Human-Centered AI: Key Considerations for Adopting AI Systems in Workflows

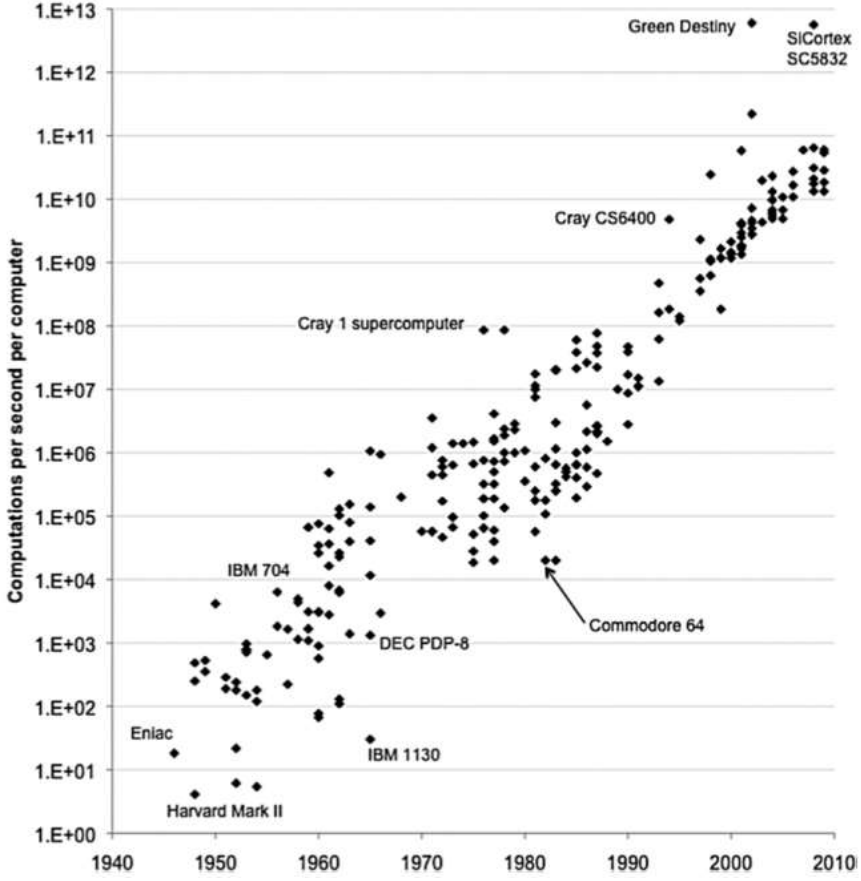
Alex Van Deusen
Design Researcher, CMU SEI
arvandeusen@sei.cmu.edu

Rachel Dzombak
Digital Transformation Lead, CMU SEI
rdzombak@sei.cmu.edu

Software Engineering Institute
Carnegie Mellon University
Pittsburgh, PA 15213

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.

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?

“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 AI engineering work aims to define and share them.

Why AI engineering?

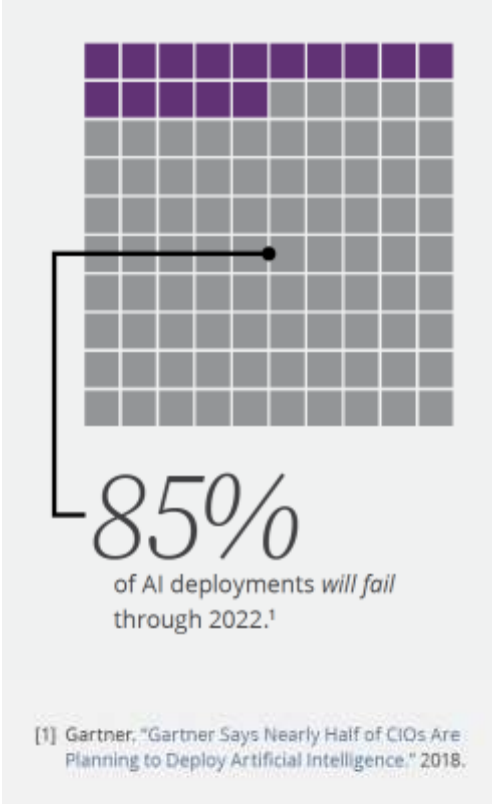
Organizations realize that AI hold great promise and power.

