Workshop, UXLx 2023

Designing Complex, Trustworthy Experiences

24 MAY 2023

Carol J. Smith
Sr. Research Scientist, AI Division

Carnegie Mellon University Software Engineering Institute

Copyright Statement

Carnegie Mellon University Software Engineering Institute

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.

Welcome!

10 min - Welcome

20 min – Topic Overview

40 min – Activity: Trustworthy Design

20 min – Research & Speculation

30 min – Break

Timing is approximate

20 min - Frameworks

60 min – Activity: Abusability Testing

10 min - Take action and Close

Designing Complex, Trustworthy Experiences

Carnegie Mellon University Software Engineering Institute

Rapid Adoption of Emerging Technologies

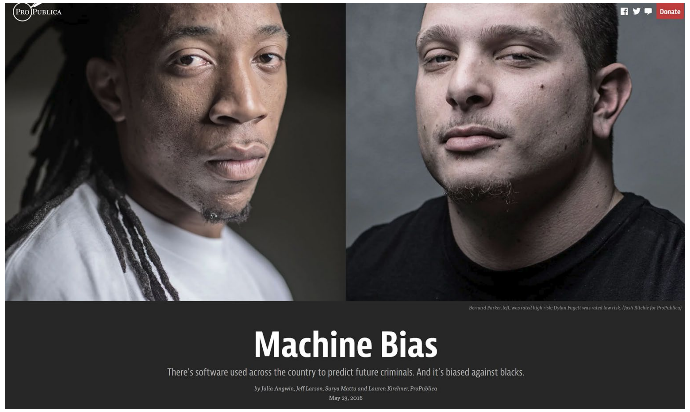
Lack of research results in…

Widely used algorithm for follow-up care in hospitals is racially biased, study finds
PBS News Hour Weekend

Amazon scraps secret AI recruiting tool that showed bias against women
Reuters

The IRS directed 7 million Americans to sign up with ID.me face-scan service, according to congressional letter

The Washington Post



Berkeley News
Mortgage algorithms perpetuate racial bias in lending, study finds

The New York Times
Thermostats, Locks and Lights: Digital Tools of Domestic Abuse

THE WALL STREET JOURNAL.
Slack Backtracks on New Way to Message Over Harassment Concern

Make complex systems trustworthy

User Experience Honeycomb
Peter Morville, et al.

Trustworthy
Systems

Complexity

Sources of Complexity

- Human's capabilities
- Environmental context
- Information
- Society, government
- System design

DIFFERENT NEEDS, GOALS, PURPOSE…

Dynamic systems, complex situations

Video: Tesla Autopilot in Heavy LA Traffic by Scott Kubo https://youtu.be/m3-QzTFxoUg?t=14


Tools Scroll Evaluation Options System Help
0 Worklist item(s) received

Trust and Decision Making

Carnegie Mellon University Software Engineering Institute

Does trust really matter?

Trust

Trust is complex, transient, and personal.

Trust is a psychological state

- Use evidence to determine risk.
- Gained with enough confidence in positive outcomes, to give control of something significant, to the system.

Trust is contextual

Calibrated based on personal experiences, current context, and the available evidence of the system's capability and integrity.

Distrust
Trust falling short of system capabilities - may lead to disuse.

Calibrated Trust
Trust matches system capabilities - leading to appropriate use.

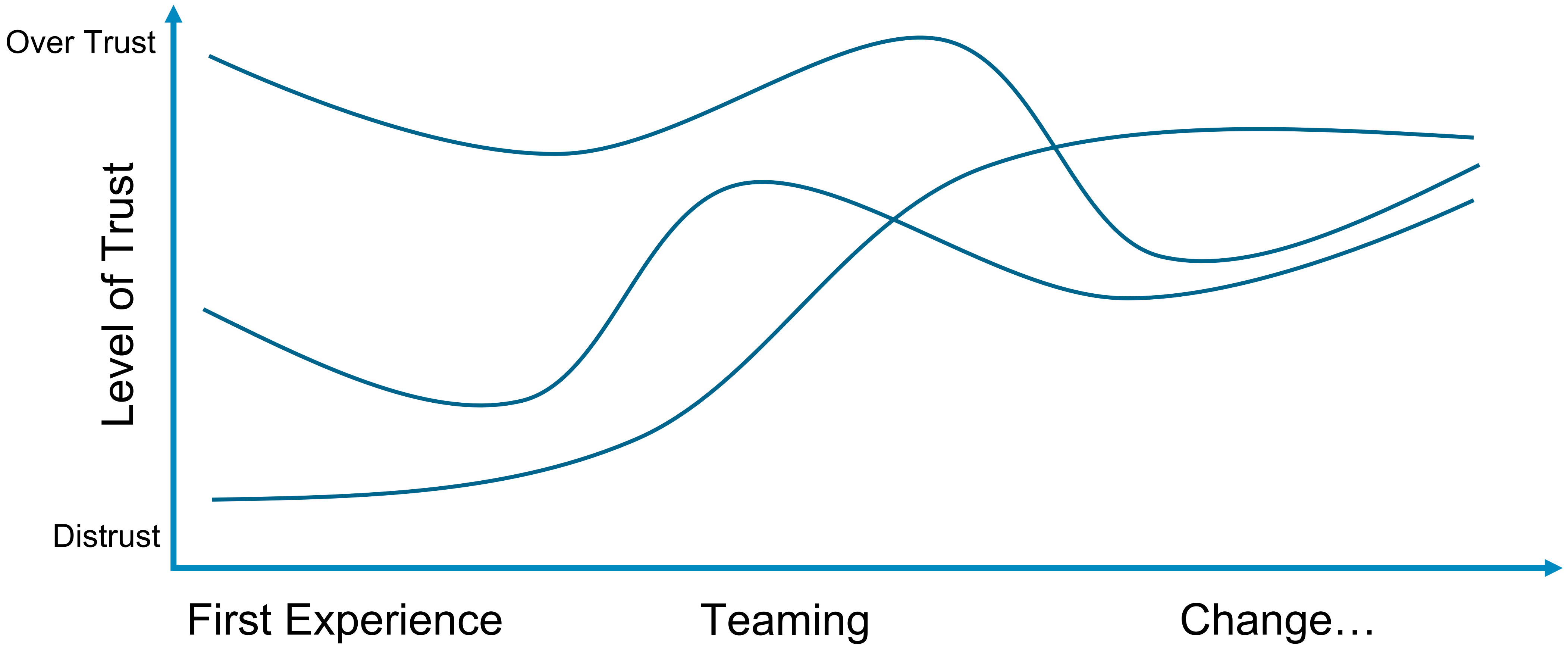
Over Trust
Trust exceeding system capabilities - may lead to misuse.

Rejection. ⟷ **Automation bias.**

John D. Lee and Katrina A. See. 2004. Trust in Automation: Designing for Appropriate Reliance. Hum Factors 46, 1 (March 2004), 50–80. DOI:https://doi.org/10.1518/hfes.46.1.50_30392
Bobbie Seppelt and John Lee. 2012. Human Factors and Ergonomics in Automation Design. In Handbook of Human Factors and Ergonomics (Fourth Edition) Chapter 59. Wiley. DOI: https://doi.org/10.1002/9781118131350.ch59

Trust changes over time

Carnegie Mellon University
Software Engineering Institute



Kun Yu, Shlomo Berkovsky, Ronnie Taib, Dan Conway, Jianlong Zhou, and Fang Chen. 2017. User Trust Dynamics: An Investigation Driven by Differences in System Performance. IUI 2017 (March 2017), 307-317. DOI: http://dx.doi.org/10.1145/3025171.3025219

