CMU Tepper School of Business
December 5, 2022

Human Centered AI

Considerations for Developing AI Technologies

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

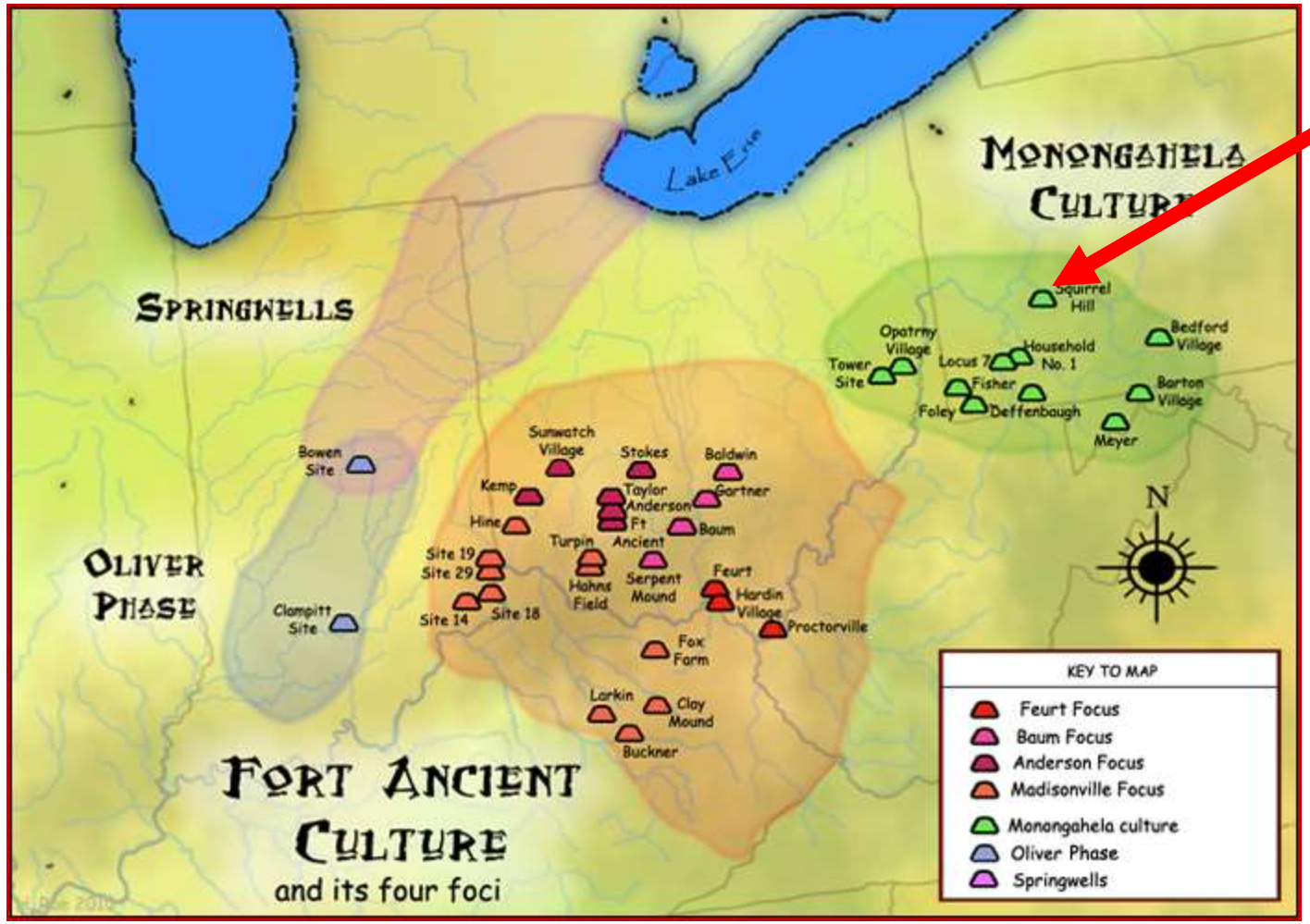
AI Division
Software Engineering Institute
Carnegie Mellon University
Pittsburgh, PA 15213

Copyright Statement

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.

DM22-1158

Acknowledgement: The Land I Speak On



Land of Monongahela,
Adena and Hopewell Nations;

Seneca, Lenape
and Shawnee lands;

Osage, Delaware
and Iroquois lands.

Now known
as Pittsburgh, PA, USA.

Map by Herb Roe via Wikipedia https://en.wikipedia.org/wiki/Monongahela_culture

Carol's Roles

Software Engineering Institute

AI Division Staff

- Sr. Research Scientist in Human-machine interaction
- Government customers
- AI, autonomy, emerging technology

Adjunct Instructor

Interaction Design Overview

- Human-centered design
- Prototyping
- Design and iteration

What is artificial intelligence?

