ACM
Distinguished Speaker Program

AI and Machine Learning Demystified

Carol J. Smith

Sr. Research Scientist - Human-Machine Interaction

Twitter: @carologic @sei_etc

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.

DM21-0572

About ACM

- ACM, the Association for Computing Machinery (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

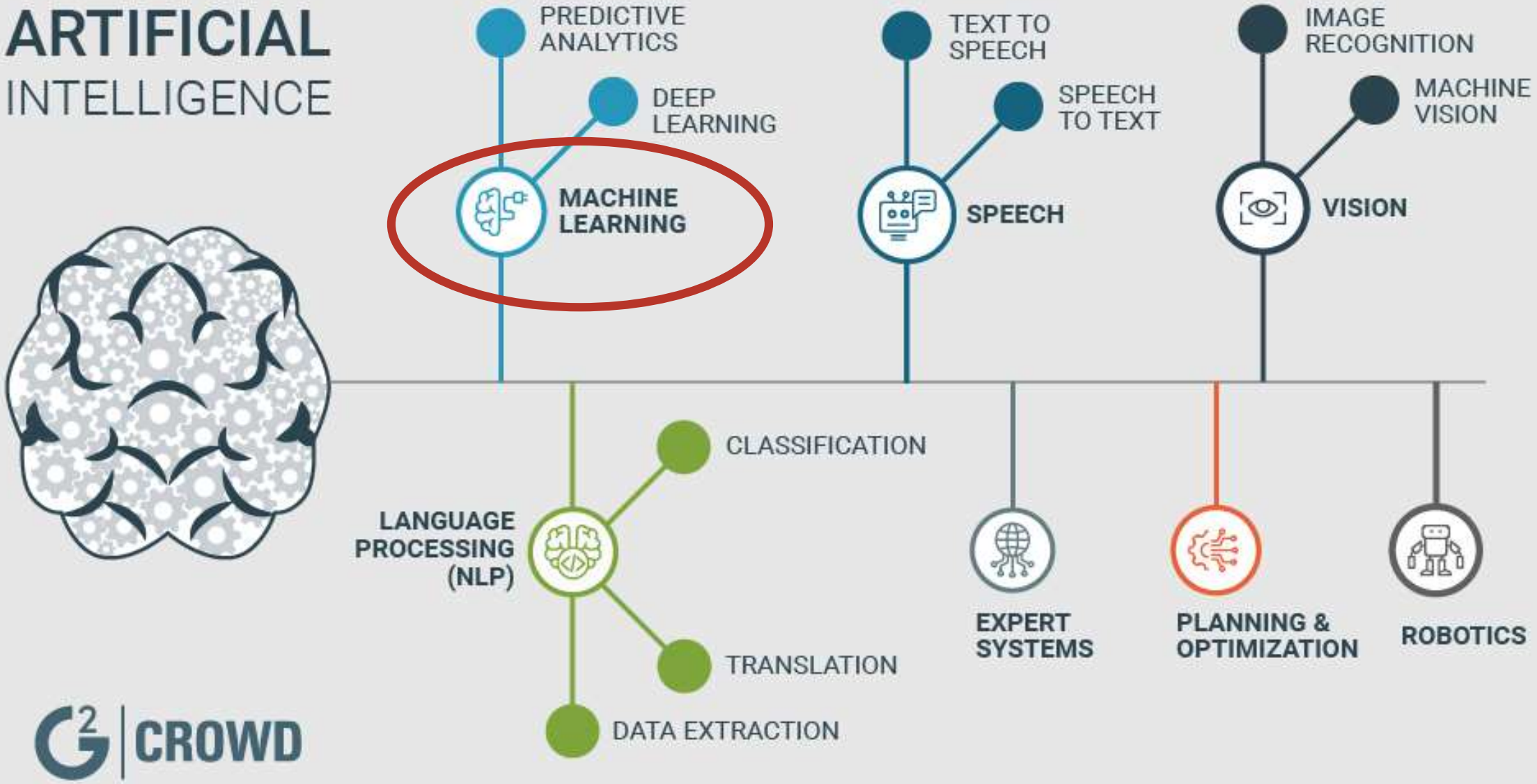
For additional information, please visit http://dsp.acm.org/

AI is as imperfect as the humans making it

What is artificial intelligence?

