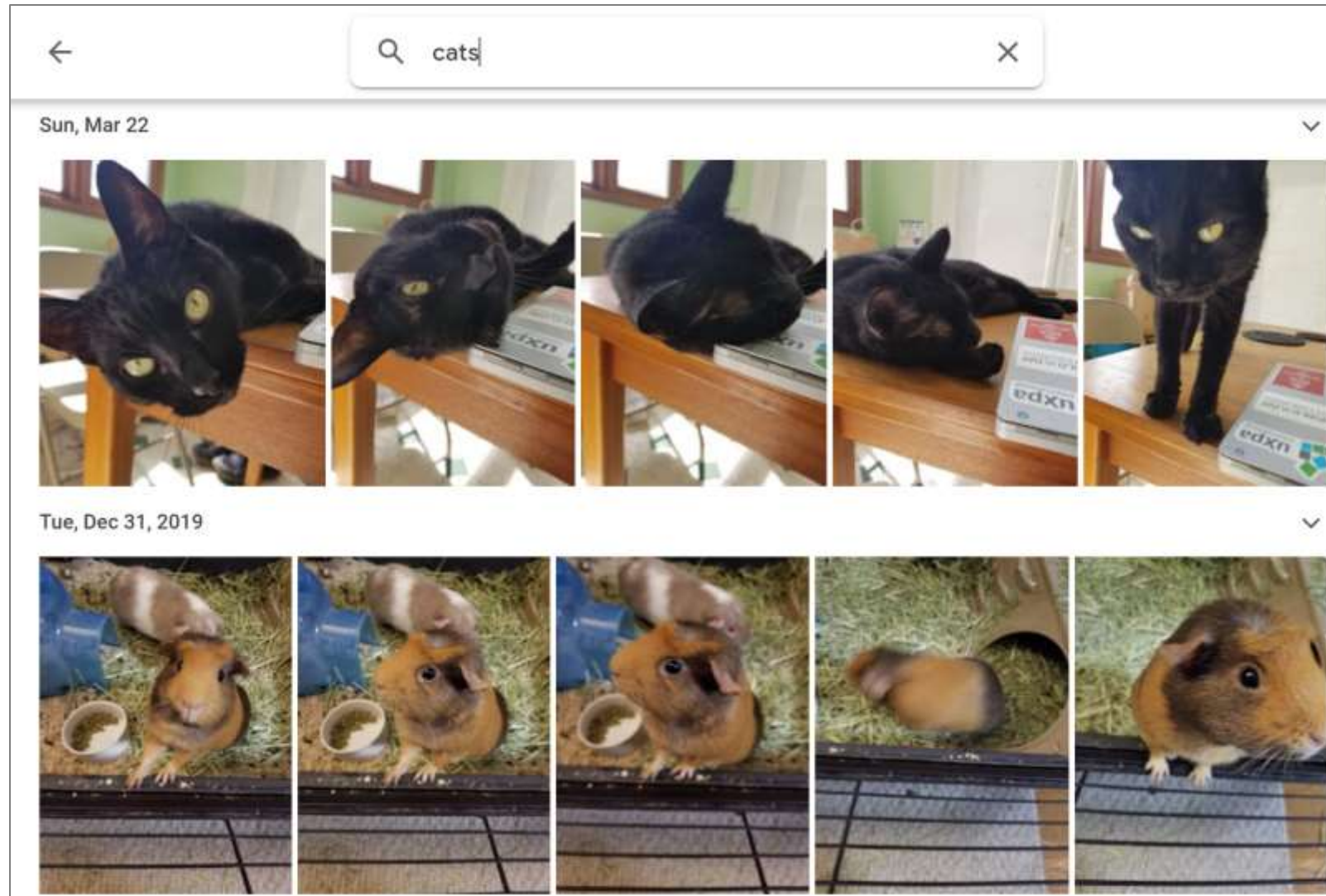
Ethics?



