Create human-centered AI with the Human-AI eXperiences (HAX) Toolkit

Saleema Amershi, Ph.D.
Mihaela Vorvoreanu, Ph.D.
Dr. Saleema Amershi
Sr Principal Research Manager
HAX Group, Microsoft Research
Co-chair Aether Working Group on Human-AI Interaction & Collaboration
Agenda

01 Human-centered AI – what is it and why do we need it?

02 The HAX Toolkit for creating human-centered AI
Microsoft’s Project Tokyo

Facial recognition used to identify Lafayette Square protester
The Washington Post, November 2, 2020
Ira Black/Getty Images
Society is demanding change
Challenges to creating AI responsibly

Cultural—shifting mindsets and embracing diversity
Organizational—aligning business and RAI objectives
Technical—building AI technologies
We’re a long way from purely technical solutions
How do we support our AI builders?
Responsible AI is Human-Centered

It’s about ensuring that *what we build* benefits people and society, and that *how we build it* begins and ends with people in mind.
How do we build AI in a human-centered way?

Tie all technical decisions back to user needs
People & context

AI system

Interface
Models
Data
Infrastructure
How do we build AI in a human-centered way?

Tie all technical decisions back to user needs

Involve diverse perspectives early and throughout

Plan for failures so users can recover when things go wrong

...
HAX Toolkit

Tools for creating responsible human-AI experiences.
Best practices for how AI systems should behave during human-AI interaction

A tool to guide teams through planning and implementing human-AI interaction best practices

A set of flexible and reusable solutions to recurring human-AI interaction problems

A tool for generating scenarios to test based on likely human-AI interaction failures

More tools coming soon!

https://aka.ms/haxtoolkit
Guidelines for Human-AI Interaction

ACM CHI 2019, Best Paper Honorable Mention Award

Guidelines for Human-AI Interaction
Saleema Amershi, Dan Weld, Mihaela Vorvoreanu, Adam Fourney, Besmira Nushi, Penny Collisson, Jina Suh, Shamsi Iqbal, Paul N. Bennett, Kori Inkpen, Jaime Teevan, Ruth Kikin-Gil, and Eric Horvitz

Microsoft
Redmond, WA, USA
{amershi, mivorver, advom, benshi, pennbyy, jinishe, shamsip, pauben, kori, teevan, rufikag, horvitz}@microsoft.com

ABSTRACT
Advances in artificial intelligence (AI) frame opportunities and challenges for user interface design. Principles for human-AI interaction have been discussed in the human-computer interaction community for over two decades, but more study and innovation are needed in light of advances in AI and the growing uses of AI technologies in human-facing applications. We propose 18 generally applicable design guidelines for human-AI interaction. These guidelines are validated through multiple rounds of evaluation including a user study with 49 design practitioners who tested the guidelines against 20 popular AI-infused products. The results verify the relevance of the guidelines over a spectrum of interaction scenarios and reveal gaps in our knowledge, highlighting opportunities for further research. Based on the evaluations, we believe the set of design guidelines can serve as a resource to practitioners working on the design of applications and features that harness AI technologies, and to researchers interested in the further development of guidelines for human-AI interaction design.

CCS CONCEPTS
- Human-centered computing → Human computer interaction (HCI); - Computing methodologies → Artificial intelligence.

1 INTRODUCTION
Advances in artificial intelligence (AI) are enabling developers to integrate a variety of AI capabilities into user-facing systems. For example, increases in the accuracy of pattern recognition have created opportunities and pressure to integrate speech recognition, translation, object recognition, and face recognition into applications. However, as automated inferences are typically performed under uncertainty, often producing false positives and false negatives, AI-infused systems may demonstrate unpredictable behaviors that can be disruptive, confusing, offensive, and even dangerous. While some AI technologies are deployed in explicit, interactive uses, other advances are employed behind the scenes in proactive services acting on behalf of users such as automatically filtering content based on inferred relevance or importance. While such attempts at personalization may be
AI is fundamentally changing how we interact with computing systems...
...and creating new challenges for people

Tracy Brisson @tracybriss...  ·  Nov 18, 2019
Me: “Alexa, put bay leaves on our shopping list.”
Alexa: “I’ve put babies on your shopping list.”