AI is present when machines:

- Exhibit intelligence
- Perceive environment
- Take action / make decision to maximize chance of success at goal



AI / Machine Learning

Algorithms
(math + programming)

Know ONLY what taught

Control ONLY given control of

Aware of nuances
- can continue to learn

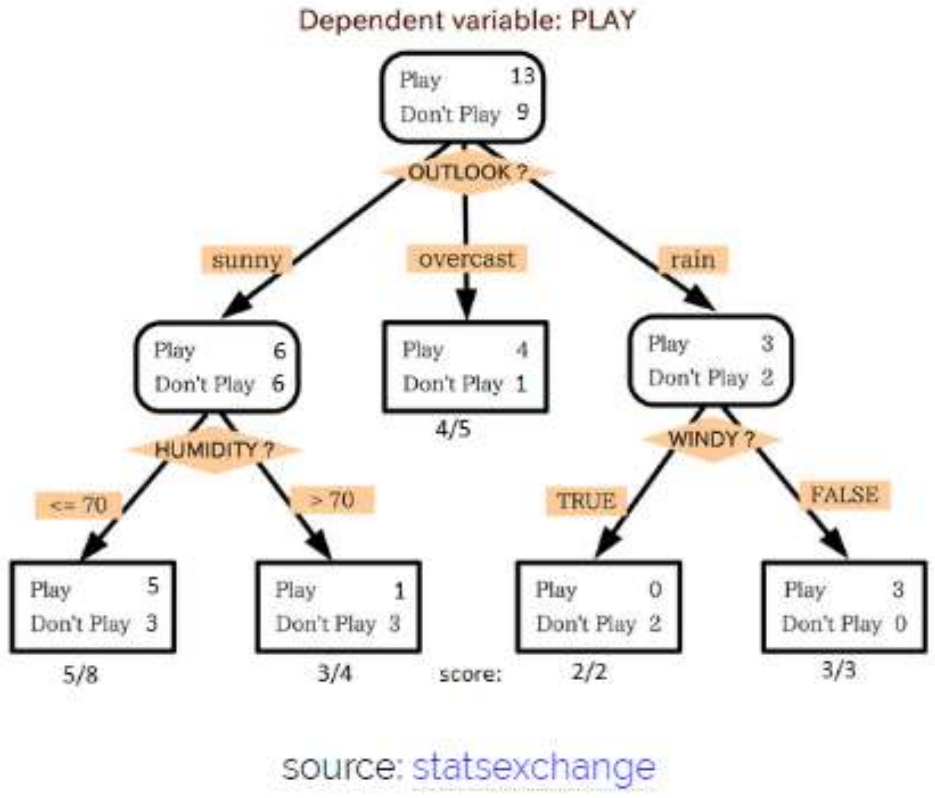


Image: https://www.analyticsvidhya.com/blog/2015/08/common-machine-learning-algorithms/

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

Photo: https://commons.wikimedia.org/wiki/File:Baby_Boy_Oliver.jpg

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

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 experts in machine learning

