

The logo for Carnegie Mellon University, featuring the text "Carnegie Mellon University" in a white serif font. The background of the slide is a dark blue grid with diagonal lines in red, green, and yellow.

**Software Engineering
Institute**

AI Engineering

Thinking through how to build AI better

OCTOBER 2021

Dr. Rachel Dzombak
rdzombak@sei.cmu.edu

Digital Transformation Lead, SEI AI Division

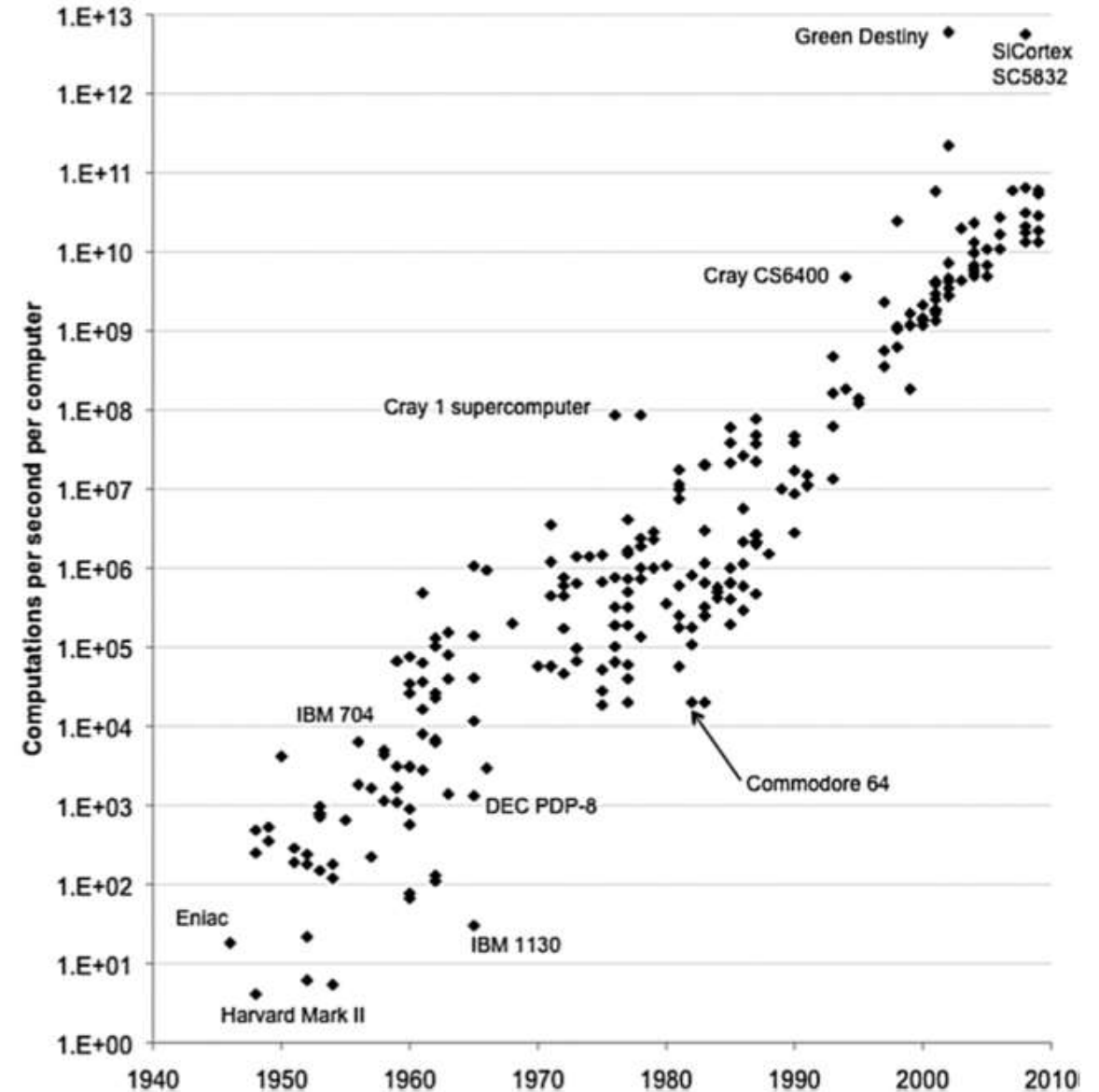
Dr. Rachel Dzombak
Lead, Digital Transformation

CMU Software Engineering Institute
AI Division



Setting the Stage

Basic building blocks of technology have been evolving at an exponential rate for some time.





Today, those basic physical, digital, and biological technologies are intersecting to create even more change.

Which is driving
large scale systems
transformations in
many industries.





We are collectively faced with designing the **systems** of the future accommodating both **technology** AND **people**.

What kind of world do you want to design?





