

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.

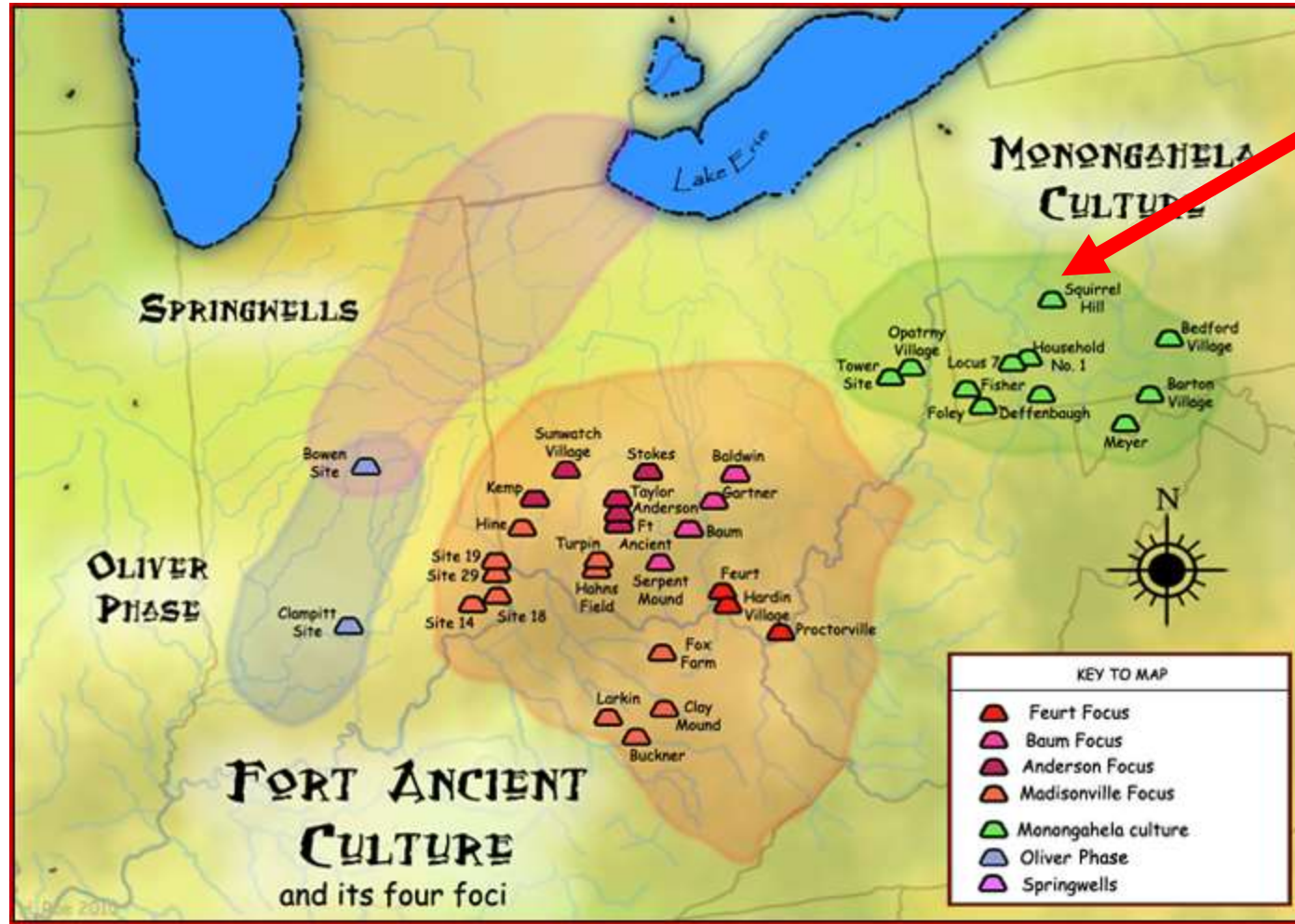
[DISTRIBUTION STATEMENT A] This material has been approved for public release and unlimited distribution. Please see Copyright notice for non-US Government use and distribution.