Floor goban, 2007, By Goban1

<https://commons.wikimedia.org/wiki/File:FloorGoban.JPG>

# Image Recognition – Google Photos



Carol's search for "cats" on her Google Photos account.

# Sound recognition: Labeling of birdsongs



“Comparison of machine learning methods applied to birdsong element classification”

by David Nicholson. Proceedings of the 15<sup>th</sup> Python in Science Conference (SCIPY 2016). [http://conference.scipy.org/proceedings/scipy2016/pdfs/david\\_nicholson.pdf](http://conference.scipy.org/proceedings/scipy2016/pdfs/david_nicholson.pdf)  
Photo by Gallo71 (Own work) [Public domain], via Wikimedia Commons <https://commons.wikimedia.org/wiki/File%3ARbruni.JPG>

# Listening and understanding human speech

## Mapping Q & A + AI

- Expected language
- Appropriate automated responses
- When to escalate?
  - Searches on self harm?
  - What else?



**Hi, I'm Woebot**



Images: <https://www.pexels.com/photo/close-up-of-mobile-phone-248512/>  
<https://www.amazon.com/Amazon-Echo-Bluetooth-Speaker-with-WiFi-Alexa/dp/B00X4WHP5E>  
<https://www.ibm.com/watson/developercloud/doc/conversation/index.html>

# Decision Making: Autonomous vehicles



<https://www.uber.com/info/atg/>

# Bias in AI

# Bias is natural



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

# All systems will have some form of bias

Complete objectivity is misleading.

Bias can have purpose and can be helpful.

We must ensure we

- identify and understand bias
- reduce unintended and/or harmful bias.

# Joy Buolamwini, Algorithmic Justice League

“Data is a function of our history...  
The past dwells within our algorithms...  
Showing us the inequalities that have always been there.”

## Coded Gaze

Photo: Joy Buolamwini on The Open Mind: Algorithmic Justice League.  
Jan 12, 2019. <https://www.youtube.com/watch?v=hwHnXdoSSFY>

THE  
**OPEN MIND**



# Our responsibility is to keep people safe



# Brakes, Back doors and Buffers

Responsible, intentional design

- How are we keeping people safe?
- When unintended consequences arise, how do we deal with them?