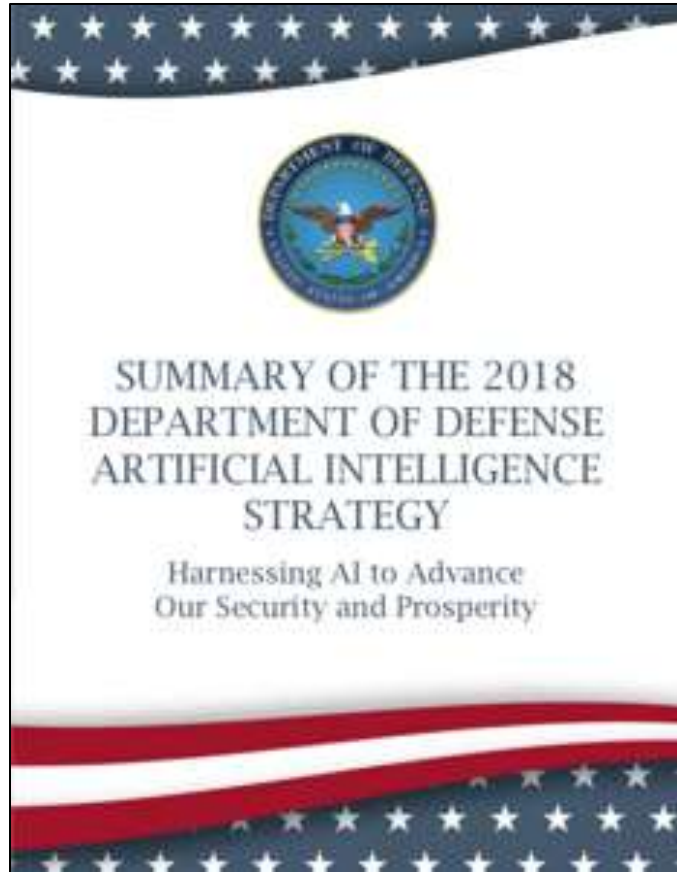
What is AI?



“It is the science and engineering of making intelligence machines, especially intelligent computer programs.”

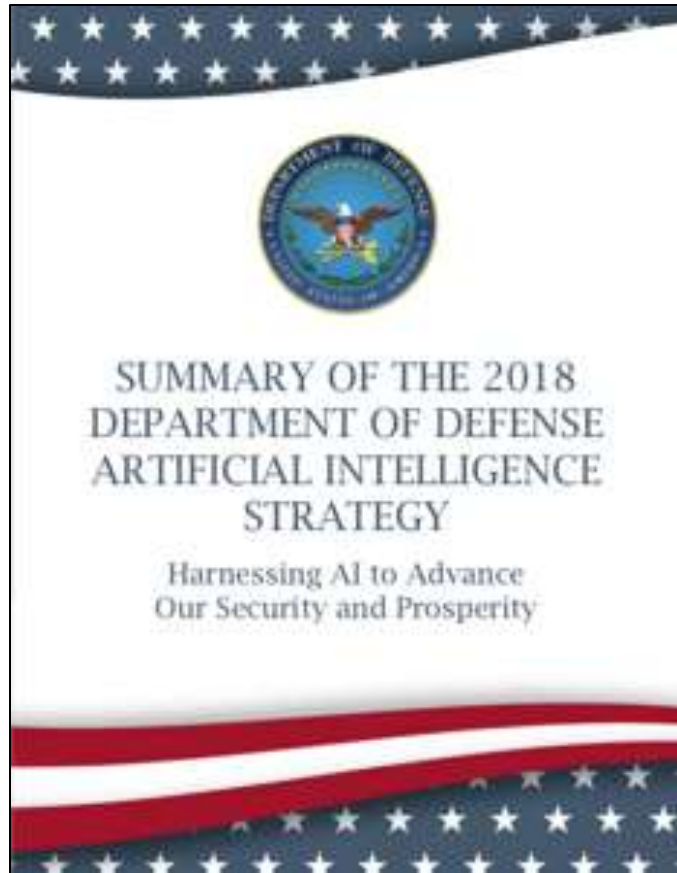
- John McCarthy, MIT, 1956

What is AI?



“AI refers to the ability of machines to perform tasks that normally require human intelligence – for example, recognizing patterns, learning from experience, drawing conclusions, making predictions, or taking action – whether digitally or as the smart software behind autonomous physical systems.”

What is AI?



“AI refers to the ability of machines to perform tasks that normally require human intelligence – for example, **recognizing patterns**, learning from experience, drawing conclusions, **making predictions**, or taking action – whether digitally or as the smart software behind autonomous physical systems.”

What is AI?