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.



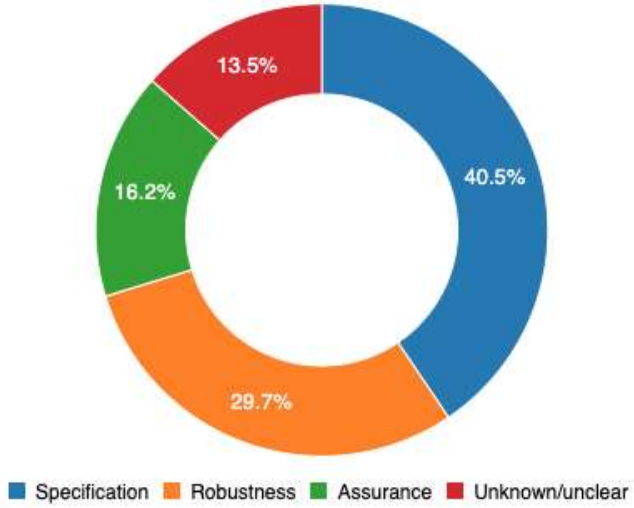
What factors cause AI system “Incidents”?

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

AI Engineering Pillars

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

How can teams harness the **power** of AI systems and design them to be **valuable** to humans?

Advancing human-centered AI

- Designers and systems must understand the context of use and sense changes over time.
- Methods, mechanisms, and mindsets to engage in critical oversight.
- Development of tools, processes, and practices to scope and facilitate human-machine teaming.

Human-Centered AI, Software Engineering Institute: https://resources.sei.cmu.edu/library/asset-view.cfm?assetid=735362

Where to Start with AI

Understand the context of use

Responsible, intentional design

- What problem are you solving?
- For whom?
- What will help them?
- What kind of improvements are expected?
- What might a machine do better or faster?
- What happens if the AI system is *not* used?
- What is *not* going to be improved (out of scope)?

Mindset… comfortable with testing unfinished ideas in order to make the best result possible.

–Kathryn McElroy

Prototyping for Designers: Developing the Best Digital and Physical Products
by Kathryn McElroy, IBM Design Thinking:
https://www.ibm.com/design/thinking/principles/restless-reinvention

Where to Start with AI

Engage in critical oversight

Ethics

Ethics are standards of expected behavior that guide the correct course of action.

- Based on well-founded standards of right and wrong
- Necessary to constantly examine one's standards to ensure that they are reasonable and well-founded
- “Ethical considerations are an inseparable part of research, design, and deployment for DoD AI systems.” –Defense Innovation Board
- What impact does my work have?

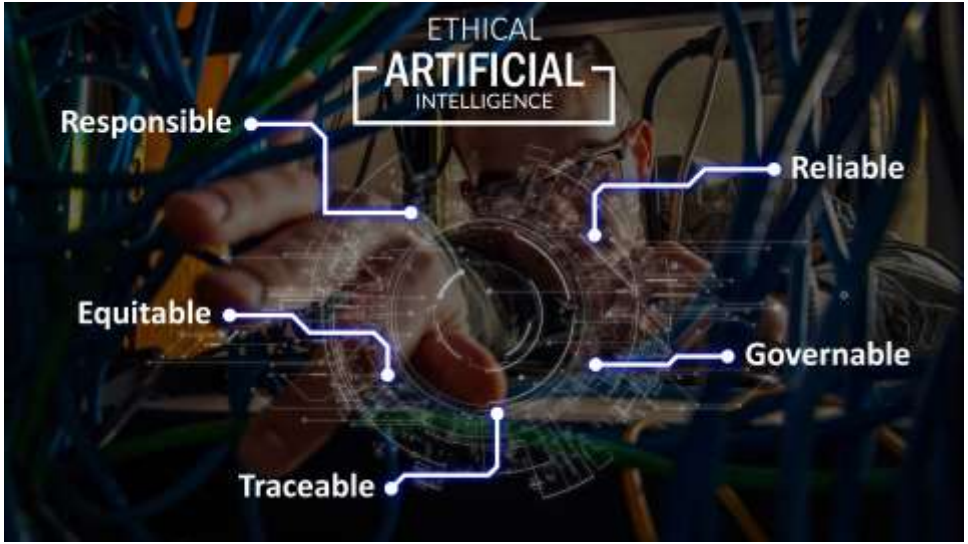
Team alignment on technical ethics…

- Improves consistency
- Provides goals and direction
- Builds on existing ethics
- Counters industry pressure
- Provides expectation and explicit permission to consider and question breadth of implications