Trust depends on psychological safety

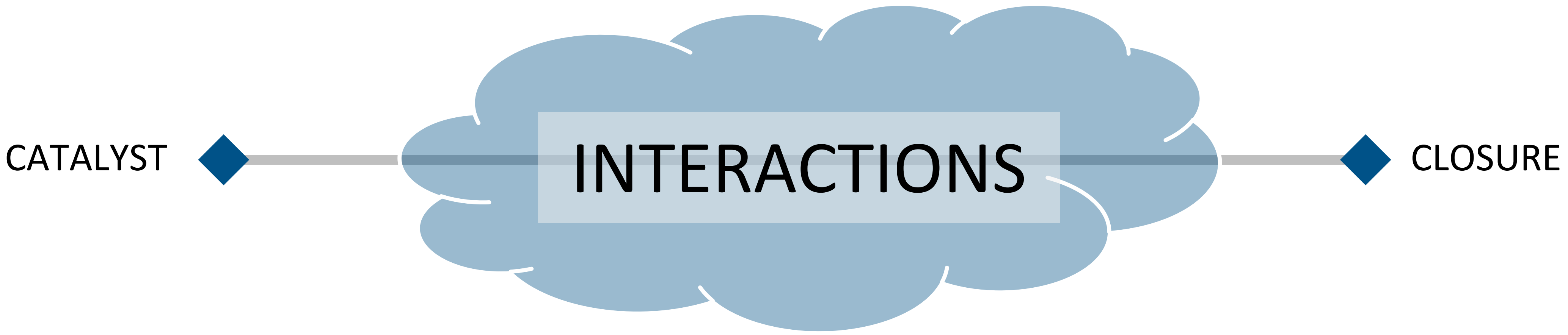
Trustworthy System Design

Decision Making

Carnegie Mellon University Software Engineering Institute

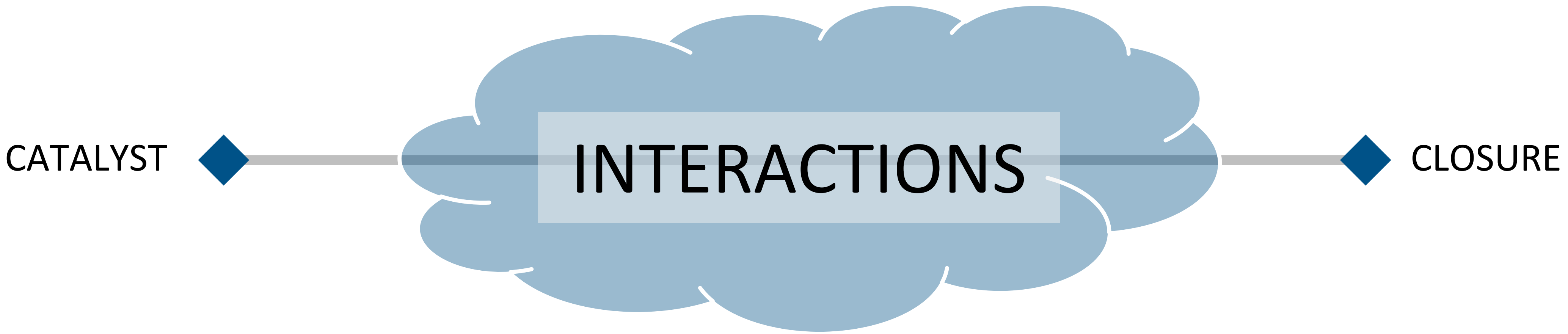
Complex Decisions

AI Systems: Exchange of information



How IAs Can Shape the Future of Human-AI Collaboration
Presented on April 28-30, 2021 at the Information Architecture Conference (IAC21)

Take turns



- Communicate
- Negotiate
- Coordinate

How IAs Can Shape the Future of Human-AI Collaboration
Presented on April 28-30, 2021 at the Information Architecture Conference (IAC21)

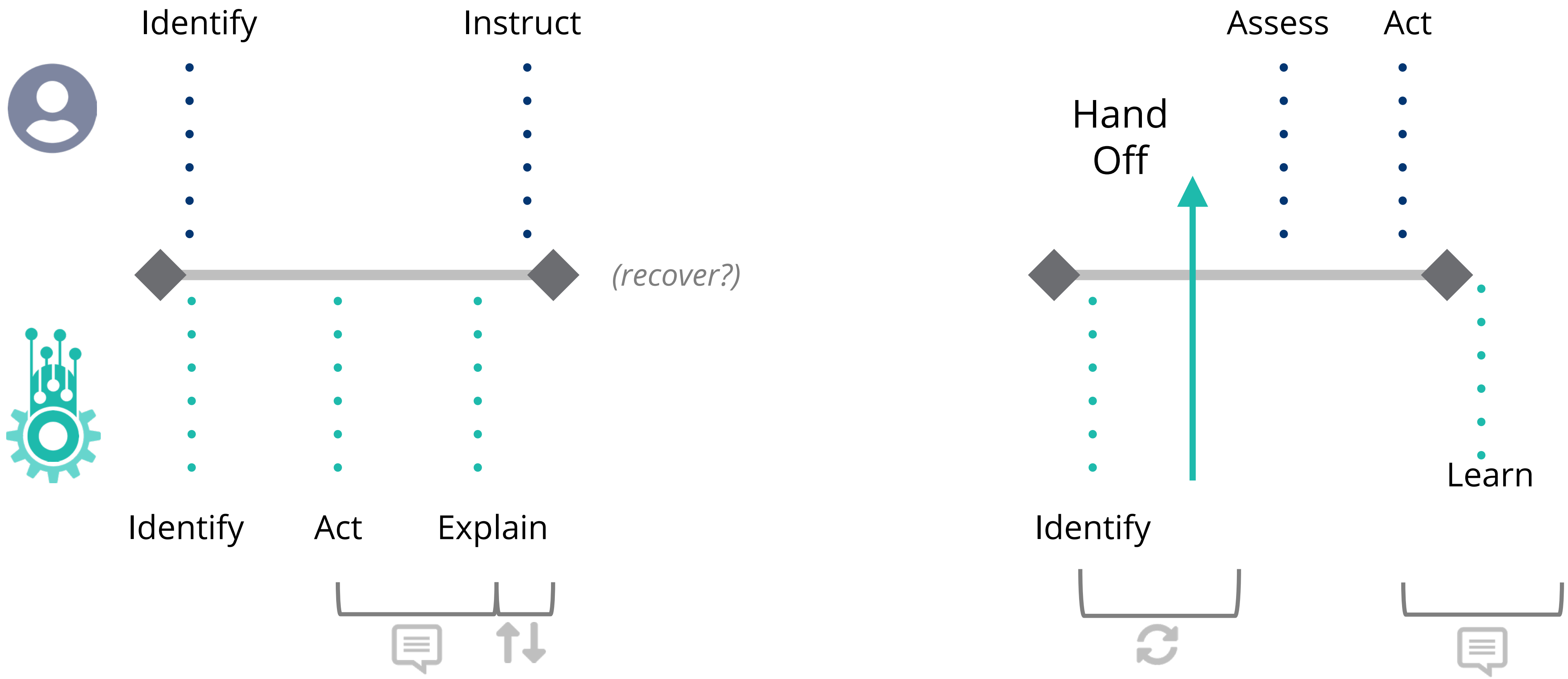
Responsibility

Significant decisions made by the system are

- explained
- able to be overridden
- appealable and reversible

Designing Trustworthy AI for Human-Machine Teaming. By Carol Smith. Software Engineering Institute Blog. March 9, 2020.

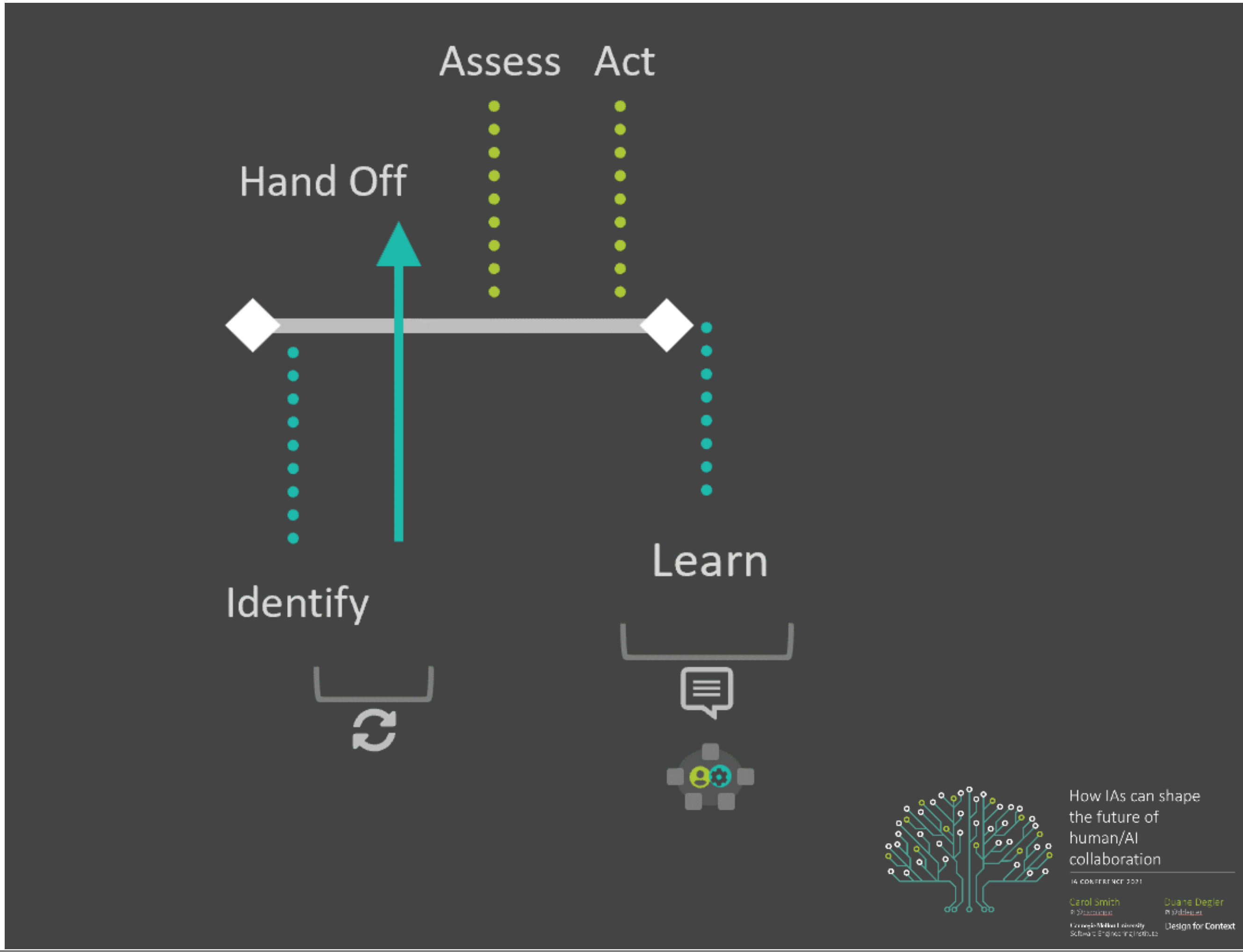
Scenario: Semi-autonomous vehicle avoiding obstruction