Experts Annotate Content

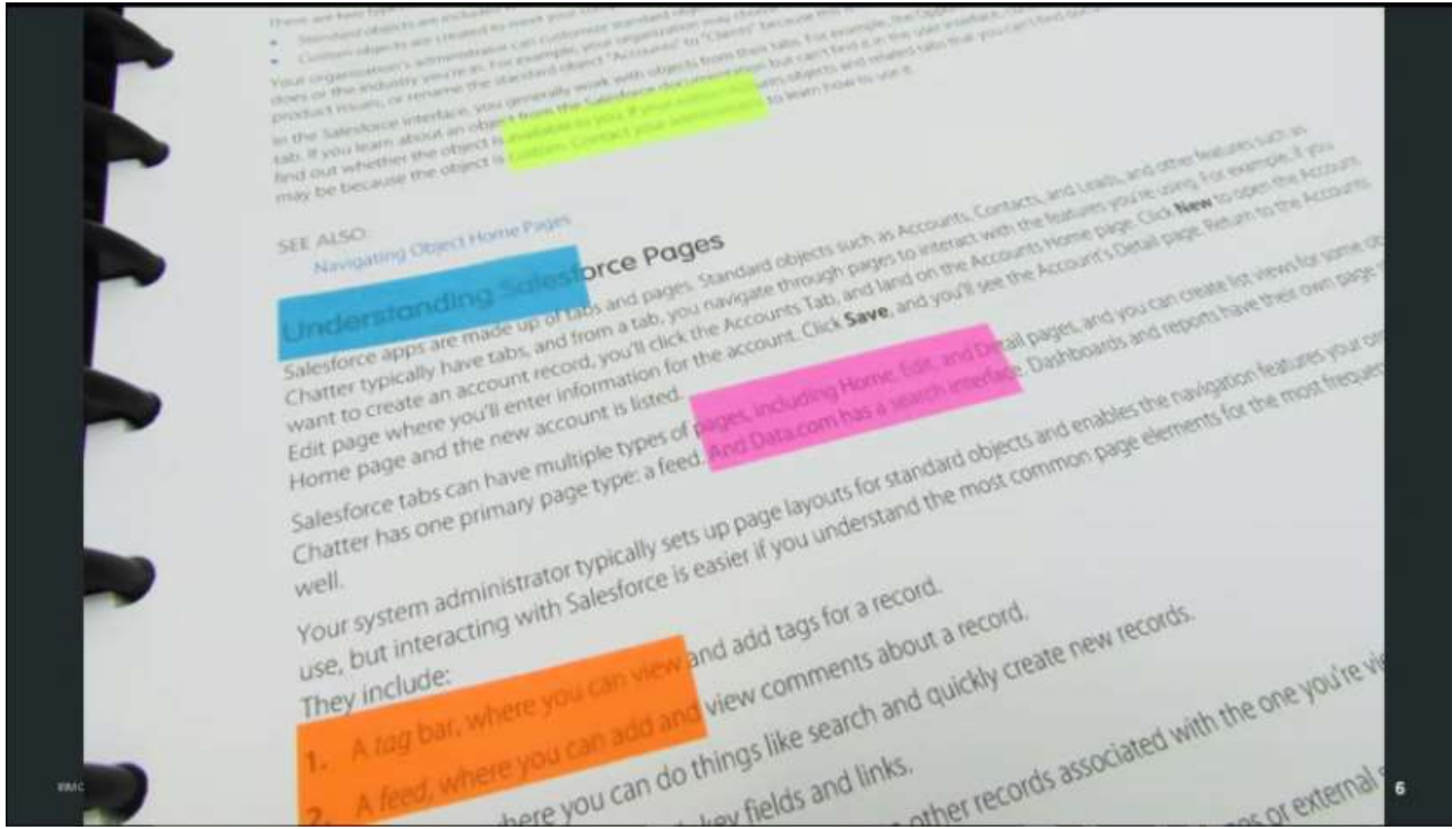


Image created by Angela Swindell, Visual Designer, IBM

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.