AI systems can

-recognize patterns

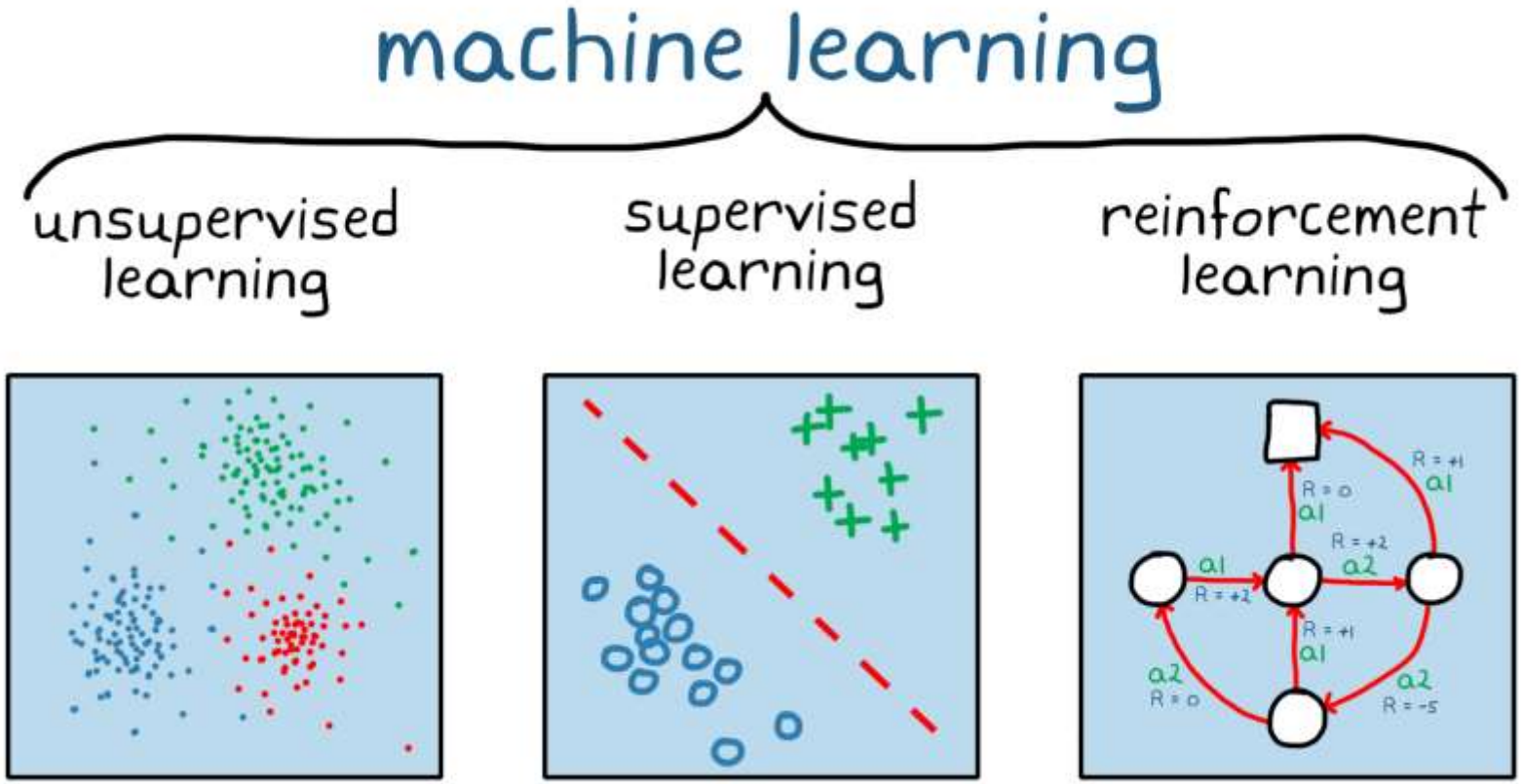
-create predictions

-make decisions, and/or

-generate new content

without being explicitly programmed to do so.

Artificial Intelligence



Deep learning, neural networks

Requirements for ML

1. **Data**: pre-existing, machine readable, relevant (amount vary)
2. **Math**: appropriate for data and context (statistics, probability, calculus…)
3. **Programming**: Python, C/C++, R, Java, JavaScript…

Math + programming = algorithm

Data + algorithm = ML model*

*The term "model" is used inconsistently. Model sometimes refers to an algorithm without data.