How IAs Can Shape the Future of Human-AI Collaboration
Presented on April 28-30, 2021 at the Information Architecture Conference (IAC21)

Dynamic situations

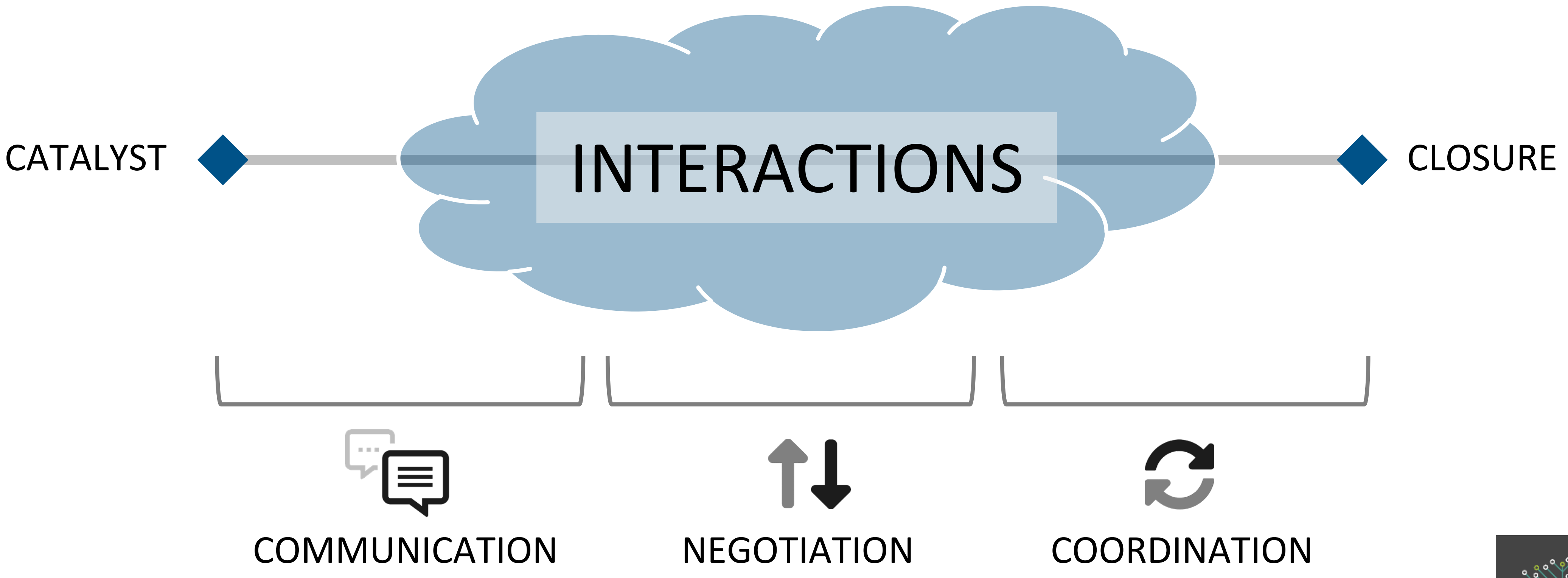
Responsibilities explicitly defined between people and systems.



How IAs Can Shape the Future of Human-AI Collaboration
Presented on April 28-30, 2021 at the Information Architecture Conference

Carnegie Mellon University Software Engineering Institute

Identify collaborative activities and interactions



How IAs Can Shape the Future of Human-AI Collaboration
Presented on April 28-30, 2021 at the Information Architecture Conference (IAC21)

Explore range of experiences

Actions to get into or maintain a **safe state** should be **easy** to do.

Actions that can lead to an **unsafe state** (hazard) should be **hard to do**.

N. G. Leveson. 2017. The Therac-25: 30 Years Later. In Computer, vol. 50, no. 11, (November 2017), 8-11. DOI: 10.1109/MC.2017.4041349
N. Leveson. 1995. Safeware: System Safety and Computers, Addison Wesley (1995).

Making Trustworthy Systems

Carnegie Mellon University
Software Engineering Institute

Identify our natural biases

Bias are shortcuts, to avoid risk and simplify problems.

Not inherently bad, may be misapplied

Implicit = invisible

Not necessarily in sync with our conscious beliefs

Can be managed and changed

Talk about biases in non-threatening, productive ways

All systems will have some form of bias

Complete objectivity is misleading.

Bias can have purpose and can be helpful.

The goal is to reduce unintended and/or harmful bias.

Conversations for Understanding

Guide teams through difficult topics

- What do we value?
- Who could be hurt?
- What lines won't our AI cross?
- How are we shifting power?*

*"Don't ask if artificial intelligence is good or fair, ask how it shifts power." Pratyusha Kalluri. https://www.nature.com/articles/d41586-020-02003-2

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

Categories of Harm

CATEGORIES OF HARM	DEFINITION
Financial	Negative impact on finances, property, or other resources
Health	Negative impact on mental, emotional, or physical health
Time	Inefficient or unproductive activities, processes, or systems
Fairness/Equity	Perpetuating or facilitating prejudice, bias and/or unfairness
Safety	Physical and/or emotional wellbeing compromised by fear, danger, or uncertainty
Privacy	Lack of control over personal information
Misinformation	Creation, spread and/or amplification of false or inaccurate information intended to deceive
Control	Inability to freely direct information, activities, or systems
Transparency	Lack of disclosure of information, activities, or systems

3Q-Do No Harm Framework, Lisa D. Dance. https://serviceease.net/3q-do-no-harm-framework

Activity 1: Trustworthy Design

SmartPackage

End-to-End home package tracking system

Integrates with most digital shopping experiences

Tracks all online orders

- Originating from address
- Shipped to address

Tracks all carriers and provides package location

- Alerts when packages arrive
- Photos and videos of delivery

Manages returns

Significant decisions

Is there an ability to:

- Turn on and off notifications?
- Control camera content/capture?

Responsibilities explicitly defined

SmartPackage or Consumer will..

- Integrate new purchases?
- Integrate new vendors?
- Determine when to send an alert?

Respectful and Secure

Who has access to shipments and contents?

Who has access to images of shipments and address?

How is that information used?

How is PII* of consumers protected?

*PII is Personally Identifiable Information (name, address, etc.)

Overview of Activity (see handout)

Introduce each other

Individually

- Consider design challenges for SmartPackage
- Use "Categories of Harm" to generate ideas

Discuss as a group

- Questions, themes, surprises, etc.
- Consider severity, frequency, risk, likelihood, etc.
- Prioritize top 3 issues/topics