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.

DM22-0133

# 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*

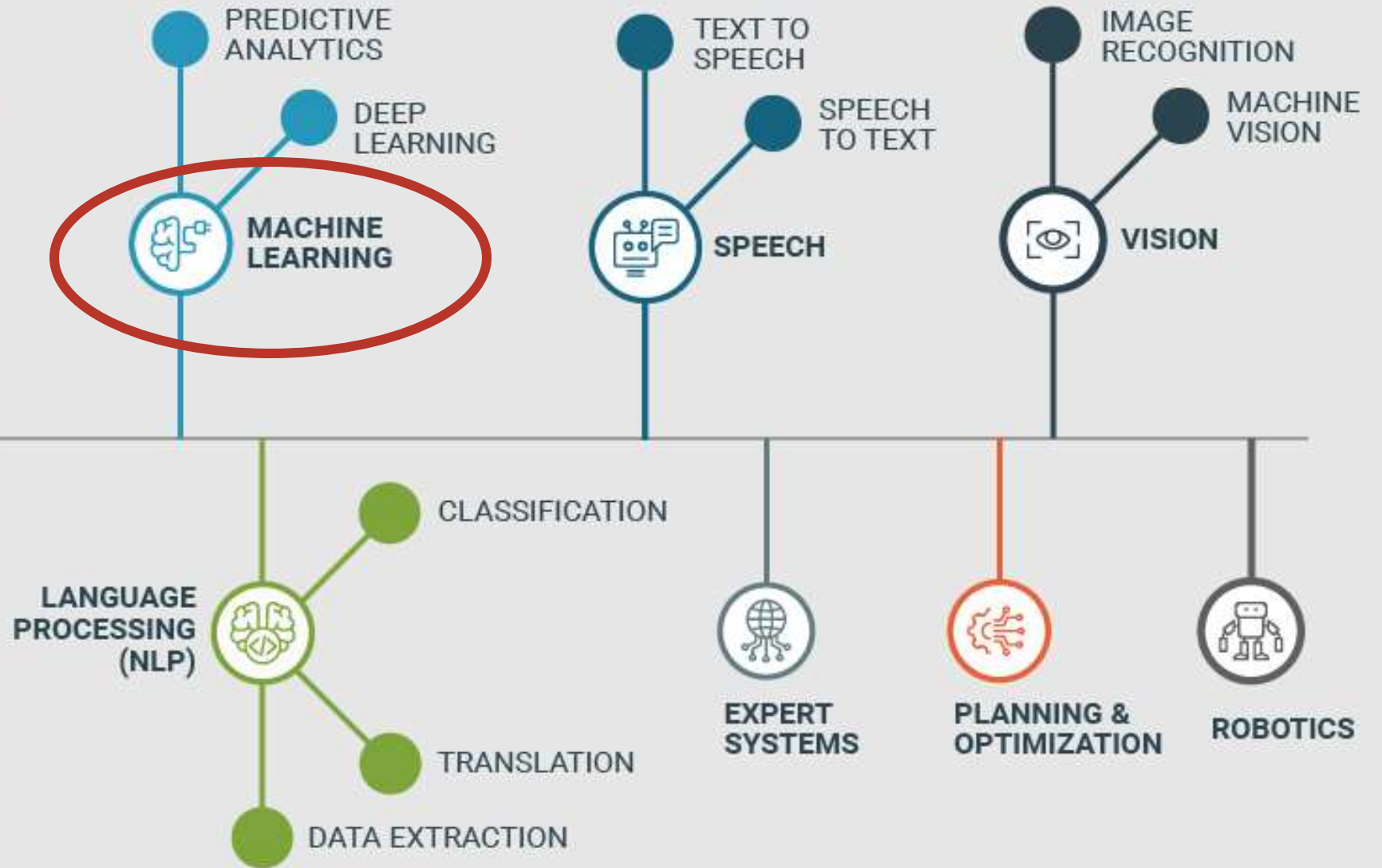
For additional information, please visit <http://dsp.acm.org/>