#alexafail

Culver City Firefighters @CC_Firefighters
While working a freeway accident this morning, Engine 42 was struck by a #Tesla traveling at 65 mph. The driver reports the vehicle was on autopilot. Amazingly there were no injuries! Please stay alert while driving! #abc7eyewitness #ktla #CulverCity #distracteddriving
Creating the Guidelines for Human-AI Interaction

- **Phase 1. Consolidation**: Identified themes across 150+ recommendations
- **Phase 2. Team Evaluation**: Modified heuristic evaluation over 13 common AI products
- **Phase 3. User Evaluation**: Systematic analysis of 20 AI products with 49 UX practitioners
- **Phase 4. Expert Review**: Final review with 11 UX practitioners
Disclaimers

The guidelines are not a checklist

Additional guidelines may be needed in some scenarios

You are using them “the right way” if you consider them during development
Guidelines for Human AI Interaction

https://aka.ms/haxtoolkit

1. Make clear what the system can do
2. Make clear how well the system can do what it can do
3. Time services based on context
4. Show contextually relevant information
5. Match relevant social norms
6. Mitigate social biases
7. Support efficient invocation
8. Support efficient dismissal
9. Support efficient correction
10. Scope services when in doubt
11. Make clear why the system did what it did
12. Remember recent interactions
13. Learn from user behavior
14. Update and adapt cautiously
15. Encourage granular feedback
16. Convey the consequences of user actions
17. Provide global controls
18. Notify users about changes
Guidelines for Human AI Interaction

https://aka.ms/haxtoolkit

1. Make clear what the system can do
2. Make clear how well the system can do what it can do
3. Time services based on context
4. Show contextually relevant information
5. Match relevant social norms
6. Mitigate social biases
7. Support efficient invocation
8. Support efficient dismissal
9. Support efficient correction
10. Scope services when in doubt
11. Make clear why the system did what it did
12. Remember recent interactions
13. Learn from user behavior
14. Update and adapt cautiously
15. Encourage granular feedback
16. Convey the consequences of user actions
17. Provide global controls
18. Notify users about changes
## Guidelines for Human AI Interaction

**INITIALLY**

1. Make clear what the system can do
2. Make clear how well the system can do what it can do

**DURING INTERACTION**

3. Time services based on context
4. Show contextually relevant information
5. Match relevant social norms
6. Mitigate social biases
7. Support efficient invocation
8. Support efficient dismissal
9. Support efficient correction
10. Scope services when in doubt
11. Make clear why the system did what it did

**WHEN WRONG**

12. Remember recent interactions
13. Learn from user behavior
14. Update and adapt cautiously
15. Encourage granular feedback
16. Convey the consequences of user actions

**OVER TIME**

17. Provide global controls
18. Notify users about changes

[https://aka.ms/haxtoolkit](https://aka.ms/haxtoolkit)
Guidelines for Human AI Interaction

1. Make clear what the system can do
2. Make clear how well the system can do what it can do
3. Time services based on context
4. Show contextually relevant information
5. Match relevant social norms
6. Mitigate social biases
7. Support efficient invocation
8. Support efficient dismissal
9. Support efficient correction
10. Scope services when in doubt
11. Make clear why the system did what it did
12. Remember recent interactions
13. Learn from user behavior
14. Update and adapt cautiously
15. Encourage granular feedback
16. Convey the consequences of user actions
17. Provide global controls
18. Notify users about changes

https://aka.ms/haxtoolkit
Guidelines for Human AI Interaction

1. Make clear what the system can do
2. Make clear how well the system can do what it can do
3. Time services based on context
4. Show contextually relevant information
5. Match relevant social norms
6. Mitigate social biases
7. Support efficient invocation
8. Support efficient dismissal
9. Support efficient correction
10. Scope services when in doubt
11. Make clear why the system did what it did
12. Remember recent interactions
13. Learn from user behavior
14. Update and adapt cautiously
15. Encourage granular feedback
16. Convey the consequences of user actions
17. Provide global controls
18. Notify users about changes

INITIALLY

DURING INTERACTION

WHEN WRONG

OVER TIME

https://aka.ms/haxtoolkit
Teams who have used the guidelines say

“[This is] all about creating trustworthy assistive systems...what really matters here is the model for human interaction with a system that has been given tremendous authority over human life.”