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

Photo: 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) availability

- Lawyers
- Machinists
- Insurance adjusters
- Physicians

Not just experts in machine learning

Experts Annotate Content

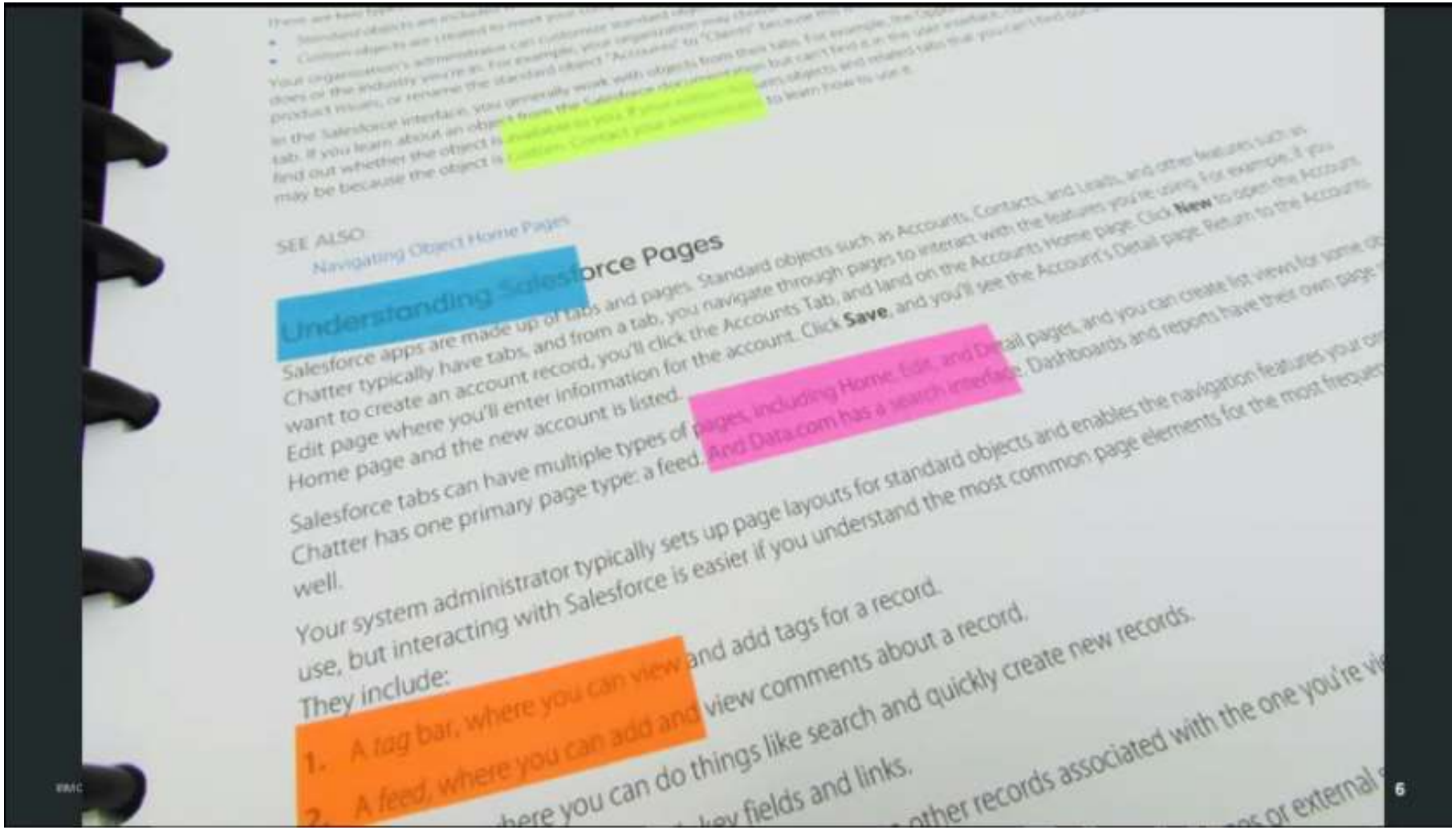


Image created by Angela Swindell, Visual Designer, IBM

Entity Type

High level concepts applied to a mention

PERSON

Amanda **Amanda Tomlin** **She**

Initial model created by Angela Swindell.

Define Entity Types

PERSON ORGANIZATION TIME

Amanda works at Carnegie Mellon University.

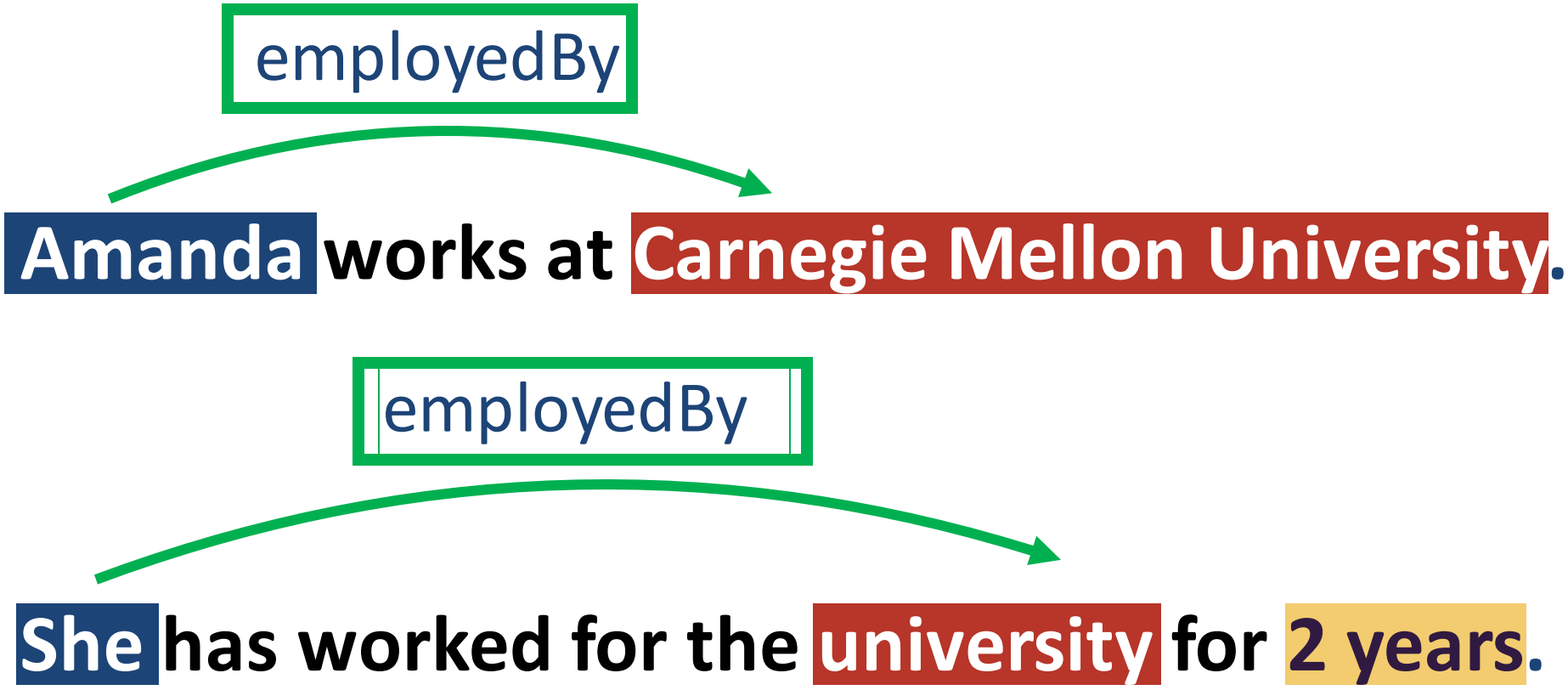
She has worked for the university for 2 years.

Initial model created by Angela Swindell.

Define Relationships

Initial model created by Angela Swindell.

Relation type



Initial model created by Angela Swindell.