medium.com



future of AI," says INFORM ...
postandparcel.info

data-flair.training



The Future of Artificial Intelligence ...
reliabilityweb.com

appypie.com



Future of Artificial Intelligence | Top ...
educba.com

cdotrends.com



AI FOR GOOD - The Future of Work - Y...
youtube.com

towardsdatascience.com



Future of Artificial Intelligence ...
hackr.io



What is the Future of AI? | Know About ...
edureka.co



AI and the Future - Design Engineering
design-engineering.com



Future of Artificial Intelligence ...
eduinpro.com



Artificial Intelligence and Google
fairobserver.com



The Future Of Artificial Intelligence ...
elearningindustry.com



The Future of Labor in an AI World
datanami.com



31 Ways AI Will Affect the Future of ...
cobizmag.com



is AI the Future of Business Intelligence
datafloq.com

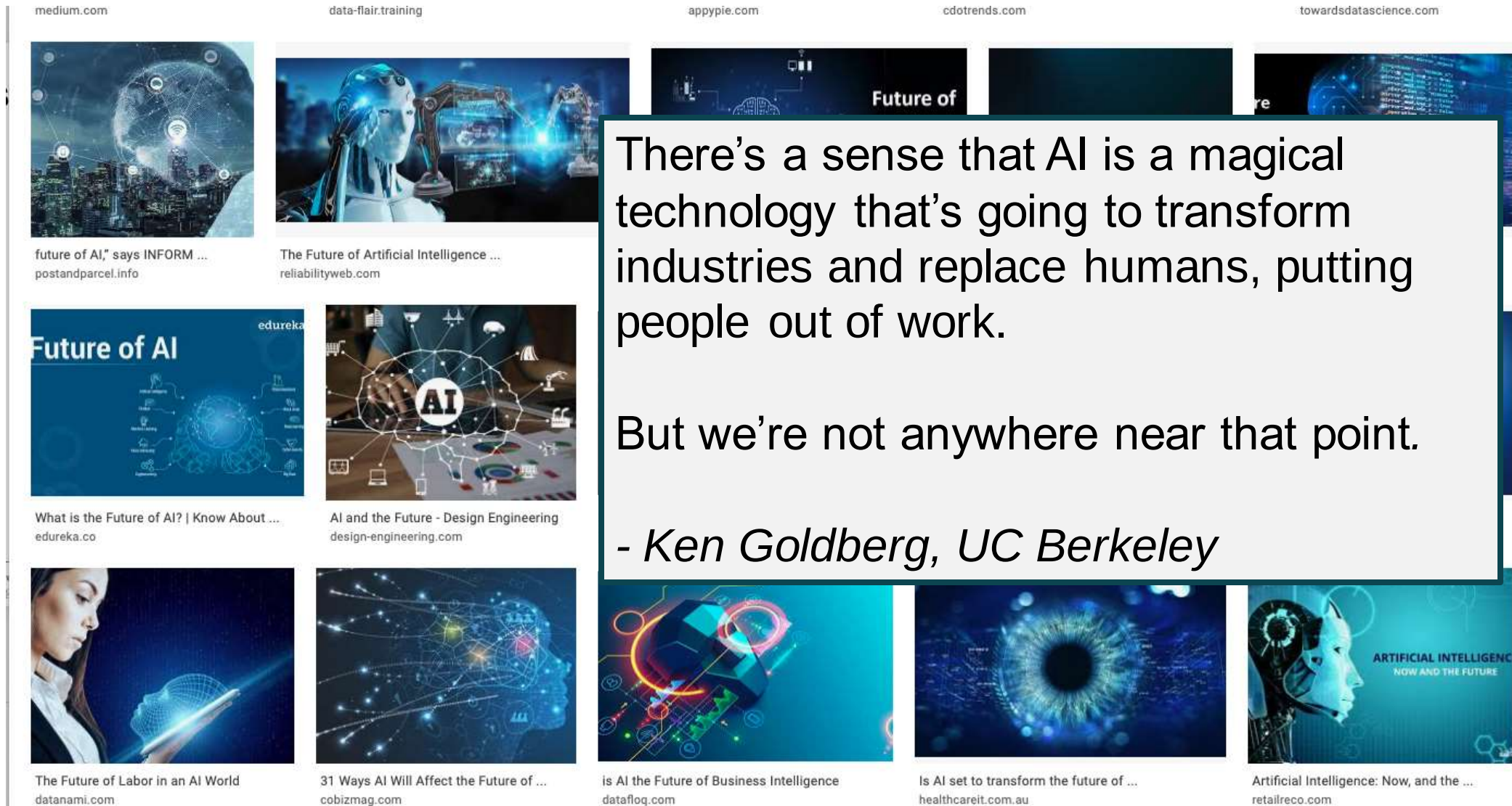


Is AI set to transform the future of ...
healthcareit.com.au

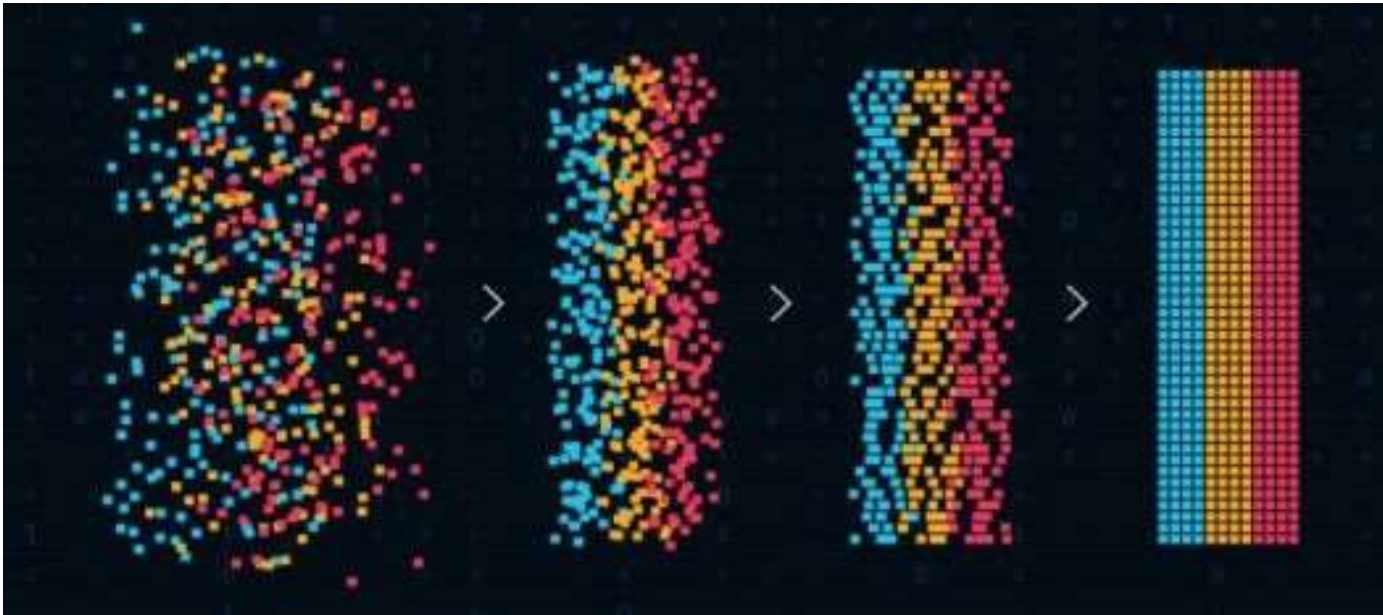


Artificial Intelligence: Now, and the ...
retailreco.com

What is AI?



What is AI?



Machines are very good at precision; they're very good at calculating numbers and pattern recognition.

- *Ken Goldberg, UC Berkeley*

At CMU, we aim to take a comprehensive view of AI development.