# What is artificial intelligence?



- AI is present when machines:
- Exhibit intelligence
  - Perceive environment
  - Take action / make decision to maximize chance of success at goal

# ARTIFICIAL INTELLIGENCE



# AI / Machine Learning

## Algorithms

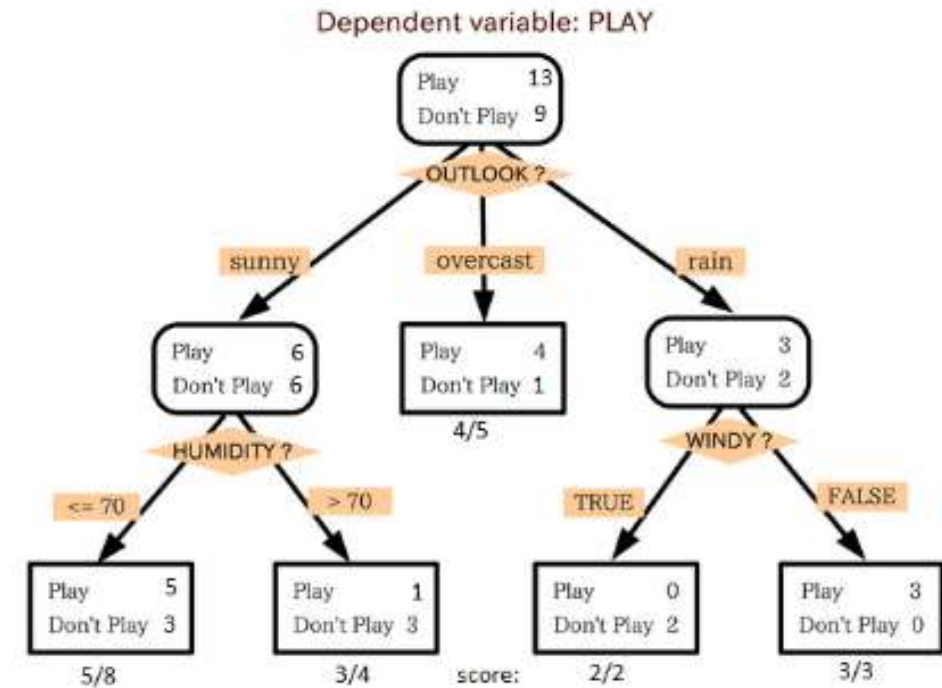
(math + programming)