Make a plan



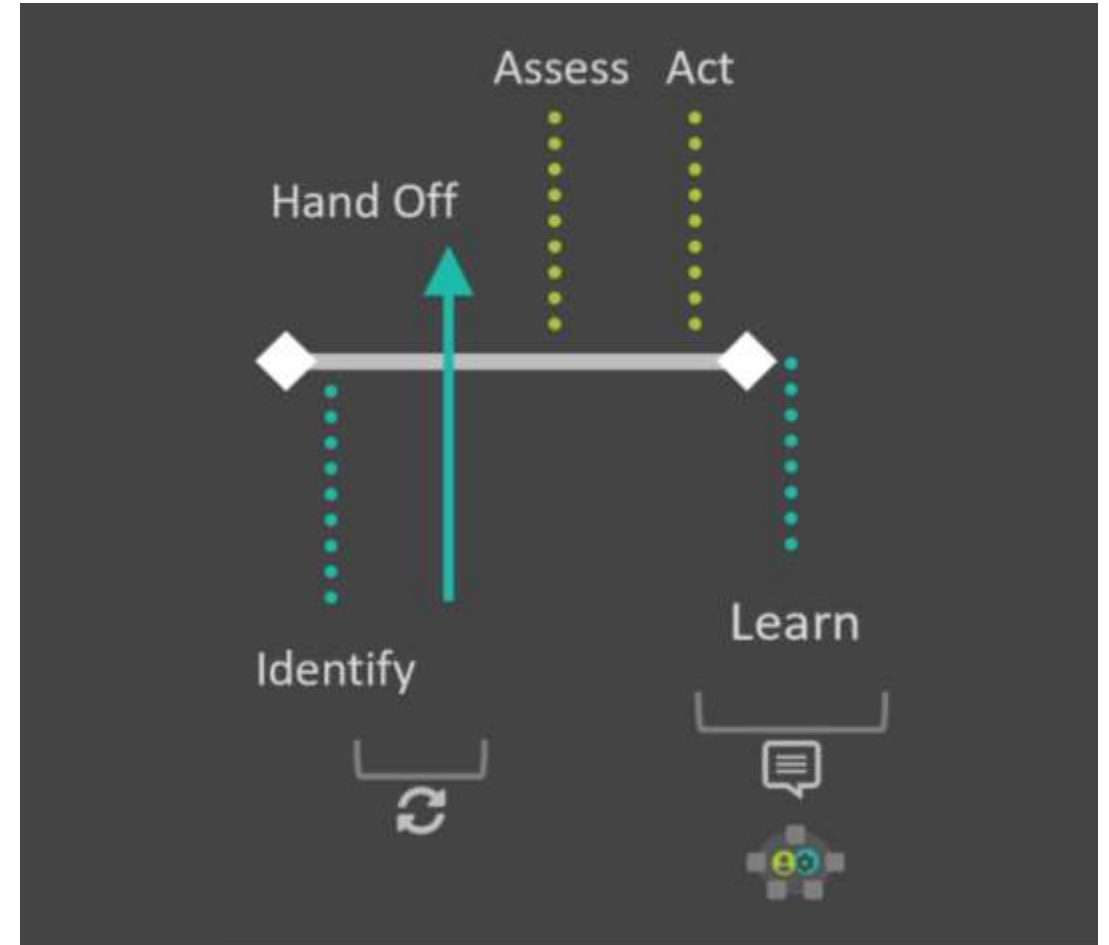
<http://www.flickr.com/photos/rockyvi/6451635085/sizes/m/in/photolist-aQ7jkF/>  
Some rights reserved by Rocky VI - <http://www.flickr.com/photos/rockyvi/>  
License: <http://creativecommons.org/licenses/by-nc-nd/2.0/>

# Significant decisions

Significant decisions made by the system

- explained
- able to be overridden
- appealable and reversible

Responsibilities explicitly defined between people and systems



How IAs Can Shape the Future of Human-AI Collaboration. Carol Smith and Duane Degler.  
Presented on April 28-30, 2021 at the Information Architecture Conference (IAC21)

# Plan for Long Term Implementation

- Cannot set and forget  
– dynamic systems
- Data curating
- Training management
- Backend system support
- Continuous monitoring  
and evaluation