1. Discussion

Carnegie Mellon University Software Engineering Institute

Identify assumptions

Consider needs for
- End-users
- Business
- Neighbors
- Society

Consider benefits and harms
- Source of data
- Photos
- Capabilities

Discuss
- Prevention and mitigation
- Tradeoffs

2. Sketch!

Select one of the top issues/topics

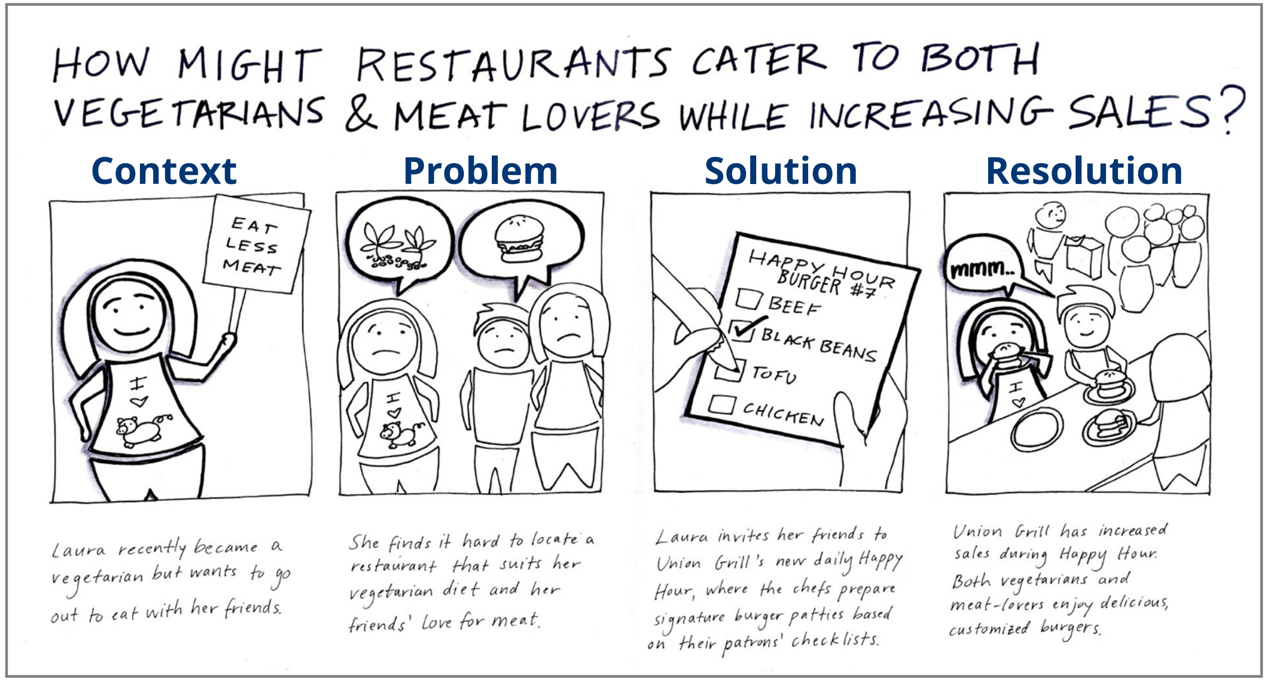
Individually

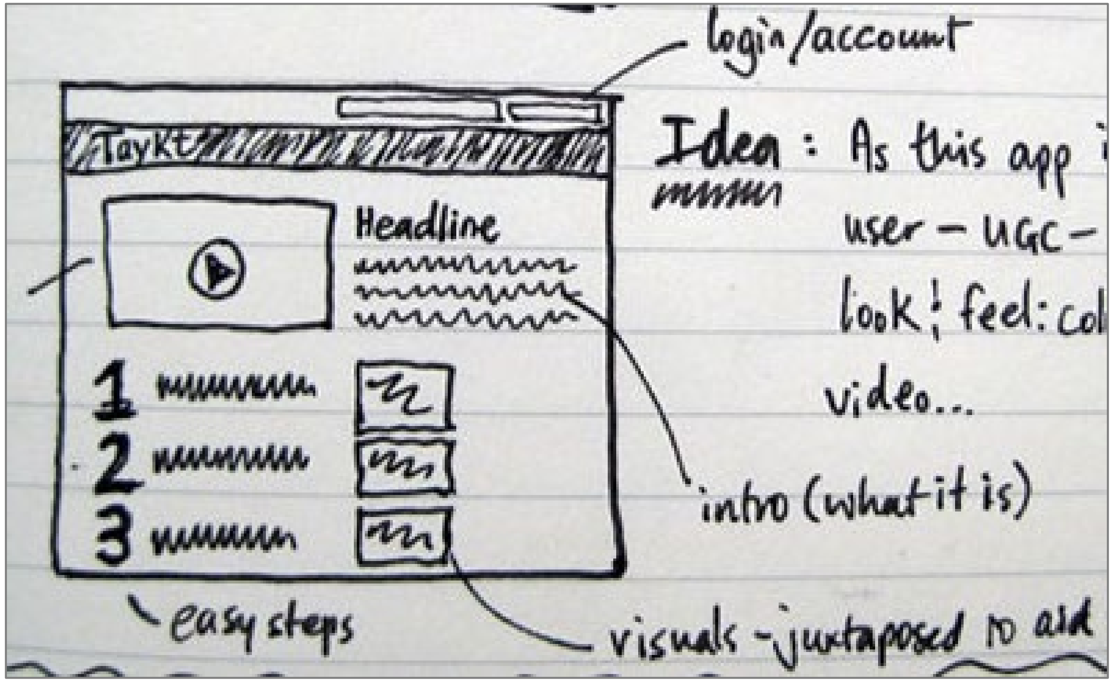
- Sketch (digital or paper) interactions
- Preventing harm?
- Providing evidence of capabilities?
- Providing evidence of source of information?

Group

- Present sketches to each other and discuss
- Selects 1-2 sketches
- One person to share with room

Fidelity can vary

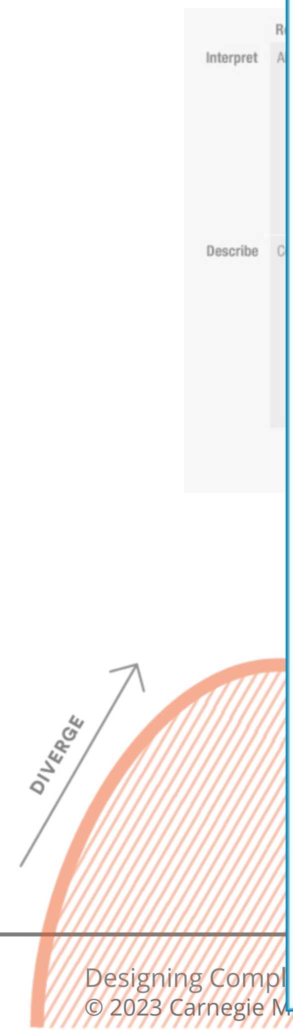


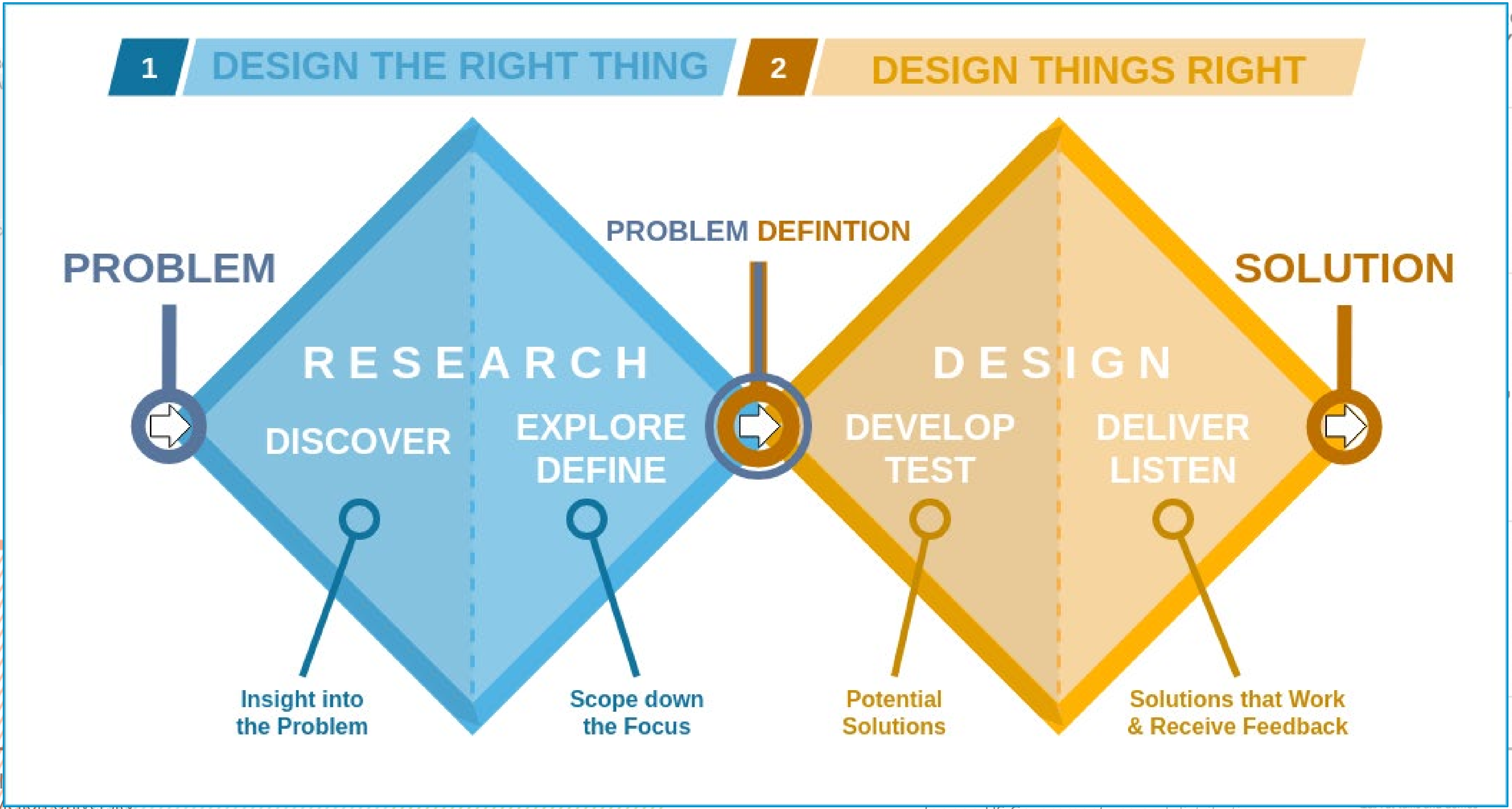


Discussion - Questions

DIFFERENT NEEDS, GOALS, PURPOSE…

Research & Speculative Activities





Activate Curiosity

Speculate about system impacts and misuse.

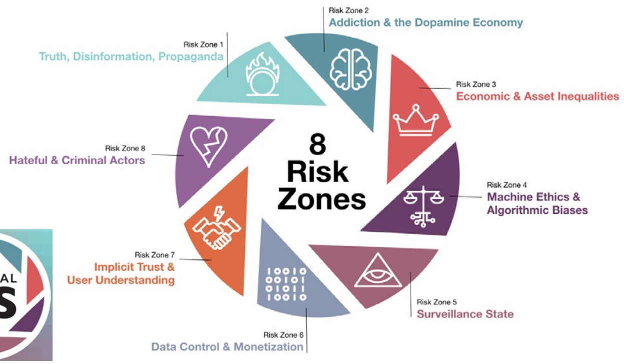
Identify potential
- unintended and/or unwanted consequences?
- severe abuse and consequences?
- negative consequences - historically and commonly marginalized groups?