Know ONLY what taught

Control ONLY given control of

Aware of nuances

- can continue to learn



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

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



# AI is NOT sentient

# Not unknowable

# Never Enough Time

Physician: 90+ hours reading  
a week

AI can bring that information  
to the physician

Enabling more  
evidence-based decisions



# Transfer human concepts and relationships

## Number Five “Needs Input”



# Supervised (by a human) machine learning

# Enormous amount of work

# Dependent on Experts

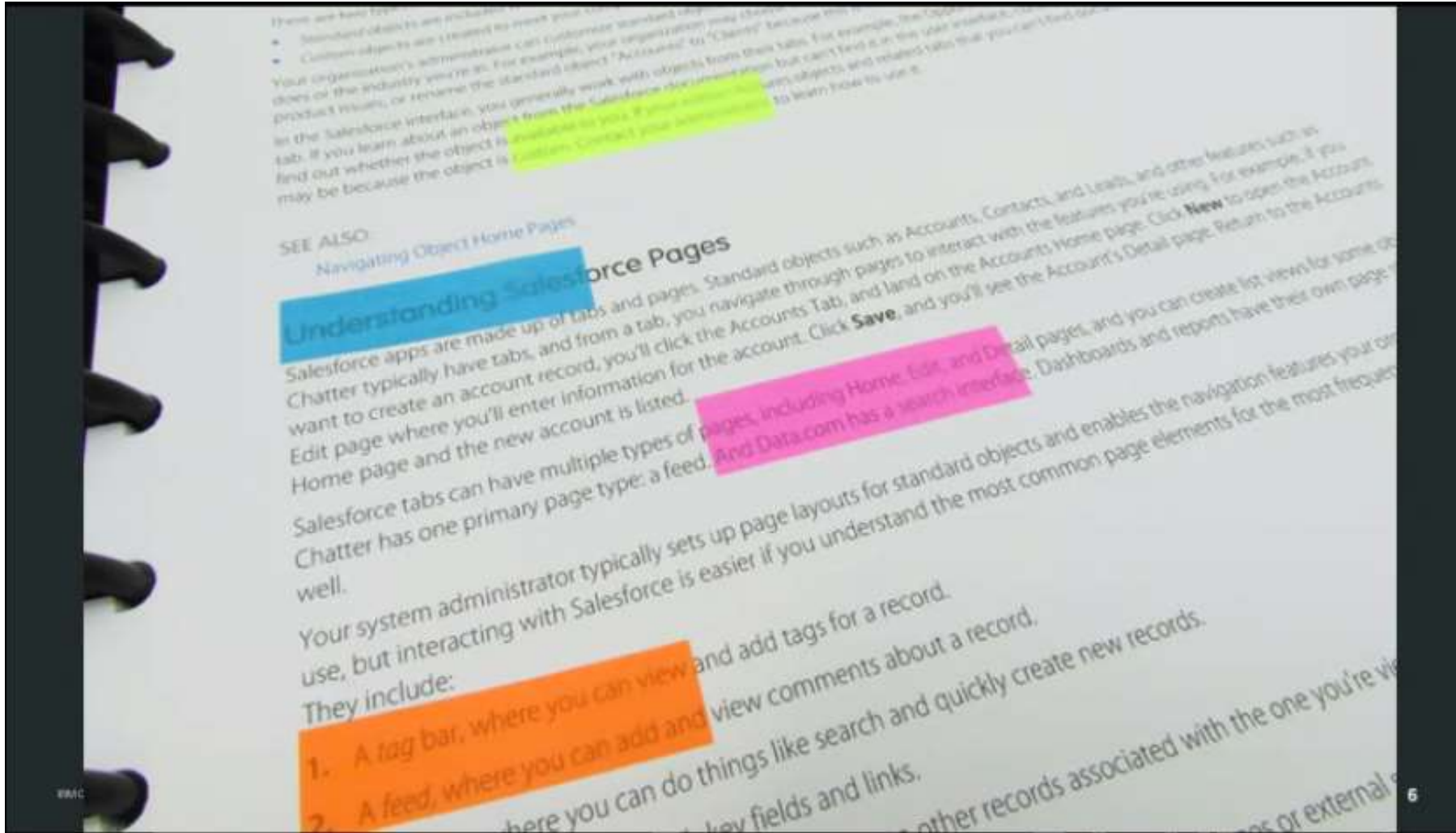
Data scientists

Subject matter experts (SME's) availability

- Lawyers
- Machinists
- Insurance adjusters
- Physicians

Not just experts in machine learning

# 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...**

# Taxonomies and ontologies

# Creating ML requires

- Content and algorithms
- Ground truth (annotation)
- Teaching and iteration
- Repetition with new content
- Time - weeks to months to start, ongoing