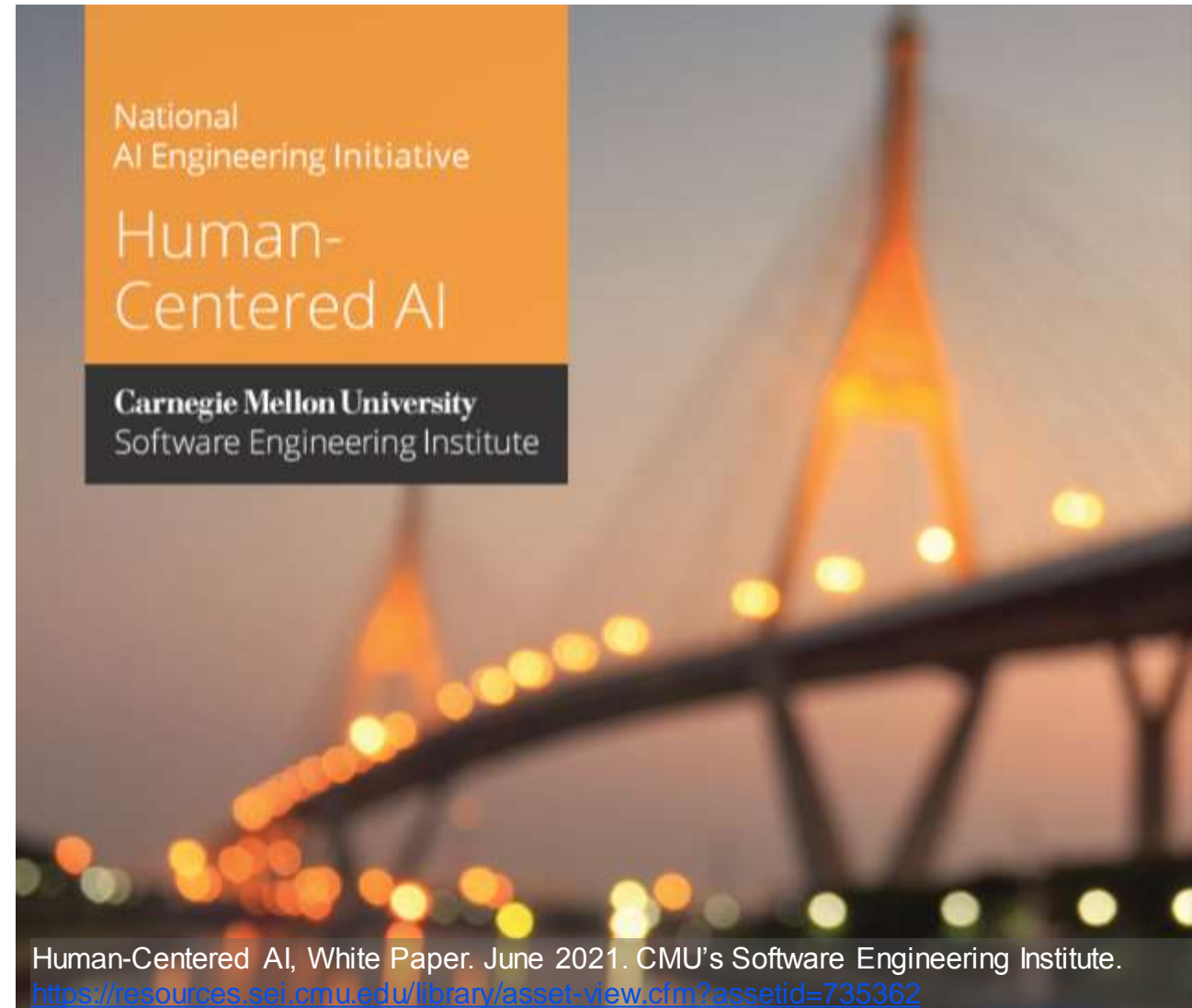
Nacho Kamenov & Humans in the Loop / Better Images of AI /  
A trainer instructing a data annotator on how to label images / CC-BY 4.0

# Design to work with, and for, people

Effective implementations

Minimize unintended  
consequences

1. Understand complexity of context
2. Design for human-machine teaming
3. Engage in critical oversight



Rob McCargow  
@robmccargow

Follow

# Learn about making Responsible AI

"Toward ethical, transparent and fair AI/ML & #MachineLearning: a critical reading list" by Eirini Malliaraki  
medium.com/@eirinimalliar ...  
#ResponsibleAI #ExplainableAI #AIethics



**Toward ethical, transparent and fair AI/ML: a critical reading list**

In the past 5 years there's been a lot of enthusiasm about AI and specifically machine learning and deep learning. As we continuously...

medium.com

7:09 AM - 26 Feb 2018 from London, England

Toward ethical, transparent and fair AI/ML: a critical reading list, by Eirini Malliaraki, Feb 19 via tweet from @robmccargow  
<https://medium.com/@eirinimalliaraki/toward-ethical-transparent-and-fair-ai-ml-a-critical-reading-list-d950e70a70ea>



SEI - Research and Capabilities - All Work - Designing Trustworthy Artificial Intell...

Designing Trustworthy Artificial Intelligence

CREATED OCTOBER 2018



SEI on HMT: [https://sei.cmu.edu/research-capabilities/all-work/display.cfm?customel\\_datapageid\\_4050=197910](https://sei.cmu.edu/research-capabilities/all-work/display.cfm?customel_datapageid_4050=197910)

# Adopt Technology Ethics

Reduce risk and unwanted bias

What do you value?