Coalesce on a shared set of technical ethics

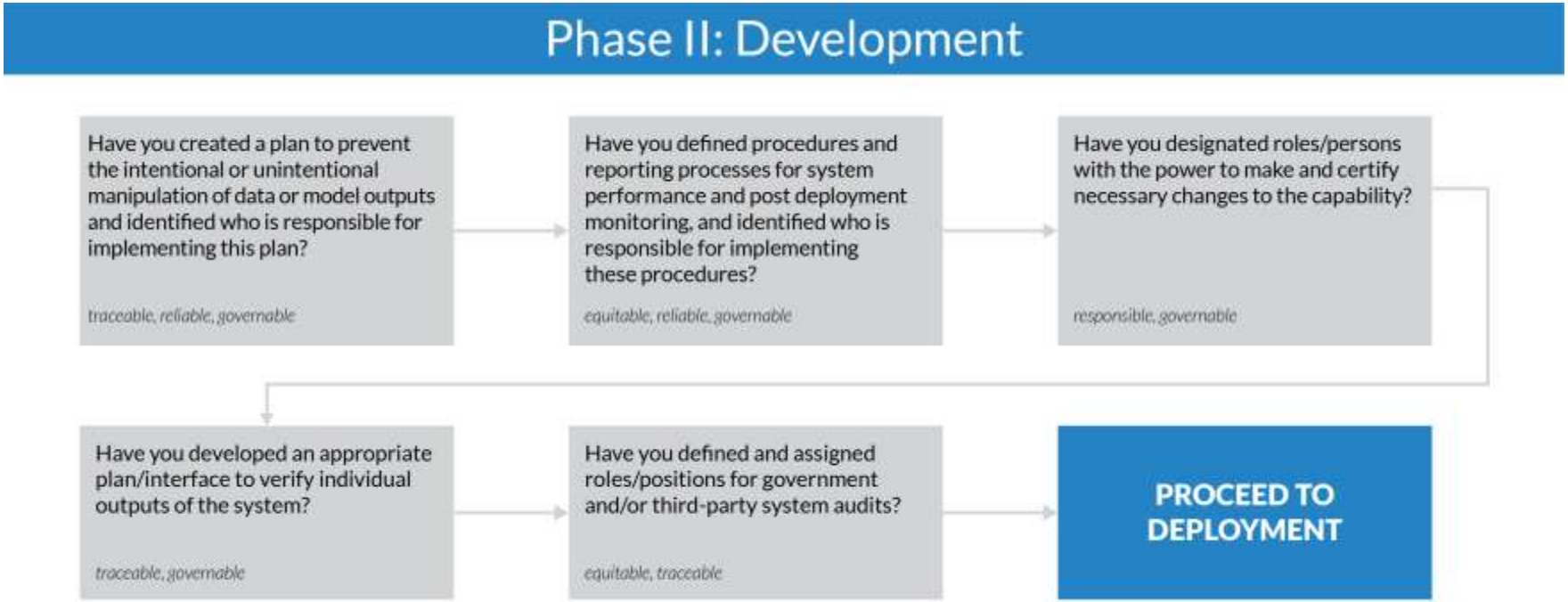
Department of Defense Ethical Principles for Artificial Intelligence

- Responsible
- Equitable
- Traceable
- Reliable
- Governable



Content adapted from Department of Defense Ethical Principles for Artificial Intelligence.
Department of Defense Adopts Ethical Principles for Artificial Intelligence - FEB. 24, 2020
https://www.defense.gov/Newsroom/Releases/Release/Article/2091996/dod-adopts-ethical-principles-for-artificial-intelligence/

Integrate your technical ethics with the development process



DEFENSE INNOVATION UNIT

Jared Dunnmon, Bryce Goodman, Peter Kirechu, Carol Smith, Alexandrea Van Deusen. "Responsible AI Guidelines in Practice: Lessons Learned from the DIU AI Portfolio." Defense Innovation Unit. https://www.diu.mil/

Start conversations by asking questions

- What do we value?
- Who could be hurt?
- How will we track our progress?
- What are potential unintended or unwanted consequences?