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=542374

Only know what taught

Training data

Unrepresentative
or incomplete training data

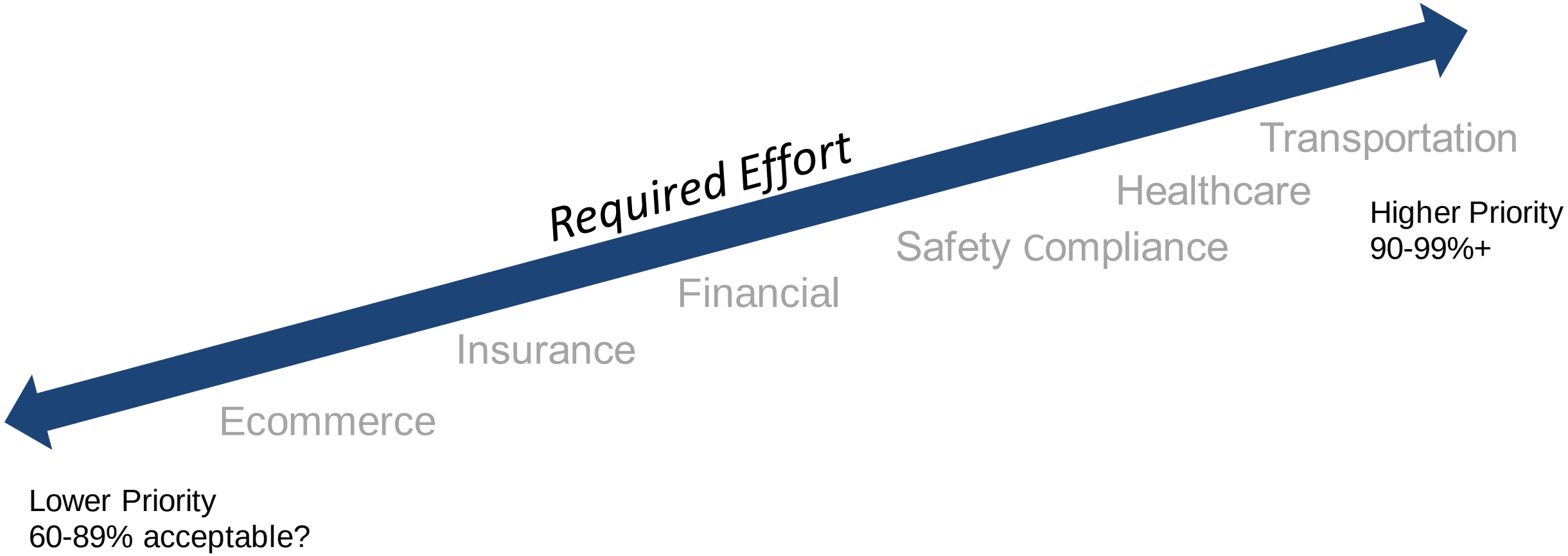
Data encountered

Unlikely to recognize

Only as good as data and time spent improving it.

Biased based on data and training.

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

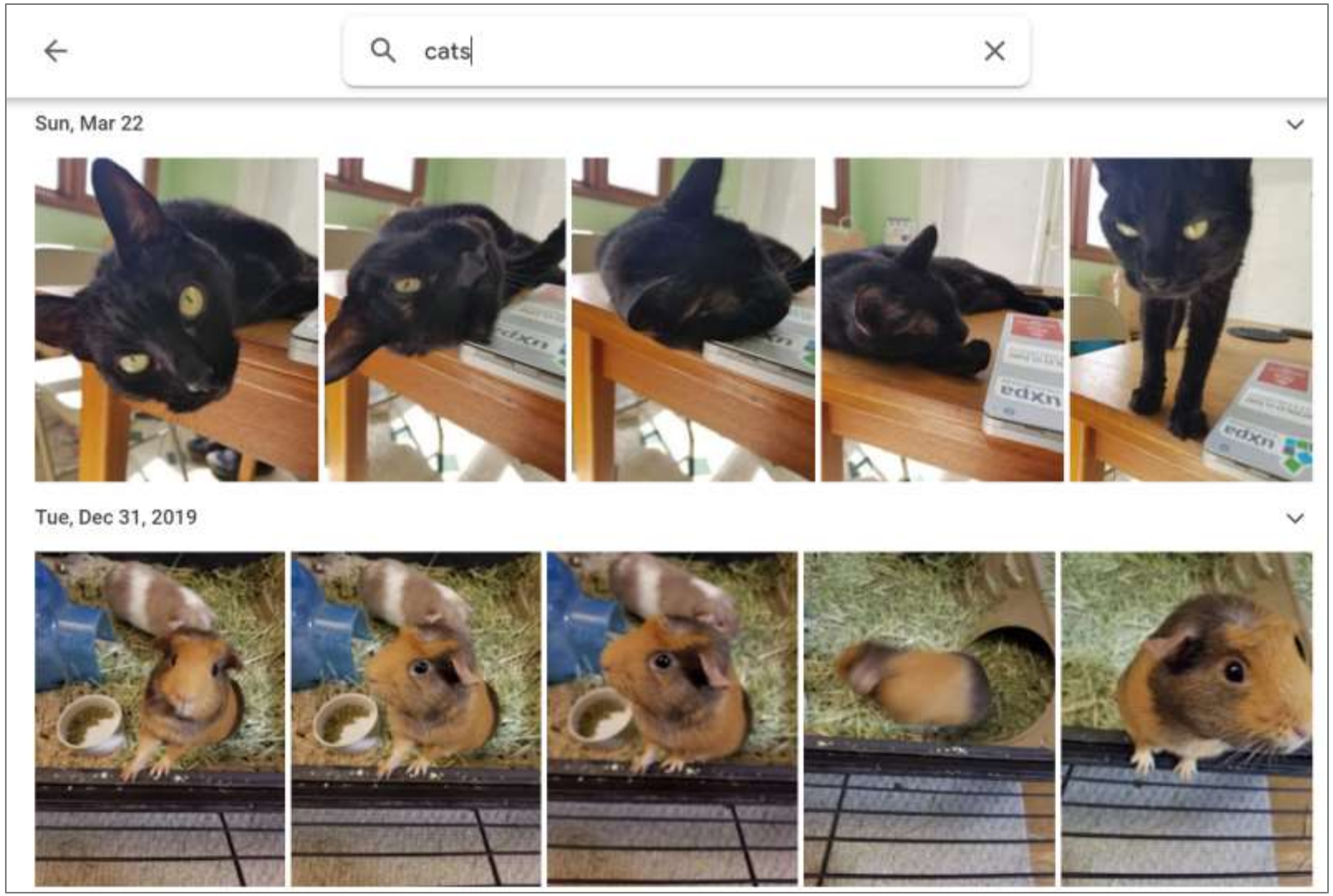
2016 Go, Google

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 15th Python in Science Conference (SCIPY 2016). 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

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.