Designing Trustworthy AI for Human-Machine Teaming. By Carol Smith. Software Engineering Institute Blog. March 9, 2020.

Reward team members for finding ethics bugs

Adopt Technology Ethics

- Harmonize cultural variations
- Balance to pace of change, industry pressure
- Explicit permission to consider breadth of implications



Montréal Declaration Responsible AI_

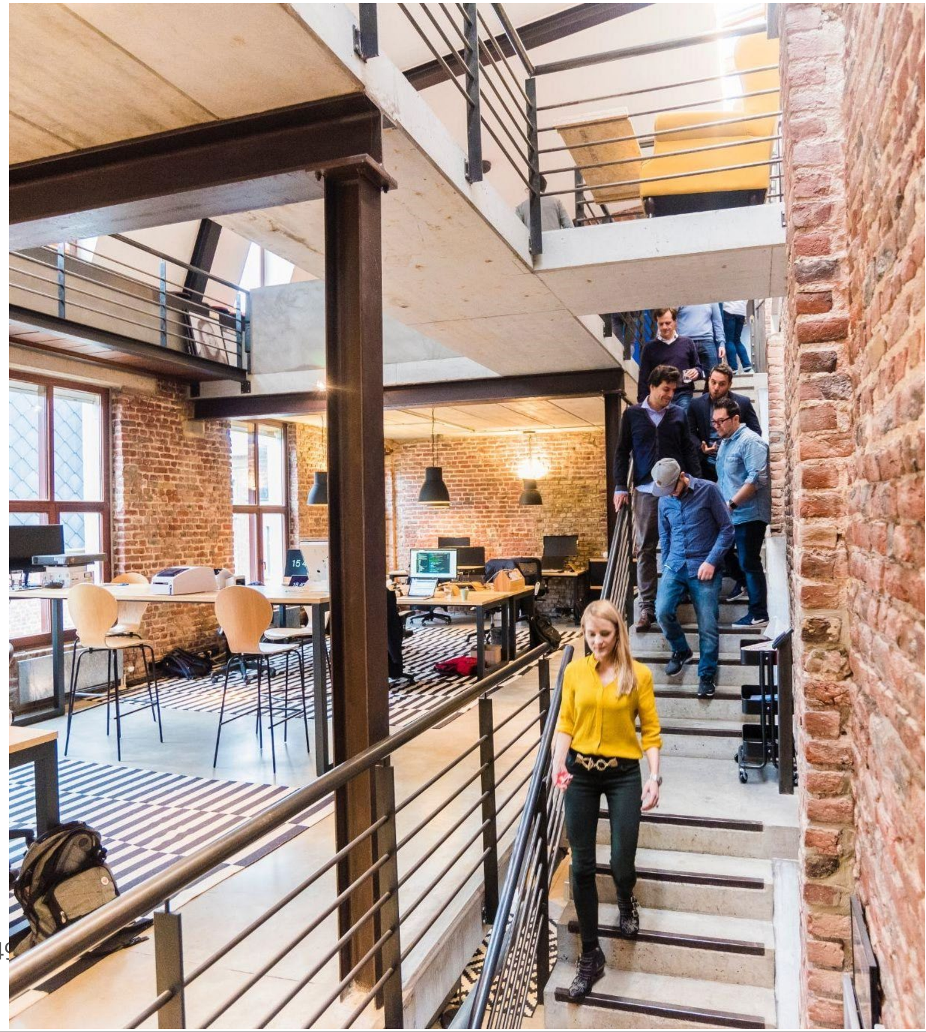
An initiative of Université de Montréal

Integrate Ethical Principles Across Organization

Carnegie Mellon University Software Engineering Institute

Recognize need for organizational change management

- Leader and stakeholder alignment
- Education
- Robust communication
- Assigned responsibility



Where Responsible AI meets Reality: Practitioner Perspectives on Enablers for Shifting Organizational Practices. https://doi.org/10.1145/3449
Photo by Proxyclick Visitor Management System: https://www.pexels.com/photo/photo-of-woman-in-yellow-top-walking-down-stairs-in-a-office-room-2451567/

Resolve misalignment

- Proactive identification of potential harms
- Ensure projects align with organization’s mission and values
- Include social impact in performance/success metrics
- Correct misaligned incentives
- Mitigate focus on fast release of technology

Bogdana Rakova, Jingying Yang, Henriette Cramer, and Rumman Chowdhury. 2021. Where Responsible AI meets Reality: Practitioner Perspectives on Enablers for Shifting Organizational Practices. Proc. ACM Hum.- Comput. Interact. 5, CSCW1, Article 7 (April 2021), 23 pages. https://doi.org/10.1145/344908

Broaden our work

- Is the technology right for the challenge?
- What kind of improvements are expected?
- What are the benefits and risks?
- How will we know we've made improvements?

Great Potential - Develop With Caution

Technology cannot fix human or social issues

Math does not reduce bias – assuming so increases risks:
- Perpetuating bias
- Embedding existing bad behaviors
- Dehumanizing

AI systems are dynamic - require constant monitoring.

Break

Return at 11:00

Carnegie Mellon University Software Engineering Institute

Welcome!

10 min - Welcome

20 min – Topic Overview

40 min – Activity: Trustworthy Design

20 min – Research & Speculation

30 min – Break

Timing is approximate

20 min - Frameworks

60 min – Activity: Abusability Testing

10 min - Take action and Close

Frameworks

Use Frameworks to prompt conversations

Pair with technical ethics

- Bridge gaps between “do no harm” and reality

Reduce risk and unwanted bias

Support inspection and mitigation planning

3Q-Do No Harm Framework

WHO'S NOT HERE?	HOW WILL VULNERABLE GROUPS BE NEGATIVELY IMPACTED?	WHEN THINGS DON'T WORK, HOW WILL THEY BE QUICKLY RESOLVED?
Create Inclusive Teams and Diversify Research Participants	Identify Unintended Consequences and Mitigate beforehand	Ensure the path to resolving problems is clear and fast

3Q-Do No Harm Framework, Lisa D. Dance. https://serviceease.net/3q-do-no-harm-framework

Who’s not here?

Consider team and research participants

3Q-Do No Harm Framework, Lisa D. Dance. https://serviceease.net/3q-do-no-harm-framework

How will vulnerable groups be negatively impacted?

Identify unintended consequences and mitigate beforehand

3Q-Do No Harm Framework, Lisa D. Dance.

When things don't work, how will it be quickly resolved?

Ensure the path to resolving problems is clear and fast

3Q-Do No Harm Framework, Lisa D. Dance. https://serviceease.net/3q-do-no-harm-framework

Categories of Harm

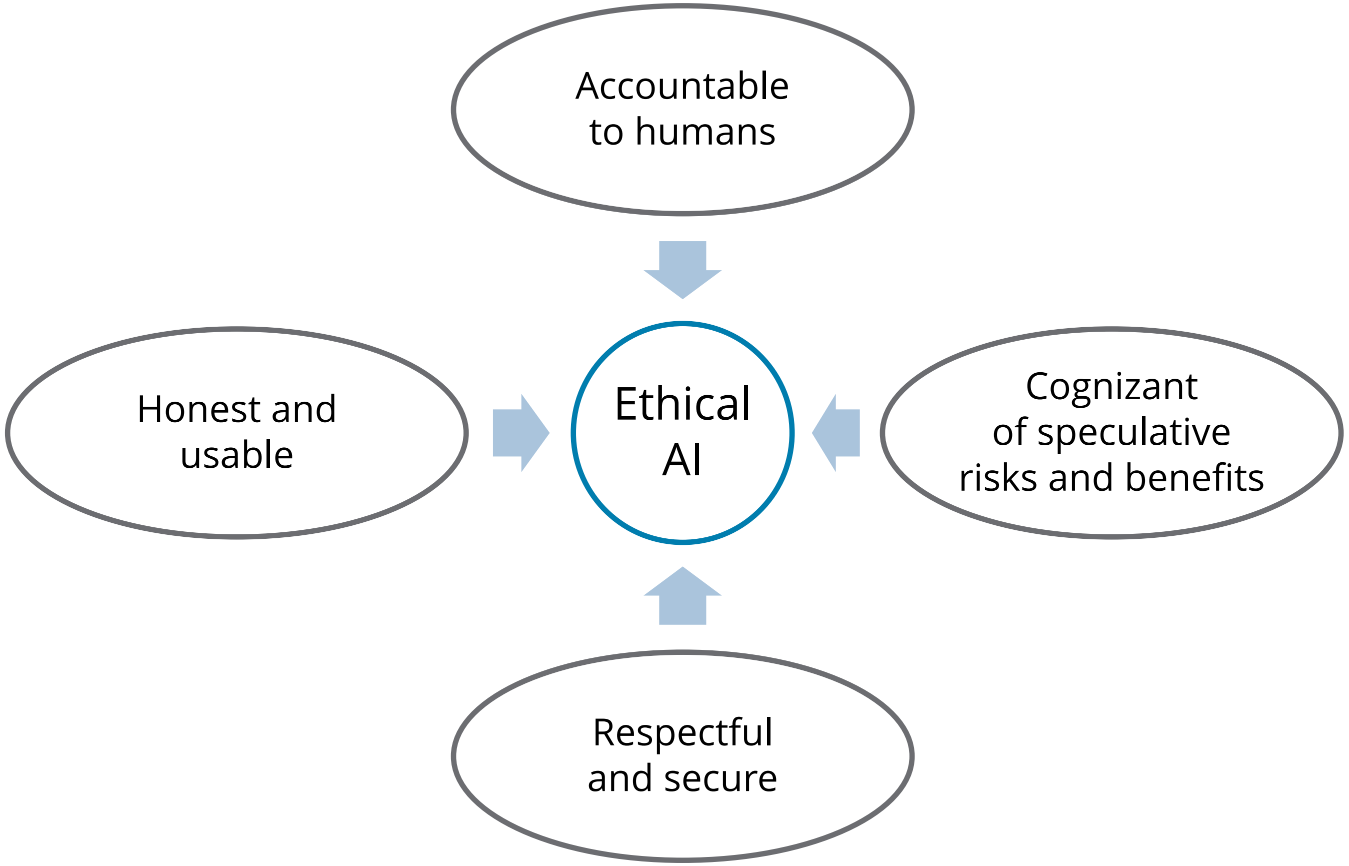
Use these categories of harm to evaluate how a product, service, or technology could cause harm.