Defining
Challenge

Acquiring
Data

Manipulating
Data

Developing
Model

Decision
Making

At CMU, we aim to take a comprehensive view of AI development.

Defining
Challenge

Acquiring
Data

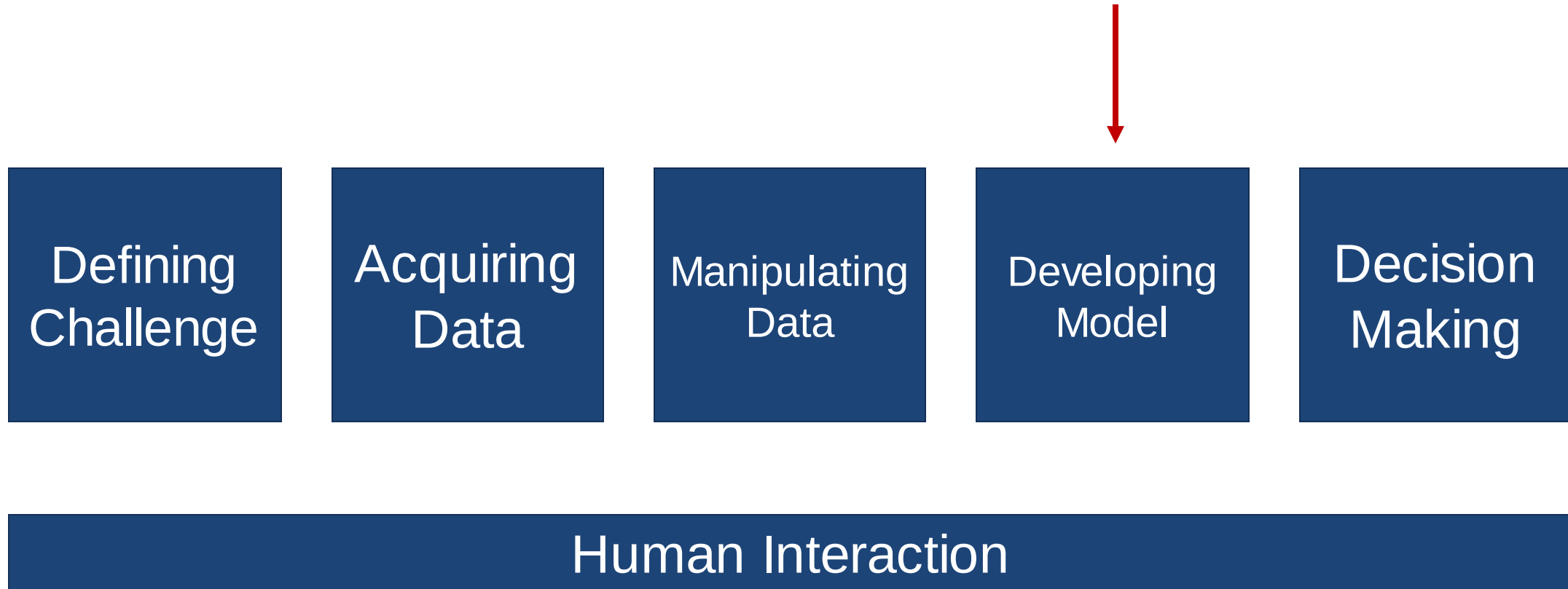
Manipulating
Data

Developing
Model

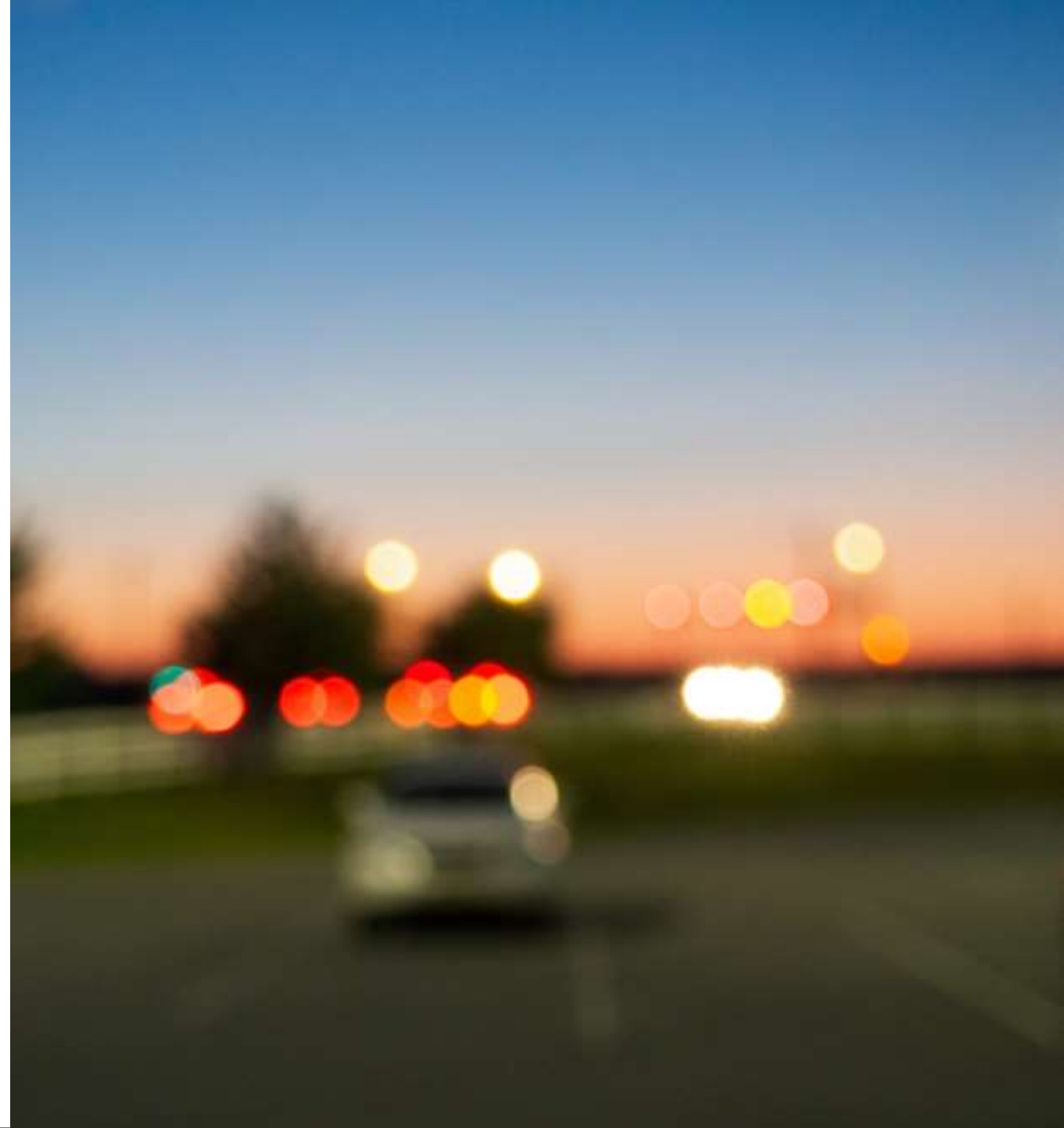
Decision
Making

Human Interaction

At CMU, we aim to take a comprehensive view of AI development.



We want to progress AI from individual tools to human-centered, robust and secure, and scalable systems.



What is AI Engineering?

AI engineering is a field of research and practice that integrates the principles of software engineering, systems, computer science, and human-centered design to create and implement AI systems in accordance with human needs for mission outcomes.



Why AI engineering?

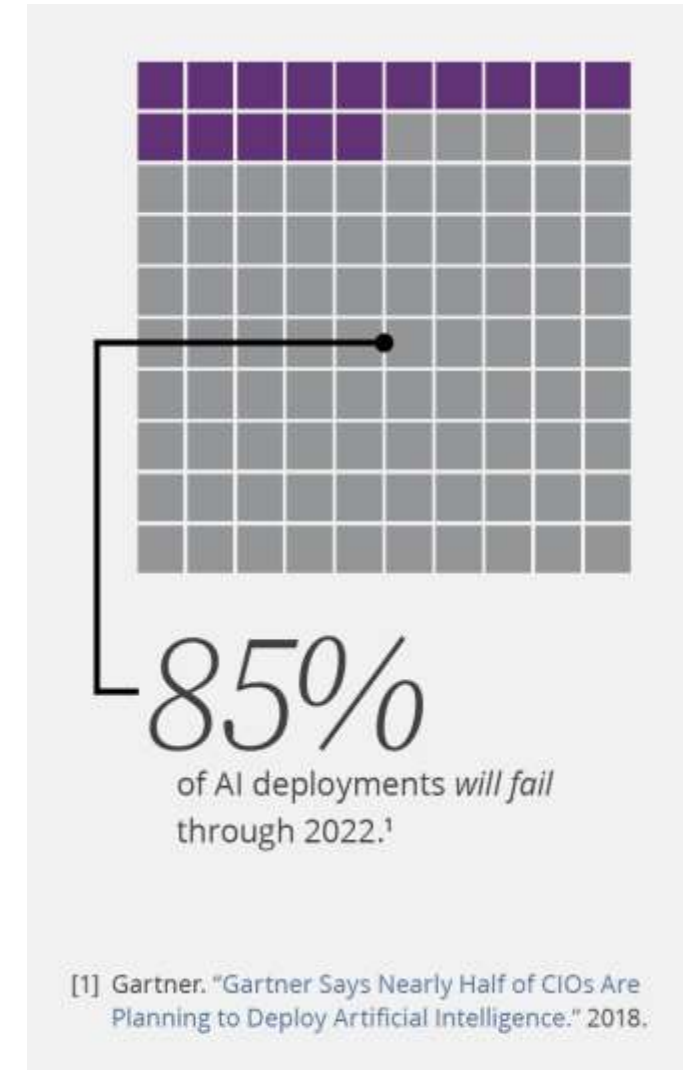
Organizations realize that AI hold great promise and power.

It is hard to get AI right.

Many organizations aren't prepared and don't have the needed expertise.

We are part of CMU – a world leader in AI.

Most work is a race to AI capability.



Why AI Engineering?

Traditional software and system engineering are critical to building reliable AI systems, but there are important differences and gaps.

Many modern AI systems are built using machine learning.

Traditional Software

- Analytical
- Explicit instructions given by programmer
- Reducible and decomposable
- Deterministic

Machine Learning

- Empirical
- Behavior learned from data or experience
- Opaque (and lots of math)
- Unpredictable

“Teaching, not micromanaging” – Peter Norvig

“There is no book of
spells, there’s just magic.”

“There is no book of
spells, there’s just magic.”

... of course, AI isn’t magic.

“There is no book of
spells, there’s just magic.”

... of course, AI isn’t magic.

We believe there are best practices,
processes, tools, and frameworks that
can improve deployment of AI and enable
trust and confidence – our National
Initiative aims to define and share them.