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

- Joy Buolamwini, Algorithmic Justice League

Movie: Coded Bias

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?

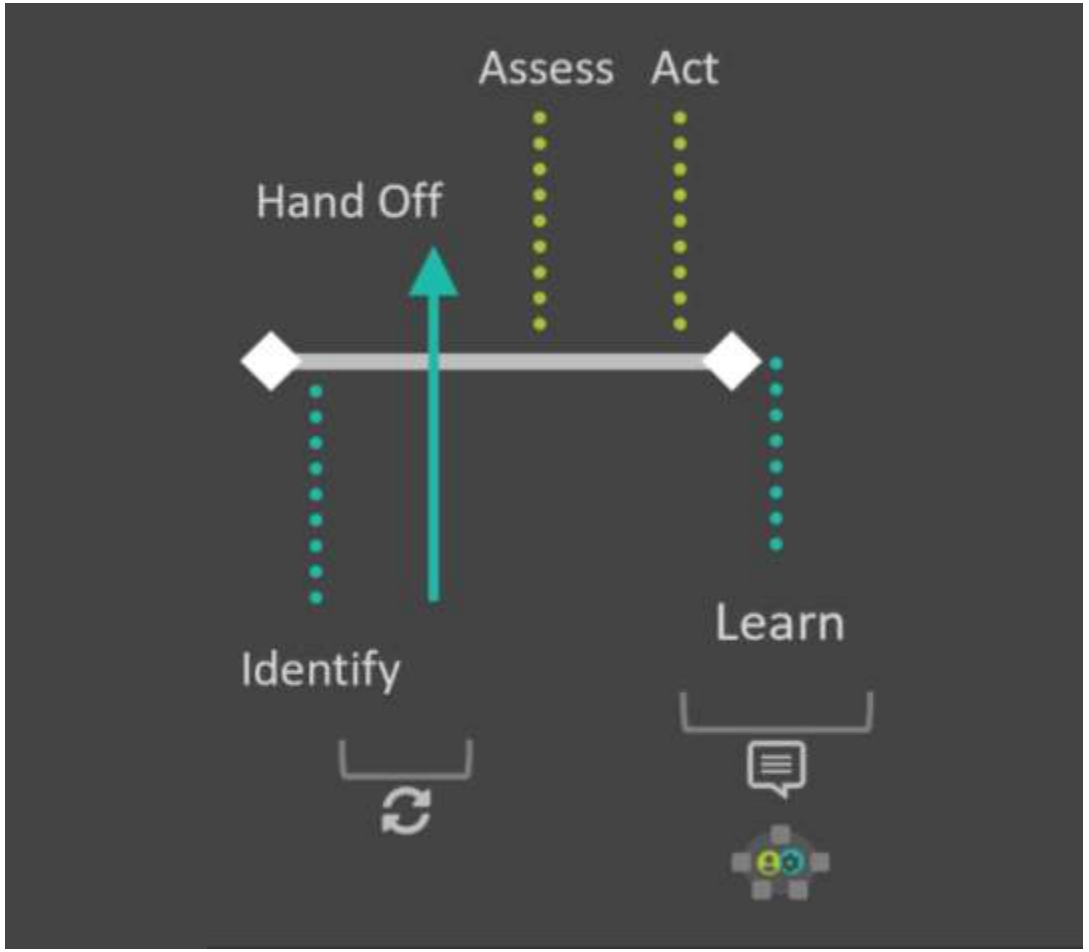
Make a plan

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)