CATEGORIES OF HARM	DEFINITION
Financial	Negative impact on finances, property, or other resources
Health	Negative impact on mental, emotional, or physical health
Time	Inefficient or unproductive activities, processes, or systems
Fairness/Equity	Perpetuating or facilitating prejudice, bias and/or unfairness
Safety	Physical and/or emotional wellbeing compromised by fear, danger, or uncertainty
Privacy	Lack of control over personal information
Misinformation	Creation, spread and/or amplification of false or inaccurate information intended to deceive
Control	Inability to freely direct information, activities, or systems
Transparency	Lack of disclosure of information, activities, or systems

UX Framework for Designing Trustworthy AI

Carnegie Mellon University
Software Engineering Institute

UX Framework for Designing Trustworthy AI



Carol J. Smith. "Designing Trustworthy AI: A Human-Machine Teaming Framework to Guide Development“ arXiv:1910.03515 [cs.AI]. Presented at AAAI (Association for the Advancement of Artificial Intelligence) Symposium FSS-19: Artificial Intelligence in Government and Public Sector, Nov 7, 2019. Arlington, Virginia, USA. https://arxiv.org/abs/1910.0351
Designing Trustworthy AI for Human-Machine Teaming. By Carol Smith. Software Engineering Institute Blog. March 9, 2020. https://insights.sei.cmu.edu/sei_blog/2020/03/designing-trustworthy-ai-for-human-machine-teaming.html

Accountable to Humans

Ensure humans have ultimate control.

Able to monitor and control risk.

Human responsibility for final decisions.

Significant decisions made by system will be:

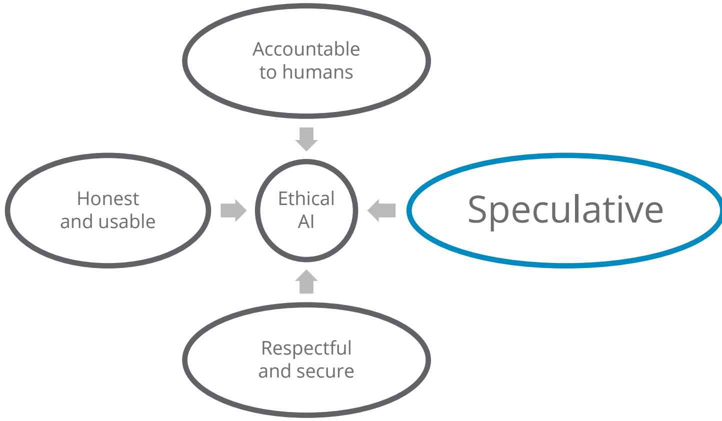
- explained
- able to be overridden
- appealable and reversible

Cognizant of Speculative Risks and Benefits

Identify full range of
- Harmful, malicious use, as well as good, beneficial use
- Unwanted/unintended consequences

Conduct UX research

Create communication & mitigation plans



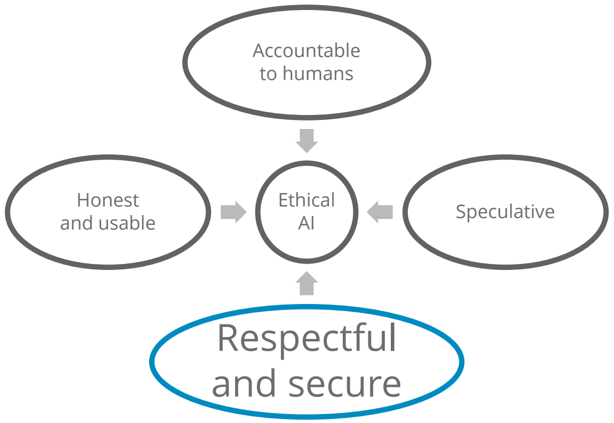
Respectful and Secure

Values of humanity, ethics, equity, fairness, accessibility, diversity and inclusion

Respect privacy and data rights

Make system robust, valid and reliable

Provide understandable security



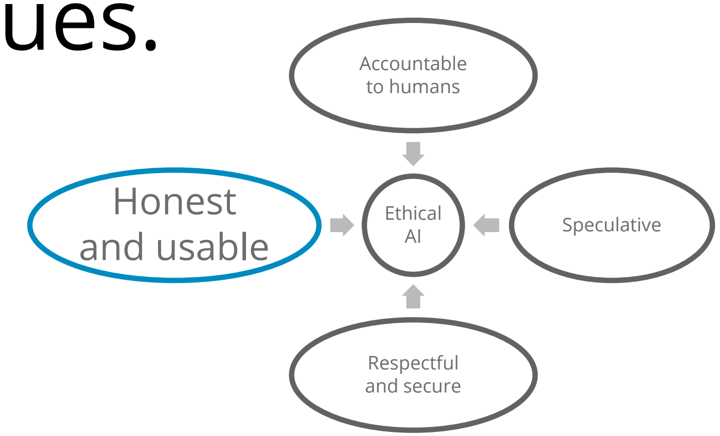
Honest and Usable

Value transparency with the goal of engendering calibrated trust.

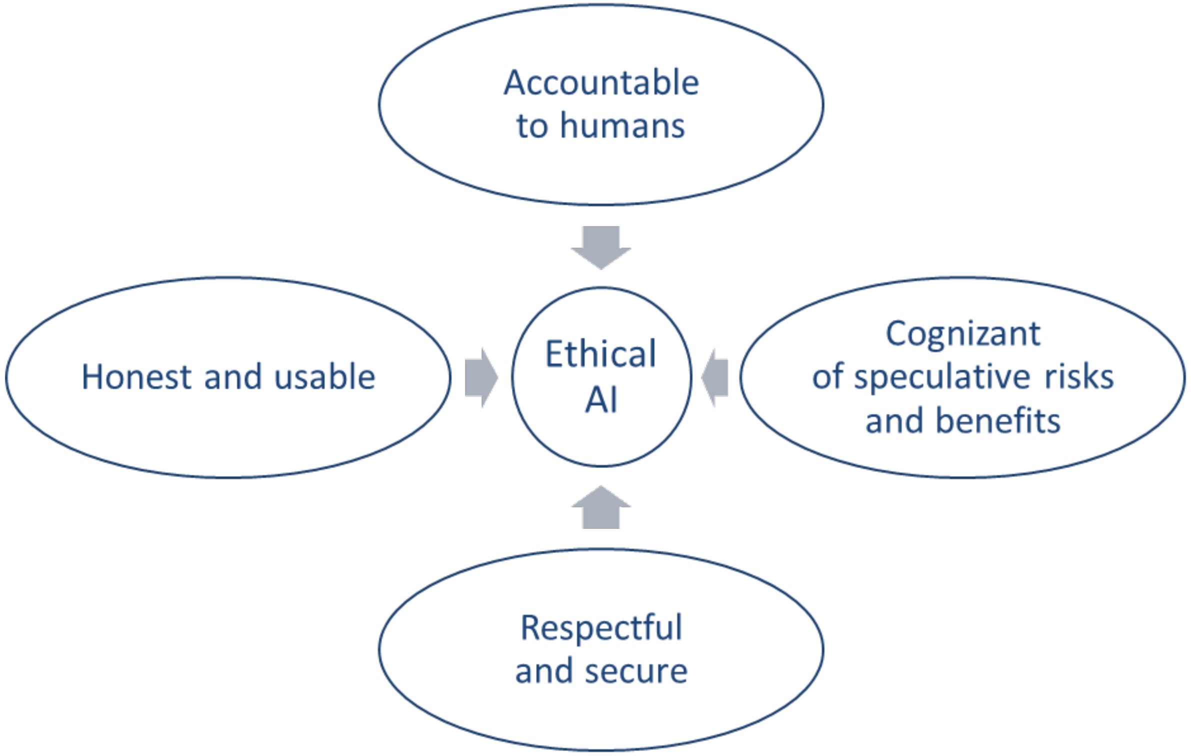
Explicitly state identity as an AI system.