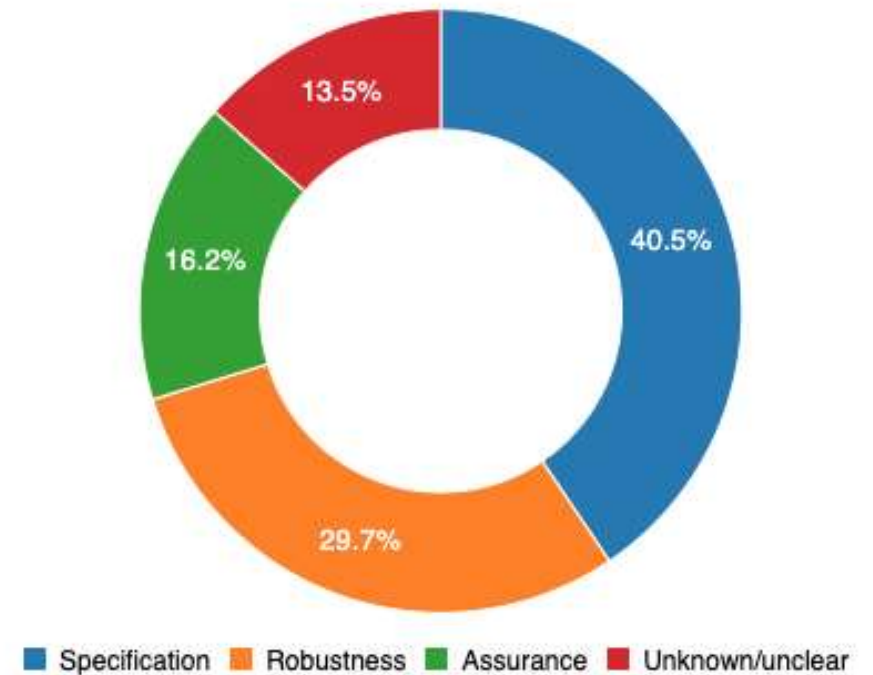
What factors cause AI system “Incidents”?

Failures in...

Specification: the system's behavior did not align with the true intentions of its designer, operator, etc.

Robustness: the system operated unsafely because of features or changes in its environment, or in the inputs the system received

Assurance: the system could not be adequately monitored or controlled during operation



74 total incidents

Source: <https://incidentdatabase.ai/taxonomy/cset>

Credit to Partnership on AI and the Center for Security and Emerging Technologies (CSET) at Georgetown University

AI Engineering Pillars

	Scalable AI <i>Accommodate the size, speed, and complexity of mission needs</i>	• Scalable management of data and models • Enterprise scalability of AI development and deployment • Scalable algorithms and infrastructure
	Robust and Secure AI <i>Operate reliably when faced with uncertainty or threat</i>	• Robustness of AI components and systems • Designing for security challenges in modern AI systems • Testing, evaluating, and analyzing AI systems
	Human-Centered AI <i>Designed with the goal of working with, and for, people</i>	• Understand context of use, sense changes over time • Scope and facilitate human-machine teaming • Methods, mechanisms, and mindsets for critical oversight

Human-Centered AI



Pair Checklist with Ethical Principles.
Reduce risk and unwanted bias.
Support inspection and mitigation planning.

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

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 substantively 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 (DFRC) that serves as a national center for research and development in software engineering and computer systems. The SEI is a national center for research and development in software engineering and computer systems. The SEI is a national center for research and development in software engineering and computer systems.

Contact Us
Carnegie Mellon University
Software Engineering Institute
4400 Fifth Avenue, Pittsburgh, PA 15213-1502
412.268.1500
412.268.1500
412.268.1500

©2019 Carnegie Mellon University | 0001 | 0.1.0.0.0001 | 0.1.0.0.0001

Robust and Secure AI

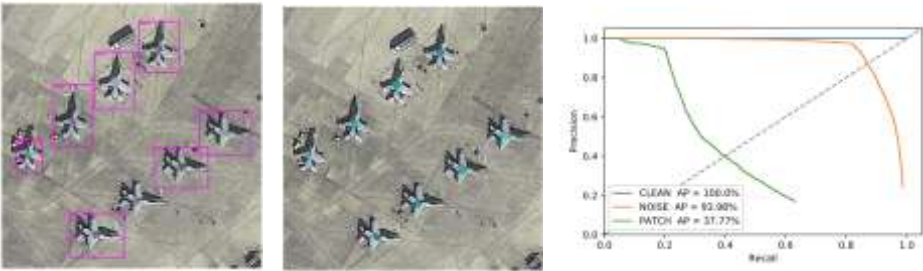


Learn the wrong thing



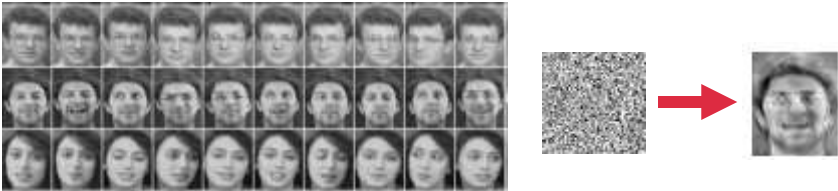
(Gu et al., 2017)

Do the wrong thing



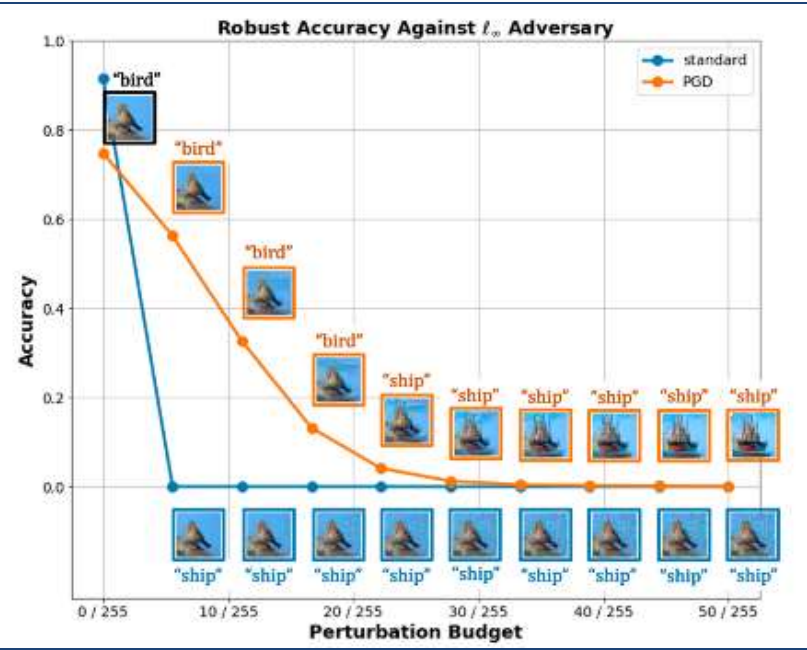
(Adhikari et al., 2020)