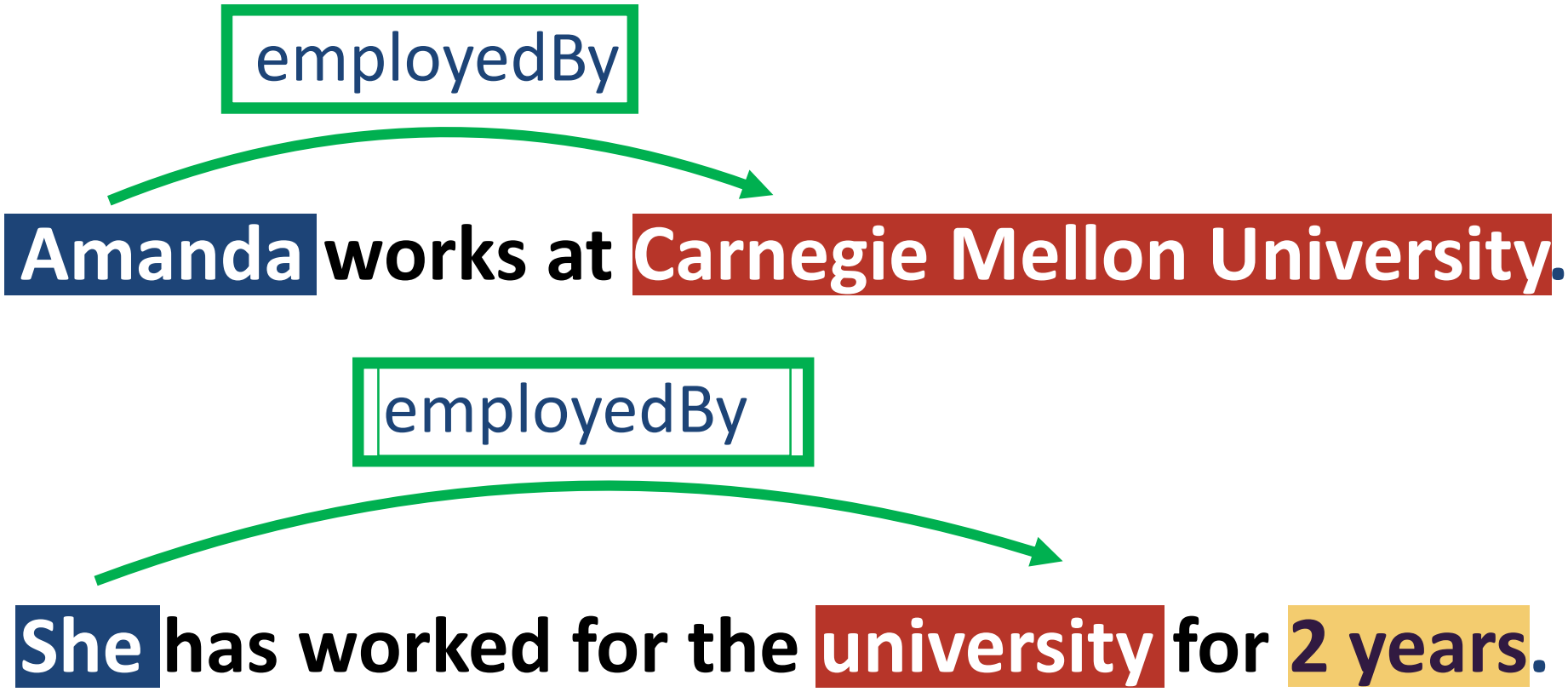
She has worked for the university for 2 years.

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…

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

Training Set and Use

Training Set

Data Encountered

Only as good as data and time spent improving it.

Biased based on what taught.

Concern varies across industries
Accuracy not always best measure!

Use Cases

Consider for each situation

Knowledge needed?

Ethical considerations?