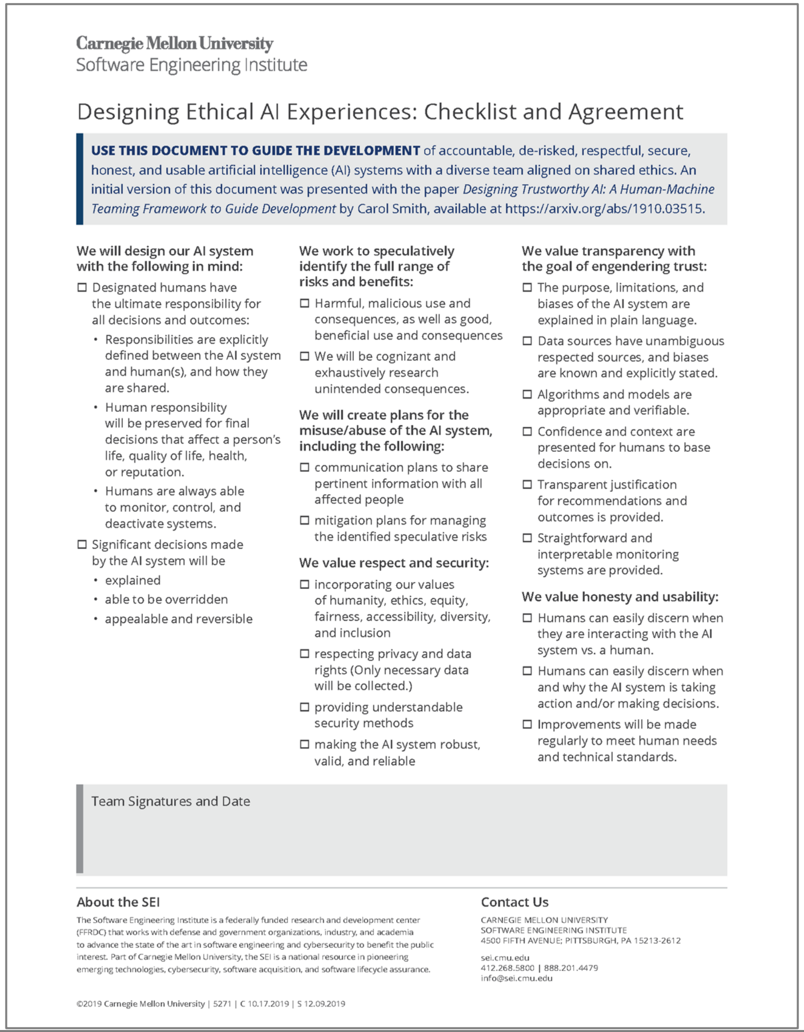
Show awareness of known and desirable bias.

Acknowledge and overcommunicate on issues.



Designing Ethical AI Experiences: Checklist and Agreement





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

The Software Engineering Institute is a federally funded research and development center (FFRDC) that works with defense and government organizations, industry, and academia to advance the state of the art in software engineering and cybersecurity to benefit the public interest. Part of Carnegie Mellon University, the SEI is a national resource in pioneering emerging technologies, cybersecurity, software acquisition, and software lifecycle assurance.

Contact Us

CARNEGIE MELLON UNIVERSITY
SOFTWARE ENGINEERING INSTITUTE
4500 FIFTH AVENUE; PITTSBURGH, PA 15213-2612

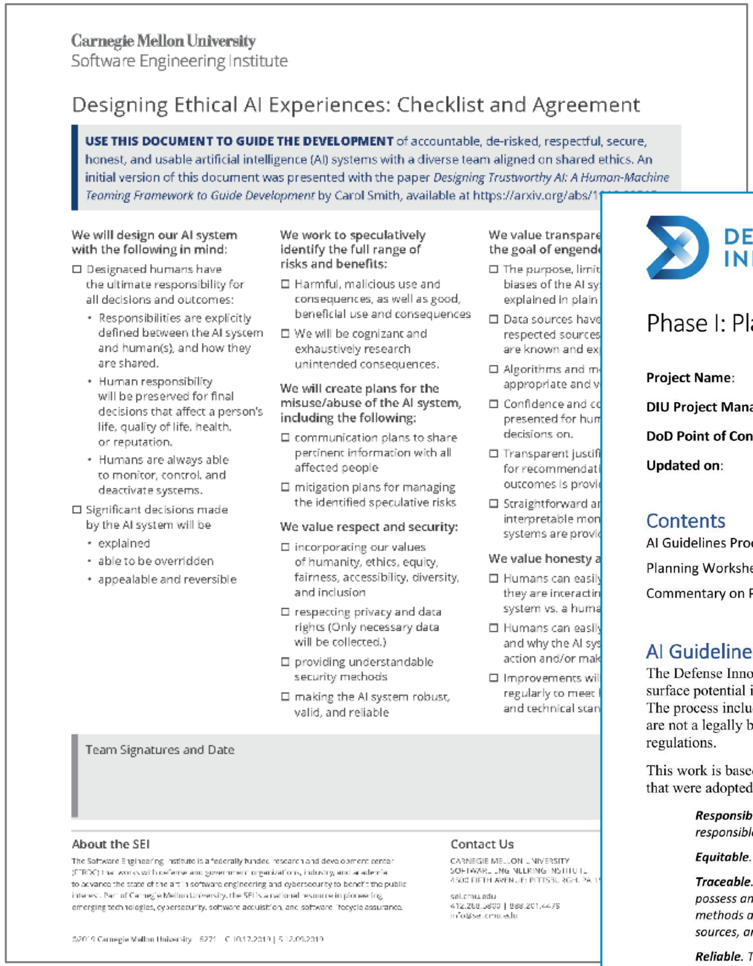
sei.cmu.edu
412.268.5800 | 888.201.4479
info@sei.cmu.edu

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

Checklists and Guides

Carnegie Mellon University Software Engineering Institute

- Pair with technical ethics
- Bridge gaps between "do no harm" and reality
- Drives conversations
- 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-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/1

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 transpare... the goal of engende...

- ☐ The purpose, limit... biases of the AI sy... explained in plain...
- ☐ Data sources have... respected sources... are known and ex...
- ☐ Algorithms and m... appropriate and v...
- ☐ Confidence and c... presented for hum... decisions on.
- ☐ Transparent justifi... for recommendati... outcomes is provi...
- ☐ Straightforward ar... interpretable mon... systems are provi...

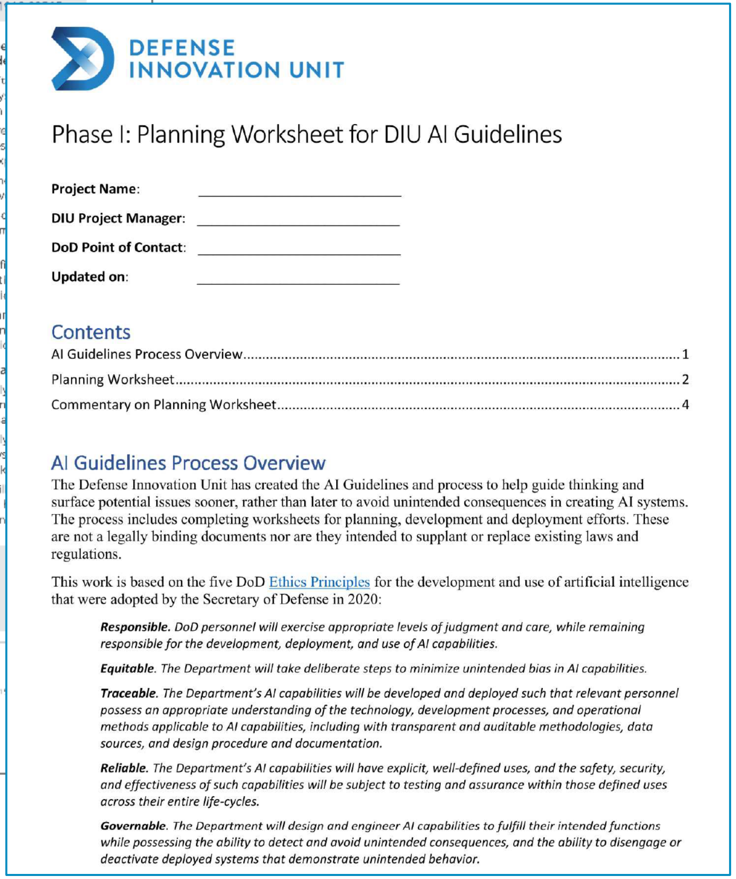
We value honesty a...

- ☐ Humans can easily... they are interactin... system vs. a huma...
- ☐ Humans can easily... and why the AI sys... action and/or mak...
- ☐ Improvements wil... regularly to meet... and technical stan...

Team Signatures and Date

About the SEI

Contact Us


DEFENSE INNOVATION UNIT

Phase I: Planning Worksheet for DIU AI Guidelines

Project Name: ____________________

DIU Project Manager: ____________________

DoD Point of Contact: ____________________

Updated on: ____________________

Contents

AI Guidelines Process Overview

The Defense Innovation Unit has created the AI Guidelines and process to help guide thinking and surface potential issues sooner, rather than later to avoid unintended consequences in creating AI systems. The process includes completing worksheets for planning, development and deployment efforts. These are not a legally binding documents nor are they intended to supplant or replace existing laws and regulations.