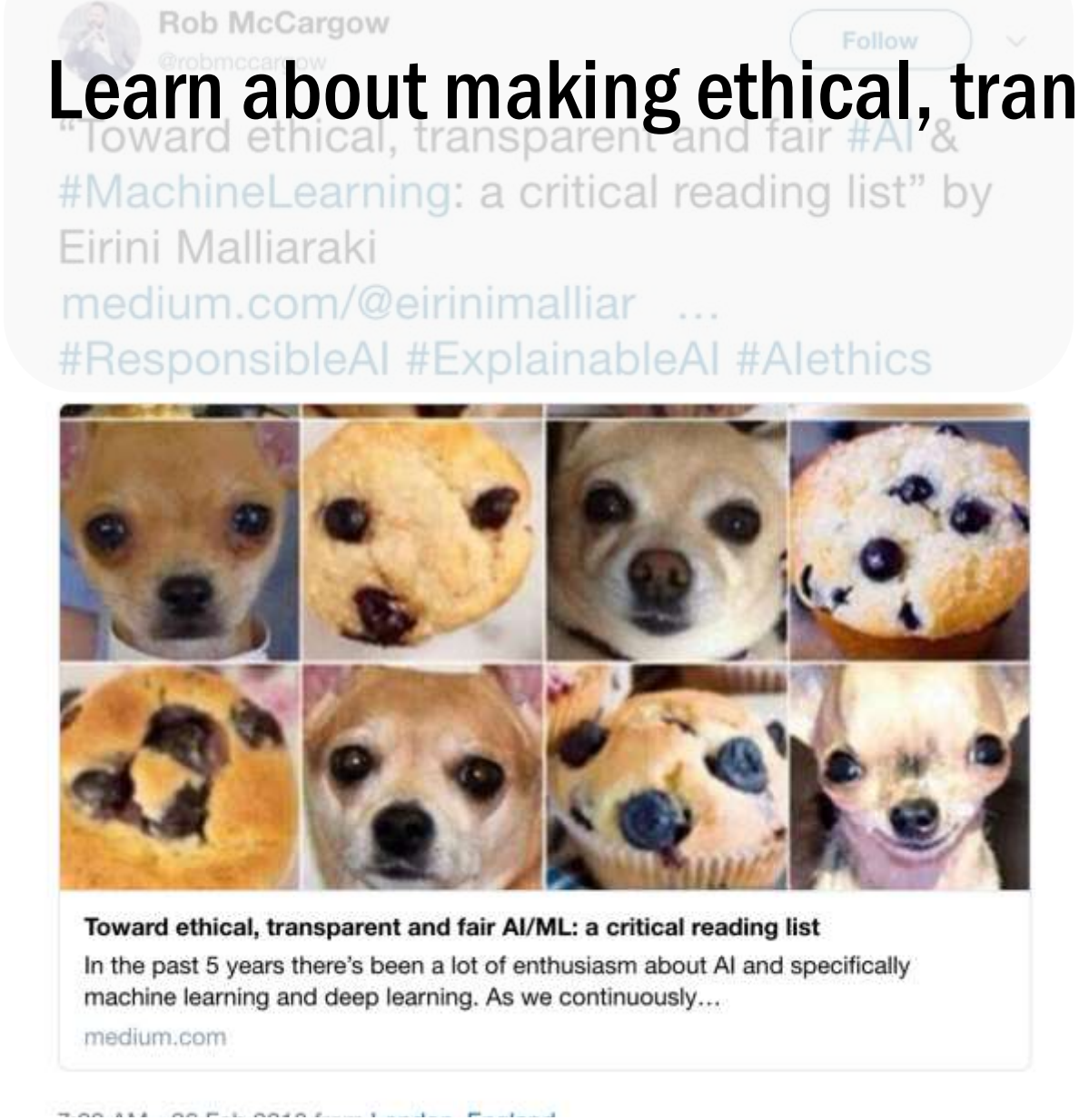
Plan for Long Term Implementation

- Cannot set and forget – dynamic systems

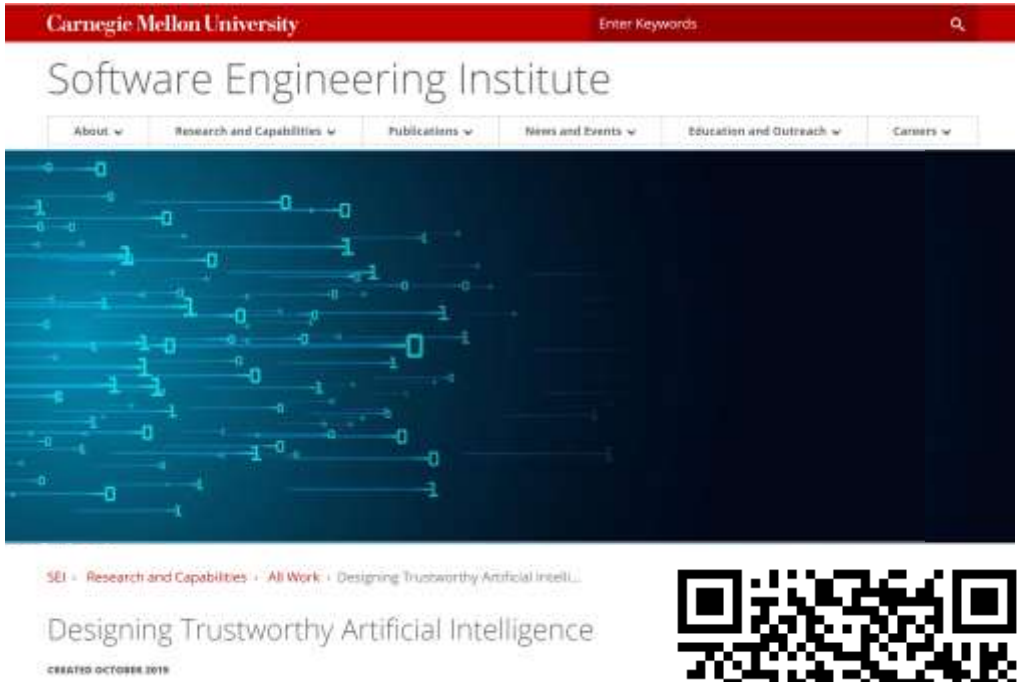
- Data curating
- Training management
- Backend system support
- Continuous monitoring and evaluation

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

Learn about making ethical, transparent and fair AI



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 on HMT: https://sei.cmu.edu/research-capabilities/all-work/display 10

Adopt Technology Ethics

Reduce risk and unwanted bias

What do you value?

What lines won't you cross?