Strategic Games

1997 Chess, IBM

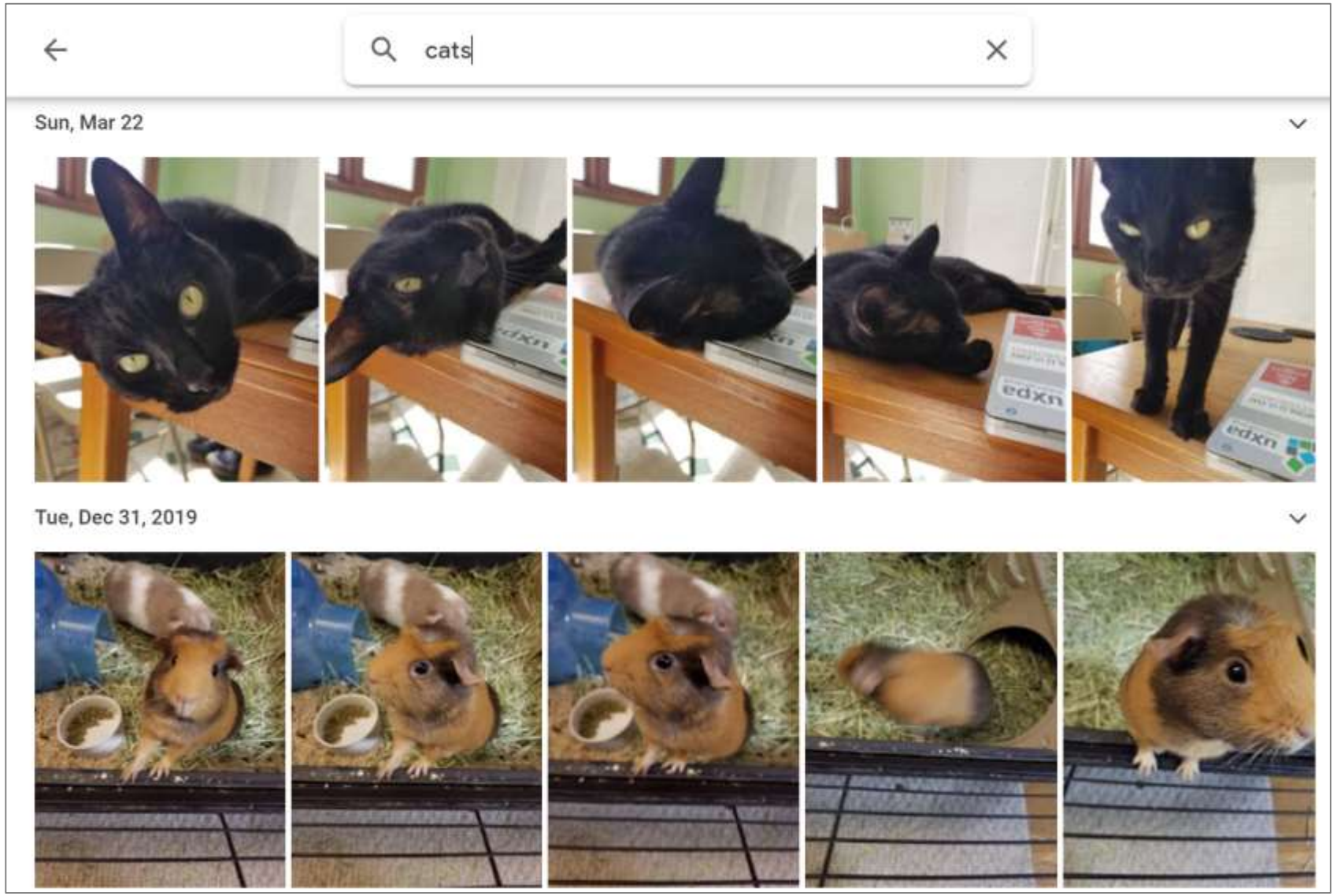
2016 Go, Google

Knowledge?

Ethics?

Floor goban, 2007, By Goban1
https://commons.wikimedia.org/wiki/File:FloorGoban.JPG