Reveal the wrong thing



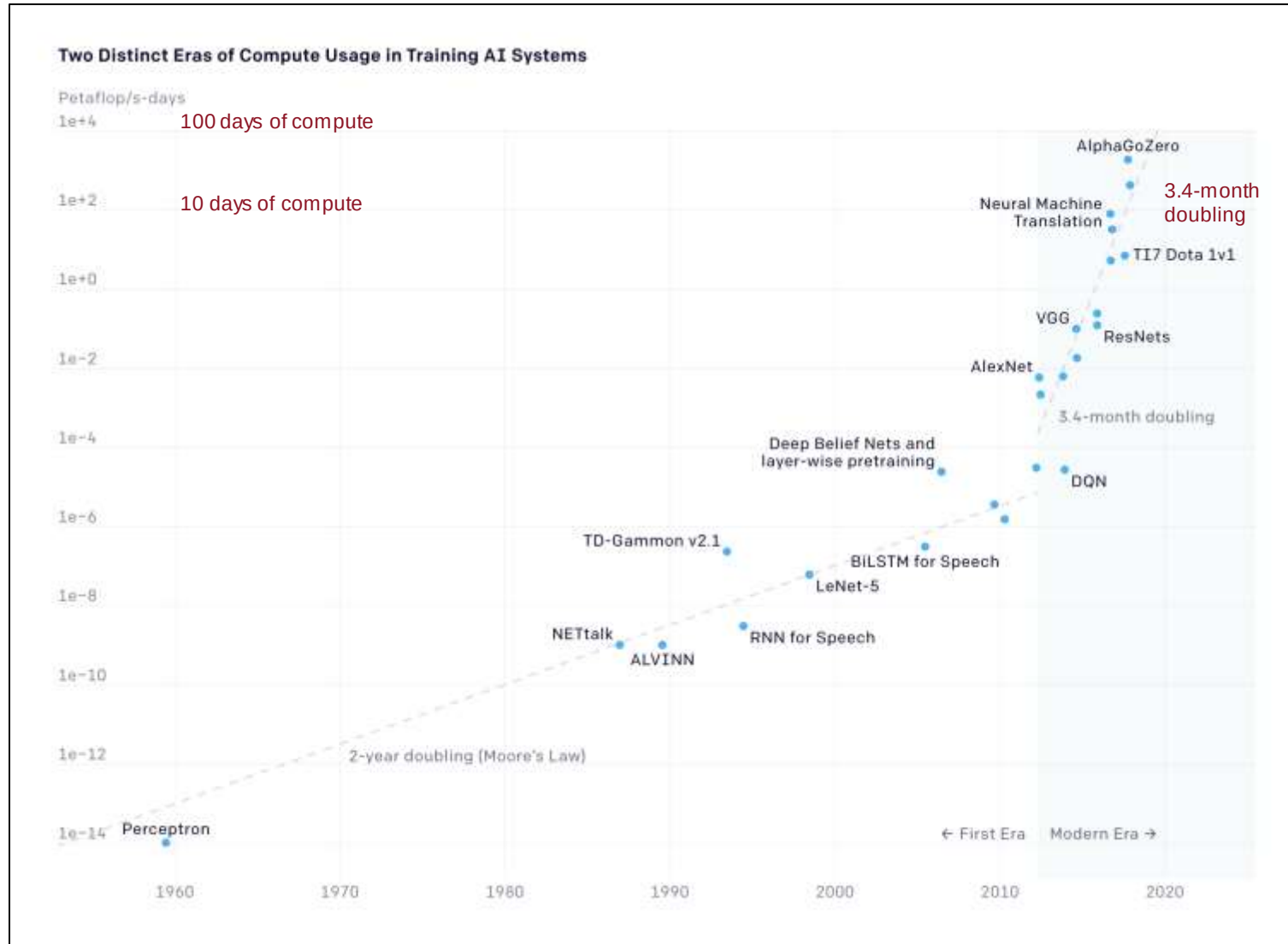
(Fredrickson et al., 2015)

Train / Verify	Learn	Do	Reveal
Learn			
Do			
Reveal			



(VanHoudnos, et al., 2020)