Association for Computing Machinery

AI NOW INSTITUTE

Microsoft

Google

Montréal Declaration Responsible AI_

An initiative of Université de Montréal

U.S. DEPT OF DEFENSE

ETHICAL OS

Conversations for Understanding

Prompts to pair with Tech Ethics

Bridge gaps between
"do no harm" and reality

Mitigation planning

Support inspection

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

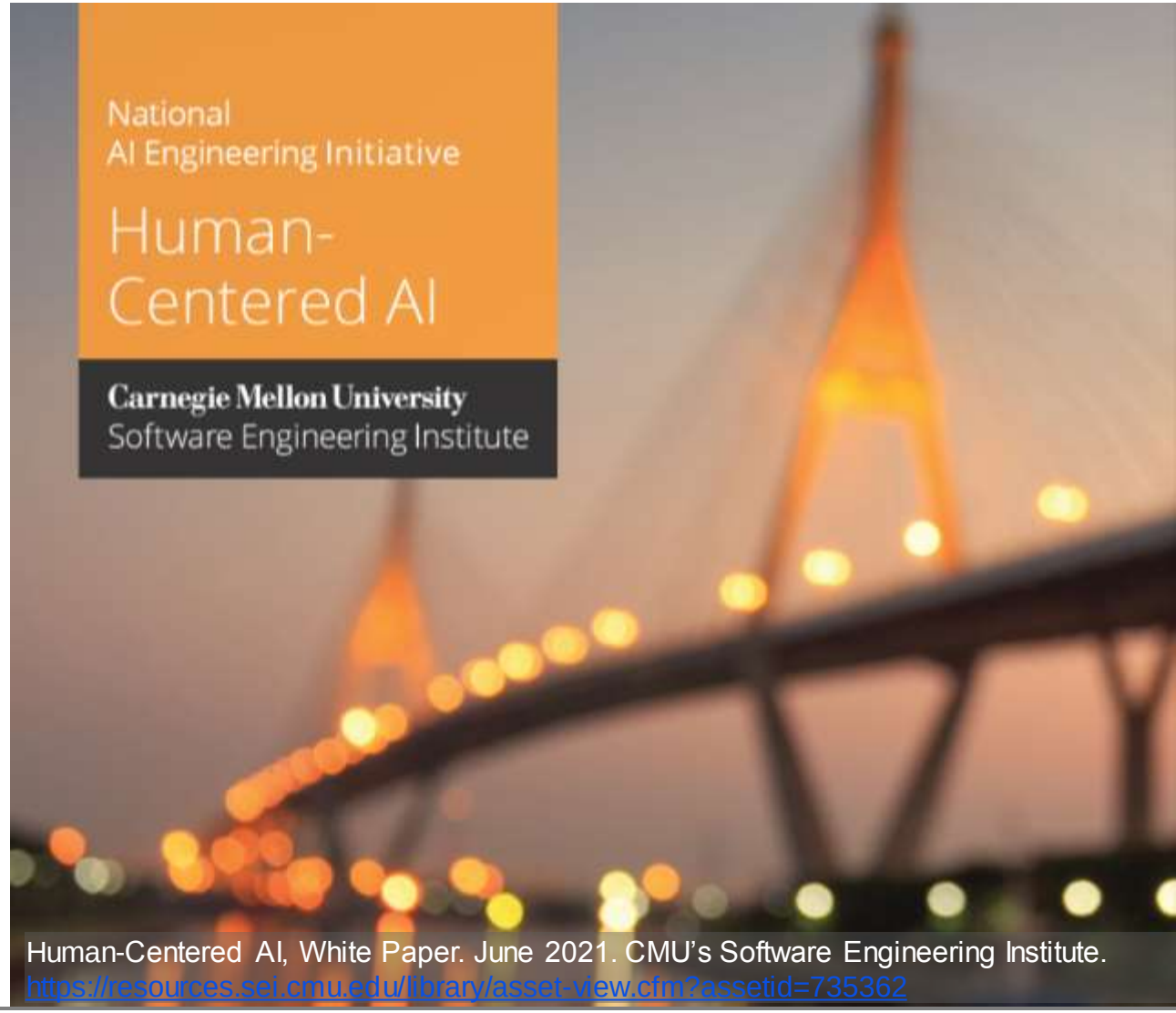
SEI Work

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



Human-Centered AI, White Paper. June 2021. CMU's Software Engineering Institute. https://resources.sei.cmu.edu/library/asset-view.cfm?assetid=735362

DEFENSE INNOVATION UNIT

ARTIFICIAL INTELLIGENCE PORTFOLIO

Responsible AI Guidelines

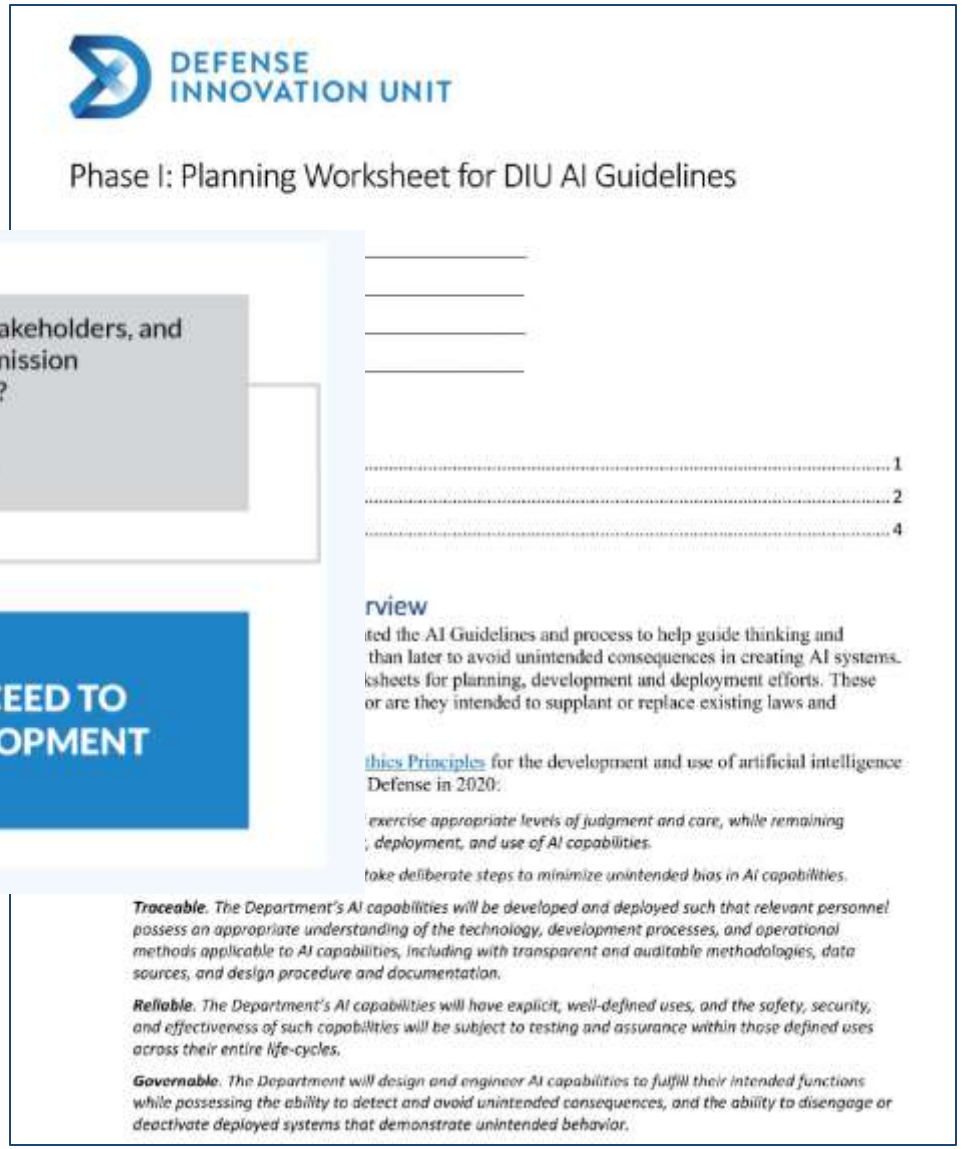
Operationalizing DoD's Ethical Principles for AI