Image Recognition – Google Photos



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

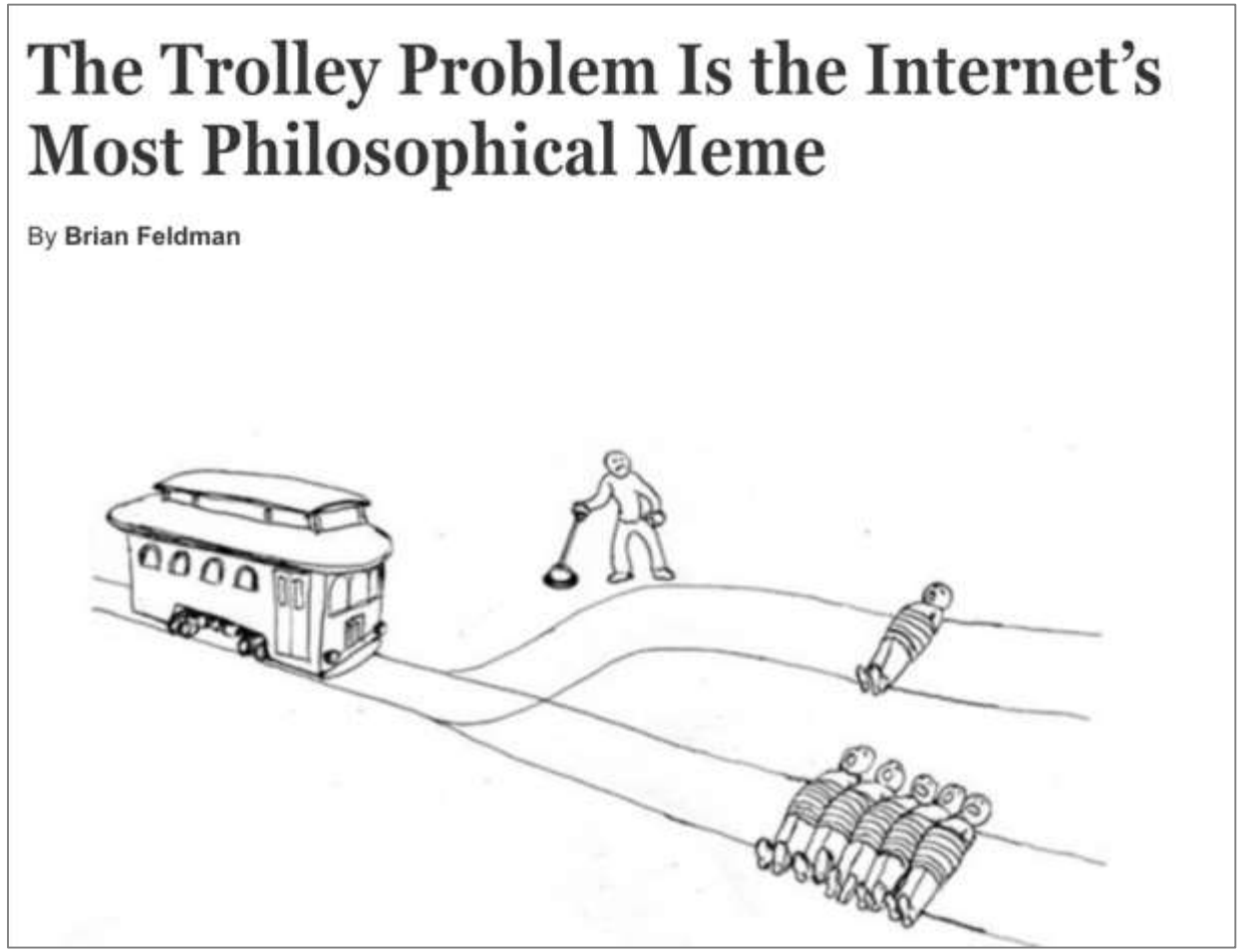
Company B

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

Neither are wrong.
Both are biased.

Our responsibility to keep people safe

Not Trolley Problems (hypotheticals in-the-moment)



The Trolley Problem Is the Internet's Most Philosophical Meme

By **Brian Feldman**

The Trolley Problem Is the Internet's Most Philosophical Meme By Brian Feldman August 9, 2016
http://nymag.com/selectall/2016/08/trolley-problem-meme-tumblr-philosophy.html

Should we trust machines as much as a well-trained human?

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

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

Adopt Technology Ethics

What do you value?

What lines won't you cross?