Incorporate user experience (UX) research and human-computer interaction (HCI) methods to start conversations

- **Identify problems** in the design
- **Uncover opportunities** to improve
- **Learn about the target user's** behavior and preferences

Test application of ethics frequently

- Pick a set of ethics that works well for you and your team, adjust as necessary
- Reduce risk and unwanted bias
- Support inspection and mitigation planning


Carnegie Mellon University
Software Engineering Institute

Designing Ethical AI Experiences: Checklist and Agreement

USE THIS DOCUMENT TO GUIDE THE DEVELOPMENT of accountable, de-biased, 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 chained.
  - 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, override, 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 important 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

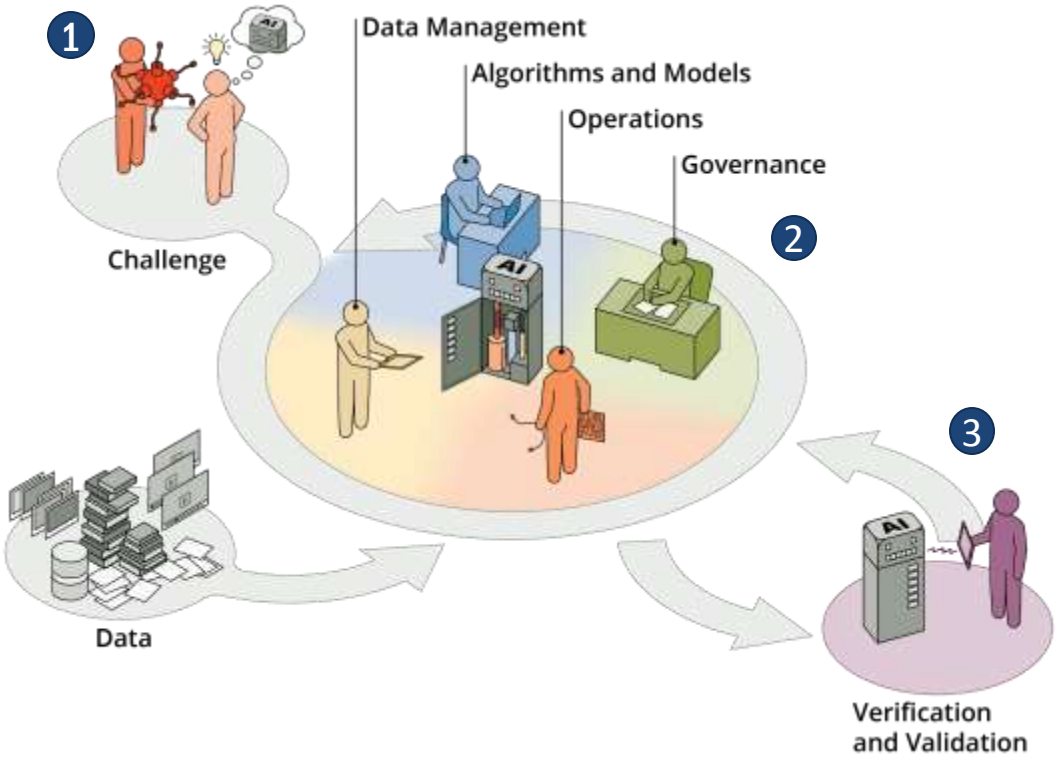
Contact Us

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

Where to Start with AI

Facilitate human-machine teaming

AI Systems and Human-Machine Teaming



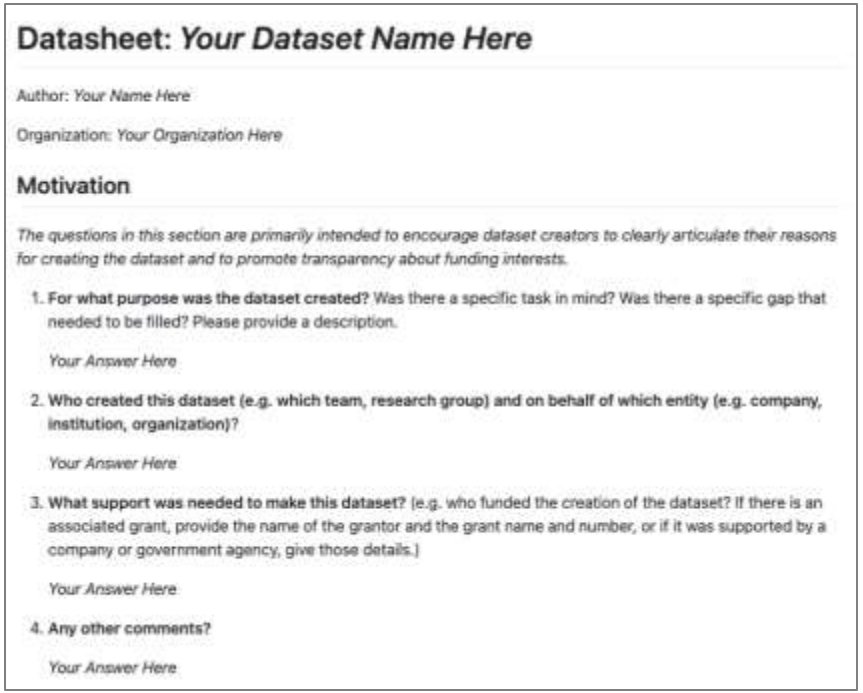
1. Focus on user needs
2. Align on technical ethics
3. Make AI interpretable, understandable, verifiable

Build confidence in AI systems

- AI capabilities should be interpretable
- Build confidence and trust for people using AI systems
- Technology and methodologies should be explained appropriately
  - Access to details
  - Rationale for decisions and recommendations
- Data should be understandable and explainable
  - Provenance
  - Curator's motivation, composition, and collection

*Datasheets for Datasets. Working Paper by Timnit Gebru, Jamie Morgenstern, Briana Vecchione, Jennifer Wortman Vaughan, Hanna Wallach, Hal Daumé III, Kate Crawford https://arxiv.org/abs/1803.09010

Datasheets for Datasets provides transparency and clarity on sources of data



Datasheet: *Your Dataset Name Here*

Author: *Your Name Here*

Organization: *Your Organization Here*

Motivation

The questions in this section are primarily intended to encourage dataset creators to clearly articulate their reasons for creating the dataset and to promote transparency about funding interests.