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

Phase I: Planning


DEFENSE INNOVATION UNIT

Phase I: Planning Worksheet for DIU AI Guidelines

rview

ted the AI Guidelines and process to help guide thinking and than later to avoid unintended consequences in creating AI systems. ksheets for planning, development and deployment efforts. These or are they intended to supplant or replace existing laws and

thics Principles for the development and use of artificial intelligence Defense in 2020:

exercise appropriate levels of judgment and care, while remaining , deployment, and use of AI capabilities.

take deliberate steps to minimize unintended bias in AI capabilities.

***Traceable**. The Department's AI capabilities will be developed and deployed such that relevant personnel possess an appropriate understanding of the technology, development processes, and operational methods applicable to AI capabilities, including with transparent and auditable methodologies, data sources, and design procedure and documentation.*

***Reliable**. The Department's AI capabilities will have explicit, well-defined uses, and the safety, security, and effectiveness of such capabilities will be subject to testing and assurance within those defined uses across their entire life-cycles.*

***Governable**. The Department will design and engineer AI capabilities to fulfill their intended functions while possessing the ability to detect and avoid unintended consequences, and the ability to disengage or deactivate deployed systems that demonstrate unintended behavior.*

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

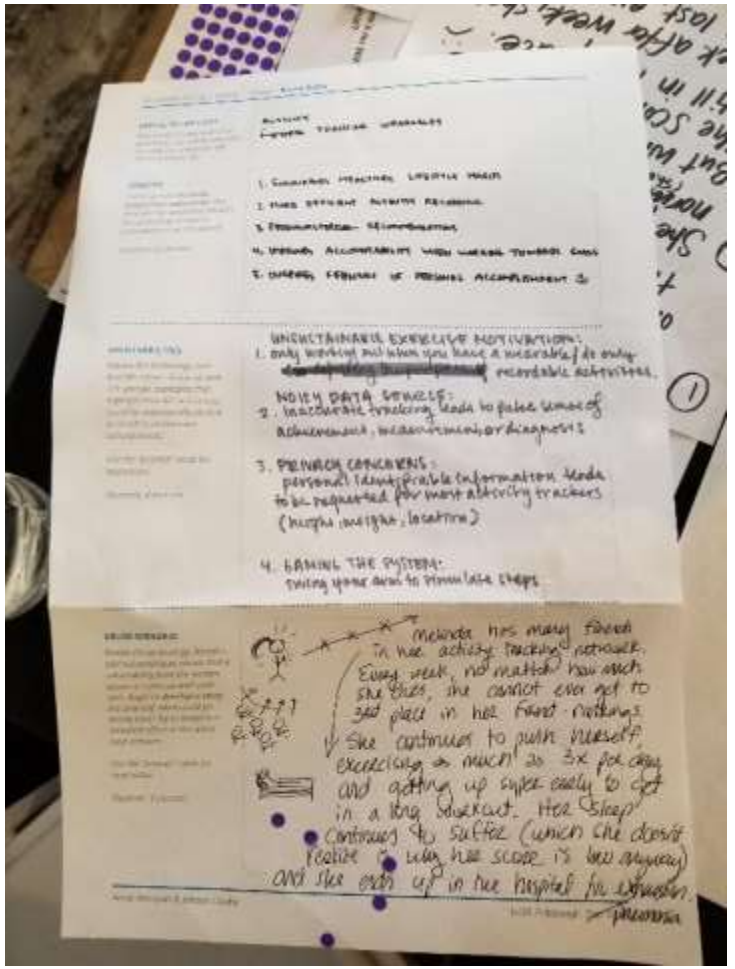
Activate curiosity; be speculative; imaginative

Don’t fear AI - Explore AI

Try out tools
Pair with others

Abusability Testing

Round Robin - online or in-person



Template by: Anna Abovyan & Allison Cosby,
IxDA Pittsburgh, Sep 2019

Overview of Abusability Testing

Pick a newish technology:

Self-checkout kiosks, drone package delivery, smart watches, etc.

Brainstorm:

- Benefits
- Vulnerabilities
- Abuse Scenario
- "Black Mirror" Episode Scenario (Casey Fiesler)

  (inspired by British dystopian sci-fi tv series of same name)

Prompt Statements

- What happens if devices can charge themselves?
- What if there's no WiFi or cellular signal?
- Is there a back door?
- How can this technology affect individuals, families, organizations, and society overall?
- Does this technology have a potential for profound negative impact on minorities (gender, sexual orientation, age, religion, nationality, etc.)?

What scenarios do you have?

- At work?
- In your personal life?
- In the news?
- Where else would you want to apply this?

Carol J. Smith

LinkedIn: https://www.linkedin.com/in/**caroljsmith**/

CMU Software Engineering Institute, AI Division