What lines won't you cross?

Track progress



# Conversations for Understanding

## Prompts to pair with Tech Ethics

## Bridge gaps between

## “do no harm” and reality

## Mitigation planning

## Support inspection



CMU SEI Checklist and Agreement - Downloadable PDF: <https://resources.sei.cmu.edu/library/asset-view.cfm?assetid=636620>

## Designing Ethical AI Experiences: Checklist and Agreement

**USE THIS DOCUMENT TO GUIDE THE DEVELOPMENT** of accountable, de-risked, respectful, secure, honest, and usable artificial intelligence (AI) systems with a diverse team aligned on shared ethics. An initial version of this document was presented with the paper *Designing Trustworthy AI: A Human-Machine Teaming Framework to Guide Development* by Carol Smith, available at <https://arxiv.org/abs/1910.03515>.

We will design our AI system with the following in mind:

- Designated humans have the ultimate responsibility for all decisions and outcomes:
  - Responsibilities are explicitly defined between the AI system and human(s), and how they are shared.
  - Human responsibility will be preserved for final decisions that affect a person's life, quality of life, health, or reputation.
  - Humans are always able to monitor, control, and deactivate systems.
- Significant decisions made by the AI system will be
  - explained
  - able to be overridden
  - appealable and reversible

We work to speculatively identify the full range of risks and benefits:

- Harmful, malicious use and consequences, as well as good, beneficial use and consequences
- We will be cognizant and exhaustively research unintended consequences.

**We will create plans for the misuse/abuse of the AI system, including the following:**

- communication plans to share pertinent information with all affected people
- mitigation plans for managing the identified speculative risks

**We value respect and security:**

- incorporating our values of humanity, ethics, equity, fairness, accessibility, diversity, and inclusion
- respecting privacy and data rights (Only necessary data will be collected.)
- providing understandable security methods
- making the AI system robust, valid, and reliable

We value transparency with the goal of engendering trust:

- The purpose, limitations, and biases of the AI system are explained in plain language.
- Data sources have unambiguous respected sources, and biases are known and explicitly stated.
- Algorithms and models are appropriate and verifiable.
- Confidence and context are presented for humans to base decisions on.
- Transparent justification for recommendations and outcomes is provided.
- Straightforward and interpretable monitoring systems are provided.

**We value honesty and usability:**

- Humans can easily discern when they are interacting with the AI system vs. a human.
- Humans can easily discern when and why the AI system is taking action and/or making decisions.
- Improvements will be made regularly to meet human needs and technical standards.

Team Signatures and Date

### About the SEI

The Software Engineering Institute is a federally funded research and development center (FFRDC) that works with defense and government organizations, industry, and academia to advance the state of the art in software engineering and cybersecurity to benefit the public interest. Part of Carnegie Mellon University, the SEI is a national resource in providing emerging technologies, expertise by software acquisition, and software lifecycle assistance.

### Contact Us

CARNEGIE MELLON UNIVERSITY  
SOFTWARE ENGINEERING INSTITUTE  
4800 FIFTH AVENUE, PITTSBURGH, PA 15213-2812  
se@cmu.edu  
412.268.1883 | 888.201.4479  
info@se.cmu.edu

©2019 Carnegie Mellon University | 5271 | E: 10/17/2019 | S: 12/08/2019

# AI has great potential, develop with caution

*“AI will ensure appropriate human judgement and not replace it”*

- Defense Innovation Board. 2019

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

Empower diverse teams, inclusive environments

Encourage deep conversations

Activate curiosity; be speculative; imaginative

# Don't fear AI - Explore AI

Try out tools  
Pair with others

Carol J. Smith

Twitter: @carologic

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

CMU Software Engineering Institute,  
AI Division

Twitter: @SEI\_CMU\_AI