Track progress

Association for Computing Machinery

AINOW INSTITUTE

Microsoft

Google

Montréal Declaration Responsible AI_

An initiative of Université de Montréal

U.S. DEPT OF DEFENSE

Use a Checklist – find concealed tasks

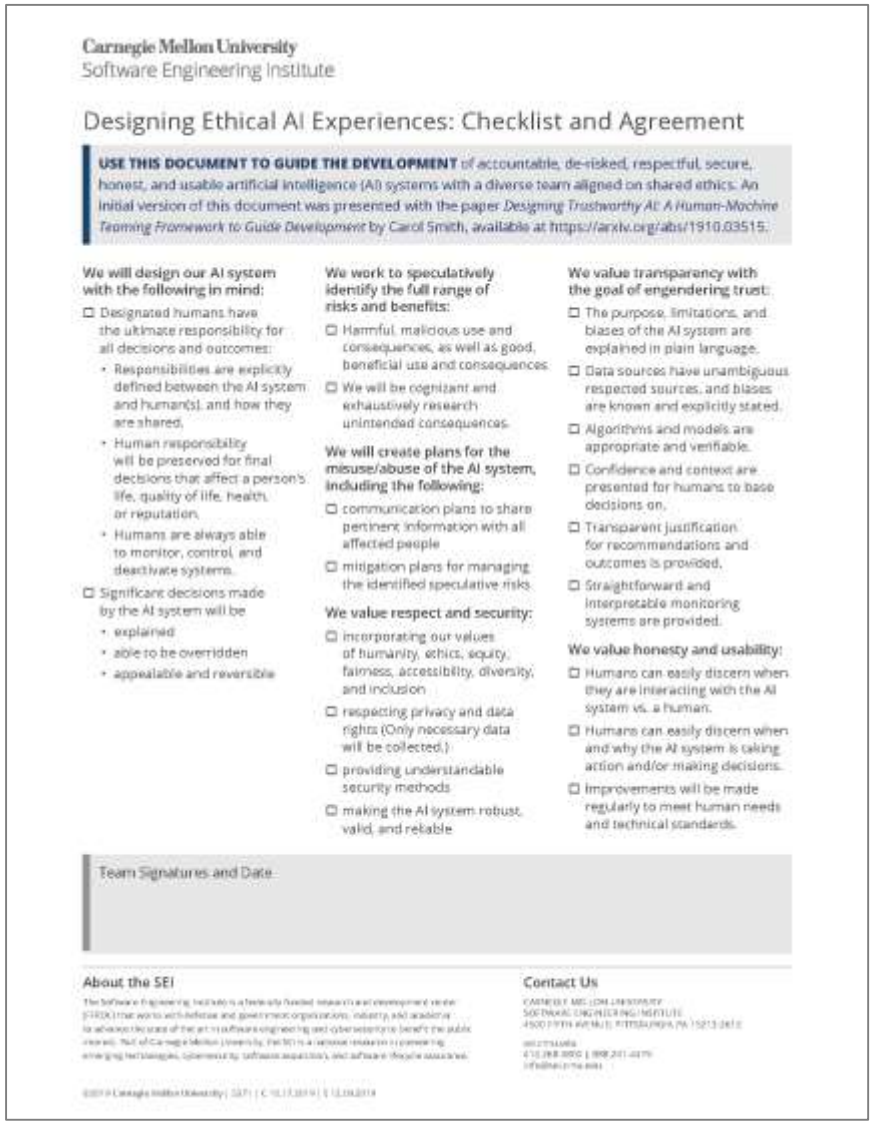
Pair with Tech Ethics

- Bridge 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.