**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=542374>

# Only know what taught

## Training data



Unrepresentative  
or incomplete training data

## Data encountered



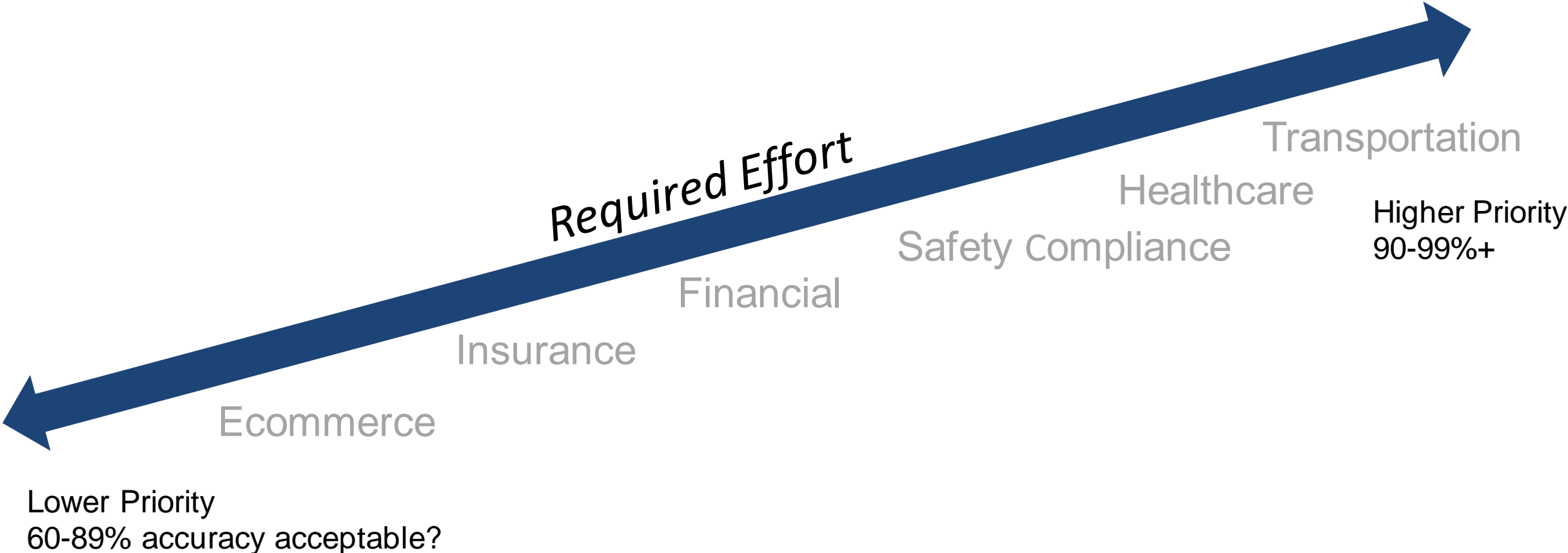
Unlikely to recognize

**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

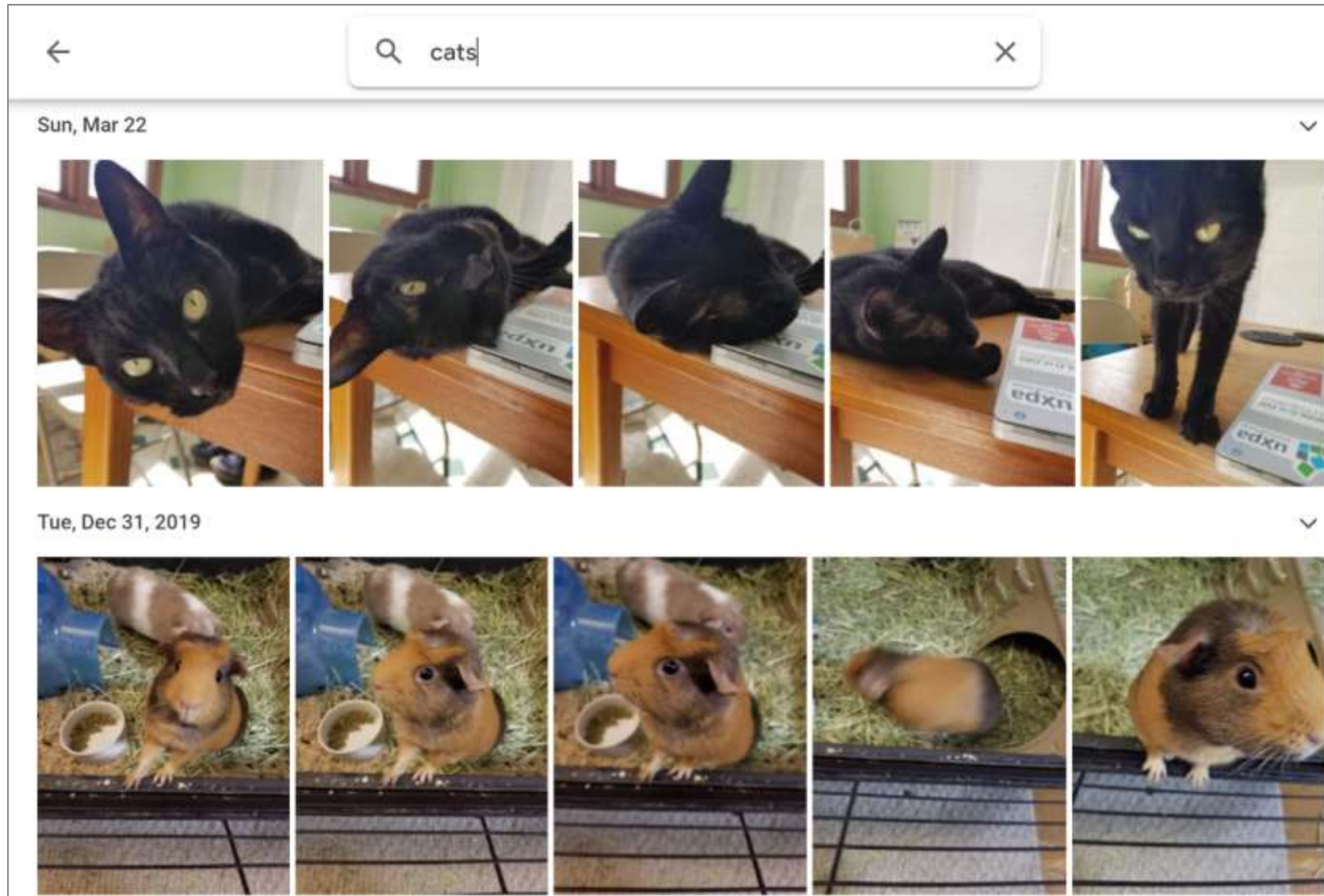
2016 Go, Google

Knowledge?

Ethics?



# Image Recognition – Google Photos



# Sound recognition: Labeling of birdsongs



# 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 |**



# Decision Making: Autonomous vehicles



# Bias in AI

# How can Data be biased?

Goal: Select the right lawn care treatment, saving humans time.

Multiple data source choices.

Whose data do you use?



# Selecting Data Source

## Company A

- Primarily uses chemicals to treat lawns
- Data likely biased towards chemical use

## Company B

- Only “all natural” treatments
- Data likely biased against chemical use

# Selecting Data Source

## Company A

- Primarily uses chemicals to treat lawns
- Data likely biased towards chemical use

## Company B

- Only “all natural” treatments
- Data likely biased against chemical use

Neither are wrong.  
Both are biased.

# 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>

# 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



Rob McCargow  
@robmccargow

Follow

Toward ethical, transparent and fair AI & #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>

# Learn about making ethical, transparent and fair AI



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

Designing Trustworthy Artificial Intelligence  
CREATED OCTOBER 2019



SEI on HMT: [https://sei.cmu.edu/research-capabilities/all-work/display.cfm?custom1\\_datapageid\\_4050=197910](https://sei.cmu.edu/research-capabilities/all-work/display.cfm?custom1_datapageid_4050=197910)

# Adopt Technology Ethics

What do you value?

What lines won't you cross?