1. **For what purpose was the dataset created?** Was there a specific task in mind? Was there a specific gap that needed to be filled? Please provide a description.

   Your Answer Here

2. **Who created this dataset (e.g. which team, research group) and on behalf of which entity (e.g. company, institution, organization)?**

   Your Answer Here

3. **What support was needed to make this dataset?** (e.g. who funded the creation of the dataset? If there is an associated grant, provide the name of the grantor and the grant name and number, or if it was supported by a company or government agency, give those details.)

   Your Answer Here

4. **Any other comments?**

   Your Answer Here

Timnit Gebru, Jamie Morgenstern, Briana Vecchione, Jennifer Wortman Vaughan, Hanna Wallach, Hal Daumé III, and Kate Crawford. 2020. Datasheets for Datasets. arXiv:1803.09010 [cs] (March 2020). Retrieved November 12, 2021 from http://arxiv.org/abs/1803.09010

Designing AI systems to team with humans

- Consider how humans team with AI, and are able to monitor systems and control risk
- Designate responsibility to humans for critical decisions and outcomes
- Significant decisions made by the AI system should be:
  - Appealable
  - Able to be overridden
  - Reversable

Development and deployment of human-centered AI systems is challenging

- State of the art in AI technology is continuously evolving
- Intended level of interdependence between humans and machines leads to trust and transparency challenges
- AI systems will be deployed alongside humans in unfamiliar and unpredictable operational contexts
- Guidelines and heuristics are needed for understanding and implementing the level of oversight needed to create and maintain ethical AI systems

Human-Centered AI, Software Engineering Institute: https://resources.sei.cmu.edu/library/asset-view.cfm?assetid=735362

Empower diverse teams in inclusive environments.

Encourage deep conversations, speculation, and imaginative thinking.

Alex Van Deusen
Design Researcher, CMU SEI
arvandeusen@sei.cmu.edu

Rachel Dzombak
Digital Transformation Lead, CMU SEI
rdzombak@sei.cmu.edu