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

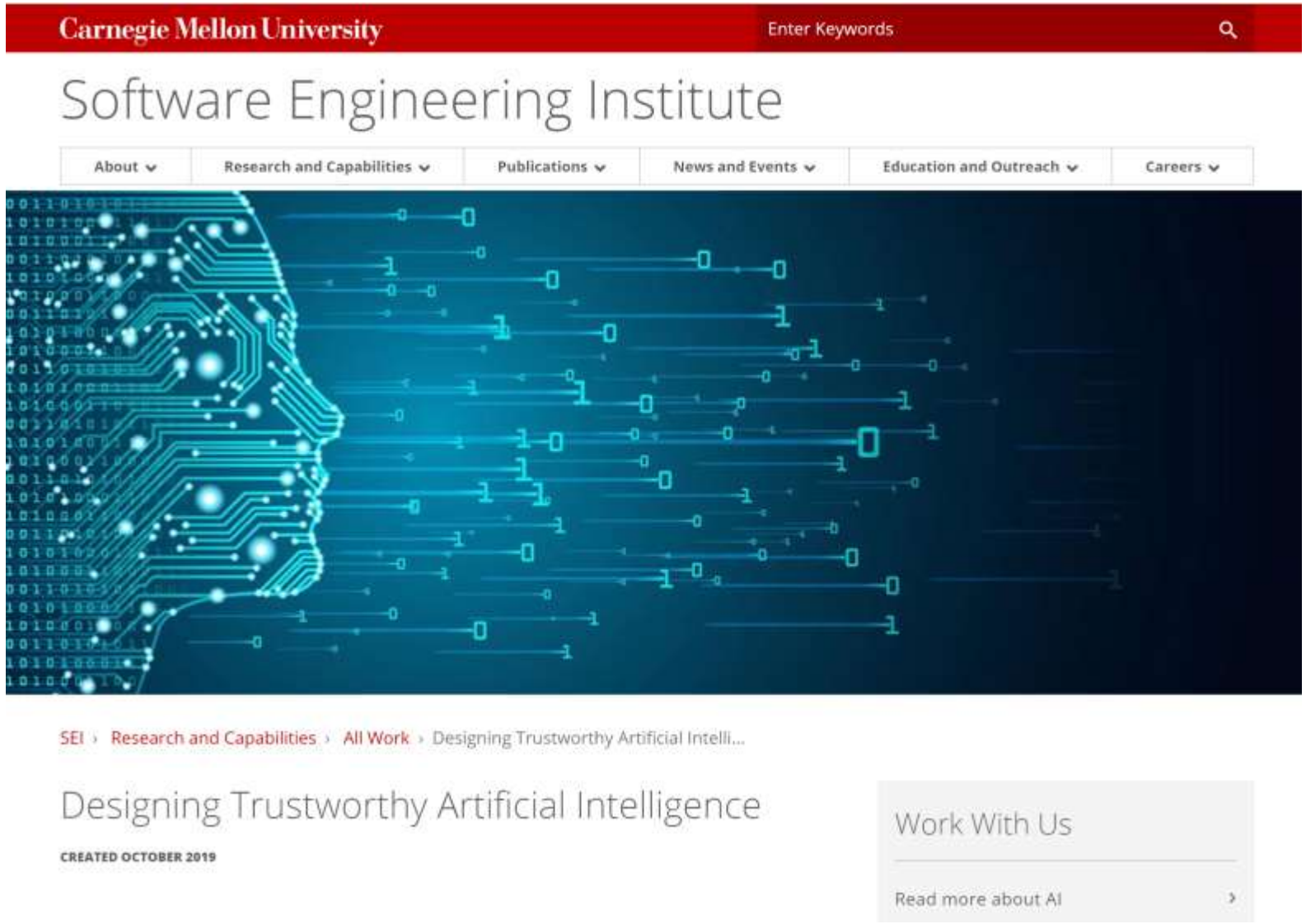
Contact Us

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

Don’t fear AI - Explore AI

Try out tools
Pair with others

Learn More



SEI on HMT: https://sei.cmu.edu/research-capabilities/all-work/display.cfm?customel_datapageid_4050=197910

Carol J. Smith

Twitter: @carologic

SEI Emerging Technology Center

Email: info@sei.cmu.edu

Twitter: @sei_etc