Track progress



# Prompts for Conversations

Use a Checklist – find concealed tasks

- Pair with Tech Ethics
- Bridges gap between “do no harm” and reality

Reduce risk and unwanted bias

Mitigation planning

Support inspection



Carnegie Mellon University  
Software Engineering Institute

## 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>.

|  |  |   |
|--|--|---|
| <p><b>We will design our AI system with the following in mind:</b></p> <ul style="list-style-type: none"><li><input type="checkbox"/> Designated humans have the ultimate responsibility for all decisions and outcomes:<ul style="list-style-type: none"><li>• Responsibilities are explicitly defined between the AI system and human(s), and how they are shared.</li><li>• Human responsibility will be preserved for final decisions that affect a person's life, quality of life, health, or reputation.</li><li>• Humans are always able to monitor, control, and deactivate systems.</li></ul></li><li><input type="checkbox"/> Significant decisions made by the AI system will be:<ul style="list-style-type: none"><li>• explained</li><li>• able to be overridden</li><li>• appealable and reversible.</li></ul></li></ul> | <p><b>We work to speculatively identify the full range of risks and benefits:</b></p> <ul style="list-style-type: none"><li><input type="checkbox"/> Harmful, malicious use and consequences, as well as good, beneficial use and consequences.</li><li><input type="checkbox"/> We will be cognizant and exhaustively research unintended consequences.</li></ul> <p><b>We will create plans for the misuse/abuse of the AI system, including the following:</b></p> <ul style="list-style-type: none"><li><input type="checkbox"/> communication plans to share pertinent information with all affected people</li><li><input type="checkbox"/> mitigation plans for managing the identified speculative risks</li></ul> <p><b>We value respect and security:</b></p> <ul style="list-style-type: none"><li><input type="checkbox"/> incorporating our values of humanity, ethics, equity, fairness, accessibility, diversity, and inclusion</li><li><input type="checkbox"/> respecting privacy and data rights (Only necessary data will be collected.)</li><li><input type="checkbox"/> providing understandable security methods</li><li><input type="checkbox"/> making the AI system robust, valid, and reliable</li></ul> | <p><b>We value transparency with the goal of engendering trust:</b></p> <ul style="list-style-type: none"><li><input type="checkbox"/> The purpose, limitations, and biases of the AI system are explained in plain language.</li><li><input type="checkbox"/> Data sources have unambiguous respected sources, and biases are known and explicitly stated.</li><li><input type="checkbox"/> Algorithms and models are appropriate and verifiable.</li><li><input type="checkbox"/> Confidence and context are presented for humans to base decisions on.</li><li><input type="checkbox"/> Transparent justification for recommendations and outcomes is provided.</li><li><input type="checkbox"/> Straightforward and interpretable monitoring systems are provided.</li></ul> <p><b>We value honesty and usability:</b></p> <ul style="list-style-type: none"><li><input type="checkbox"/> Humans can easily discern when they are interacting with the AI system vs. a human.</li><li><input type="checkbox"/> Humans can easily discern when and why the AI system is taking action and/or making decisions.</li><li><input type="checkbox"/> Improvements will be made regularly to meet human needs and technical standards.</li></ul> |
|--|--|---|

Team Signatures and Date

---

**About the SEI**  
The Software Engineering Institute is a federally funded research and development center (FDRC) that works with defense and government organizations, industry, and academia to advance the state of the art in software engineering and operations by to benefit the public interest. Part of Carnegie Mellon University, the SEI is a national leader in pioneering emerging technologies, cyber security, software assurance, and software program execution.

**Contact Us**  
CARNegie MELLon UNIVERSITY  
SOFTWARE ENGINEERING INSTITUTE  
4800 WYTH AVENUE, PITTSBURGH, PA 15215-2872  
412.261.4000  
412.261.5850 | 800.251.4475  
sei@sei.cmu.edu

©2019 Carnegie Mellon University | 5215 | C-10.17.2019 | 10.10.2019



ARTIFICIAL INTELLIGENCE PORTFOLIO

# Responsible AI Guidelines

Operationalizing DoD's Ethical Principles  
for AI

Download DIU's  
Responsible AI  
Guidelines report and  
learn how to  
implement ethical AI  
principles.

[Responsible AI Guidelines](#)

<https://www.diu.mil/responsible-ai-guidelines>



# 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