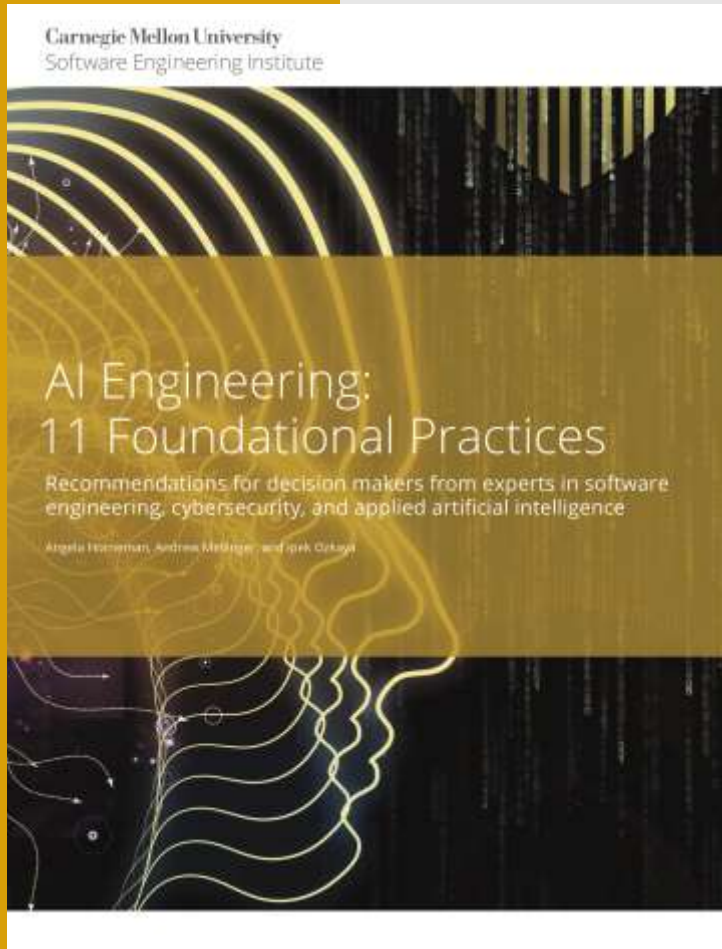
Scalable AI



Black Hornet Nano

OpenAI: AI and Compute, May 2018.
<https://openai.com/blog/ai-and-compute/>

Thompson et al., "The Computational Limits of Deep Learning," 2020. <https://arxiv.org/pdf/2007.05558.pdf>



Download Today



[resources.sei.cmu.edu/
library/asset-view.cfm?
assetid=633647](https://resources.sei.cmu.edu/library/asset-view.cfm?assetid=633647)

For more information, write to info@sei.cmu.edu

Authors

Angela Horneman, Analysis Team Lead
Carnegie Mellon University Software Engineering Institute

Andrew Mellinger, Sr. Software Developer
Carnegie Mellon University Software Engineering Institute

Ipek Ozkaya, Principal Researcher
Carnegie Mellon University Software Engineering Institute

Available for Download Today

AI Engineering: 11 Foundational Practices

“Developing viable and trusted AI systems that are deployed to the field and can be expanded and evolved for decades requires significant planning and ongoing resource commitment.”

1. Ensure you have a problem that both can and should be solved by AI.

2. Include highly integrated subject matter experts, data scientists, and data architects in your software engineering teams.

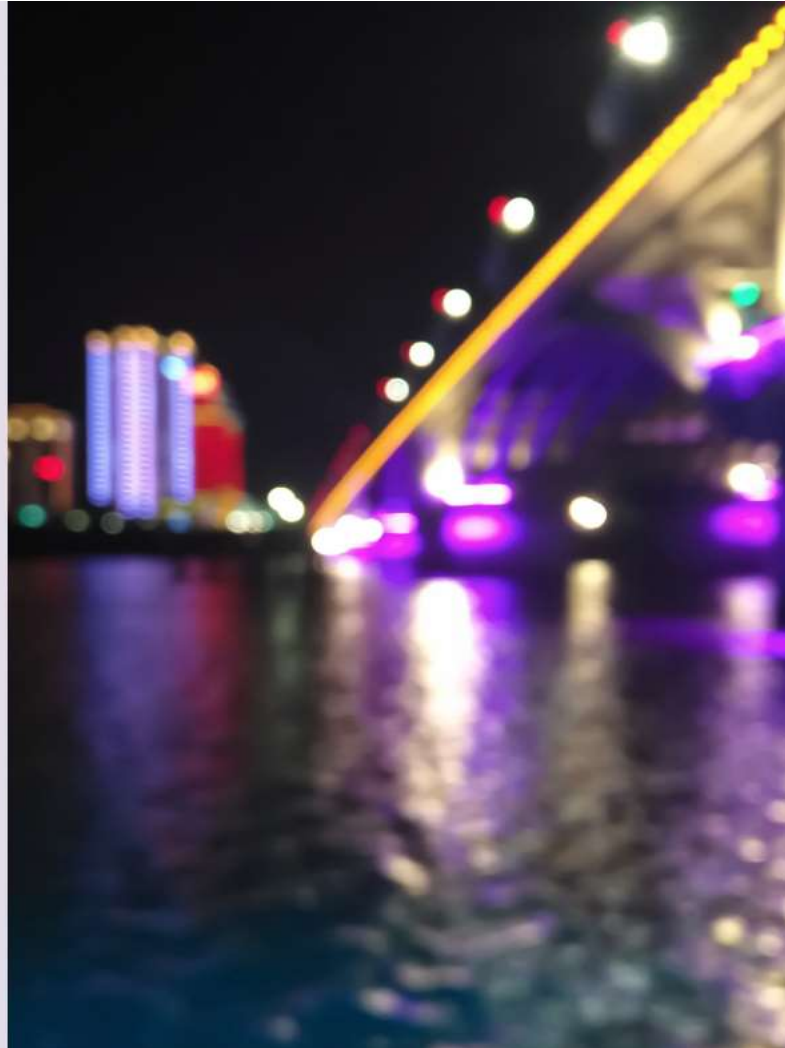
11. Treat ethics as both a software design consideration and a policy concern.

National AI
Engineering
Initiative

Carnegie
Mellon
University
Software
Engineering
Institute

AI Engineering

An Emergent Discipline for
Human-Centered, Robust and
Secure, and Scalable AI



Advocate for
AI Engineering



Collaborate to Build
the Discipline



Support the
Research Agenda

<https://www.sei.cmu.edu/our-work/artificial-intelligence-engineering/>