This work is based on the five DoD Ethics Principles for the development and use of artificial intelligence that were adopted by the Secretary of Defense in 2020:

> ***Responsible.*** *DoD personnel will exercise appropriate levels of judgment and care, while remaining responsible for the development, deployment, and use of AI capabilities.*
>
> ***Equitable.*** *The Department will 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.*

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

Activity 2: Abusability Testing

Activate curiosity

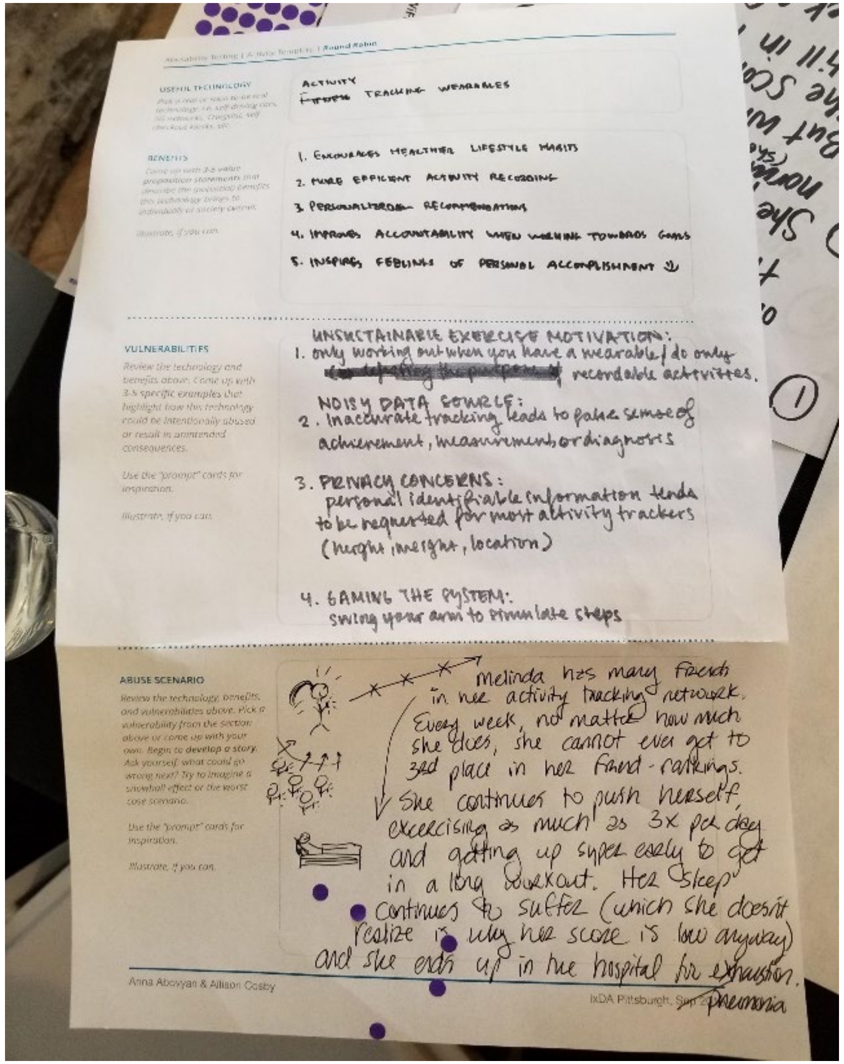
- Speculate about misuse and abuse
- Potential severe abuse and unintended/unwanted consequences
- Perspective of people in frequently marginalized groups

Abusability Testing

Carnegie Mellon University
Software Engineering Institute

1. **Value proposition**
   (Potential) Benefits this tech brings to individuals or society overall
2. **Vulnerabilities**
   Specific examples of how tech could be
   misused or intentionally abused
3. **Abuse scenario**

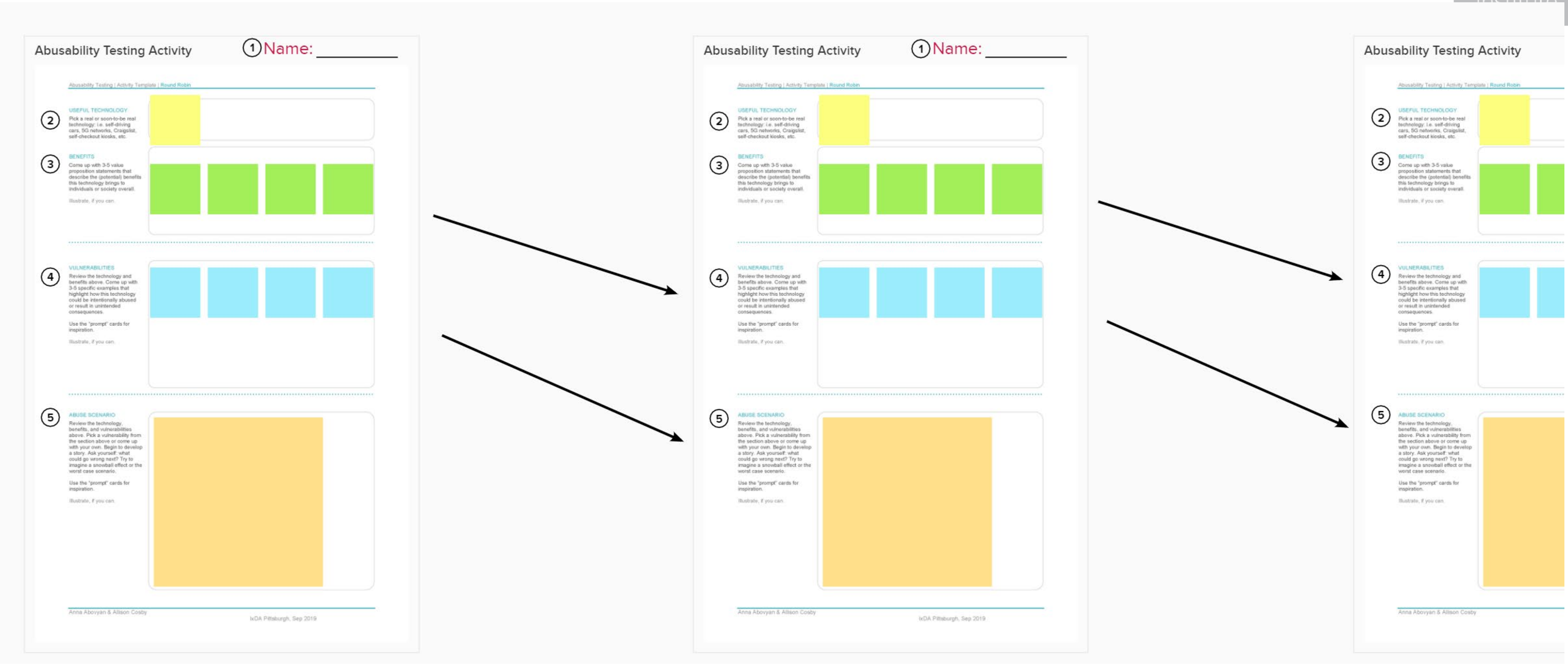
Provocation via prompt statements



UX in the Age of Abusability. The role of Composition, Collaboration, and Craft in building ethical products. Dan Brown. Sep 18, 2018. https://greenonions.com/ux-in-the-age-of-abusability-797cd01f6b13
Photo from workshop organized by Anna Abovyan, Theora Kvitka and Allison Cosby. Pittsburgh IxDA Chapter for World Interaction Design Day 2019.

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

Round Robin Activity



Pick a Technology

Individually – Describe a technology you are working on now or pick a newish technology.

Benefits

- Write value proposition statements.
- Describe the (potential) benefits technology brings to individuals or society overall (3-5).

Swap pages


Abusability Testing Activity
1 Name:
Abusability Testing Activity
1 Name:
Abusability Testing Activity

Vulnerabilities and Abuse Scenario

Specific vulnerabilities

- Examples of how this technology could be intentionally abused or result in unintended consequences (3-5).

Swap and review the technology, benefits, and vulnerabilities.

- Develop a story from a vulnerability or come up with your own.
- What could go wrong next?
- Imagine a snowball effect - the worst-case scenario.

Prompt Statements

- What happens if the electricity goes out?
- What happens if devices can charge themselves?
- What if there’s no WiFi or cellular signal?
- You’ve been hacked.
- How is status communicated?
- Is there a back door?
- What error conditions need to be considered?

More Prompt Statements

- 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.)?
- Who else might gain access?
- Would anonymity help? Or is personal accountability necessary?
- How could someone with criminal intent use this technology?
- How could governments use this technology?

“Black Mirror” Style Scenario

- Group selects one abuse scenario.
- Turns it into a “Black Mirror” style scenario (inspired by British dystopian sci-fi tv series of same name).
- Illustrate if possible.

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

Taking Action

Tools to take you forward

Carnegie Mellon University Software Engineering Institute

Encourage conversations for understanding

- 3Q-Do No Harm Framework
- UX Framework for Designing Trustworthy AI and Checklist

Activate curiosity

- Abusability Testing
- “Black Mirror” Episodes

Organizational questions

- What value do we see in defining a set of ethical guidance for our organization?
- Should we build on existing codes, guidance and best practices?
- Would different departments/products need different guidance?
- How might we reflect our culture/what we value in a code of ethics?

Organizational questions

- Who do we need to protect?
- How will we describe the “lines” our work won’t cross?
- What mitigation strategies do we need?
- What about social consequences such as job displacement?
- How might we track progress?

What scenarios do you have?

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

Responsible, Intentional Design

Just because you can, doesn't mean you should.

Carnegie Mellon University Software Engineering Institute

Carol J. Smith
LinkedIn: /in/caroljsmith/

Discussion & Questions

Carnegie Mellon University Software Engineering Institute