– Designer

“It took us four years to come up with 80% of [the guidelines]...if you can institutionalize this into the design parts of products, you give people an opportunity to build a much better products in gen one.”

– Program Manager

<p>| | | | | | | | | | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
</table>
Challenges in operationalizing the guidelines

“These guidelines are great...how can you make this as part of people’s, you know, work?”

“A lot of the guidelines are based upon the engineering more than the design...[some], would require a full-scale overhaul of backend.”

“If the spec doesn’t have that built into it, it’s gonna be too rigid to respond...if you miss the boat in the beginning you just have to sort of yell from the shore and ask them to like change where they’re steering kind of thing.”
Dr. Mihaela Vorvoreanu
Aether Director of UX Research & Responsible AI Education
Best practices for how AI systems should behave during user interaction

A tool to guide teams through planning and implementing human-AI interaction best practices

A set of flexible and reusable solutions to recurring human-AI interaction problems

A tool for generating scenarios to test based on likely human-AI interaction errors

More tools coming soon!

https://aka.ms/haxtoolkit
The Team

Funded by Aether and Microsoft Research

Meg Young
UW (now at Cornell Tech)

Andrew Anderson
OSU

Matthew Hong
Georgia Tech (now at UW)

Quan Ze (Jim) Chen
UW

Saleema Amershi
Mihaela Vovoreanu
Adam Fourney
Besmira Nushi
Tobias Schnabel
Jenn Wortman Vaughan
Hanna Wallach
Derek DeBellis
Jason Geiger

https://aka.ms/haxtoolkit
The HAX Workbook

For planning and implementing human-AI interaction best practices
### The HAX Workbook

#### Guideline 1: Make clear what the system can do.
- **Description:** Set the right expectations about the AI system's capabilities.
- **Examples:**
  - An AI-powered resume assistant.

#### Guideline 2: Make clear how well the system can do what it can do.
- **Description:** Set the right expectations about how often the AI system might make mistakes.
- **Examples:**
  - An AI chatbot that can prioritize items for a support team based on predefined urgency.

#### Guideline 3: Time actions based on context.
- **Description:** Decide when to use AI or internet based on the user's current task and environment.
- **Examples:**
  - An AI-powered system that can send push alerts.

**Step 1:** Select relevant guidelines
- Fill in each row below before going to the next step when comparing to the previous step.

**Step 2:** Imagine impact of relevant guidelines
- Fill in for each relevant guideline from the previous step.

**Step 3:** Draft implementation requirements
- Fill in for each high impact guideline from the previous step.

**Step 4:** Prioritize
- Select guidelines to implement.

**Step 5:** Track
- Or track with your own tools.

---

**Complete step 1 and then filter to “Yes” relevant before step 2**

**Filter to “High” impact before step 3**

**Track implementation progress here or in your own tools**
### The HAX Workbook

#### STEP 1: Select relevant guidelines

**Fill in each row below before going to the next step**

When complete, filter for relevance.

<table>
<thead>
<tr>
<th>GUIDELINE</th>
<th>DESCRIPTION</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Make clear what the system can do.</td>
<td>Set the right expectations about the AI system's capabilities.</td>
</tr>
<tr>
<td>2. Make clear how well the system can do what it can do.</td>
<td>Set the right expectations about how often the AI system might go wrong.</td>
</tr>
<tr>
<td>3. Time services based on context.</td>
<td>Decide when to use or not based on the user's current state and environment.</td>
</tr>
</tbody>
</table>

**Is this guideline relevant?**

Select Yes, No, Maybe, Already Done
Complete step 1 and then filter to "Yes" relevant before step 2
Filter to "High" impact before step 3
Track implementation progress here or in your own tools

**STEP 2: Imagine impact of relevant guidelines**

Fill in for each **relevant** guideline from the previous step

When complete, sort impact from high to low.

<table>
<thead>
<tr>
<th>GUIDELINE</th>
<th>EXAMPLES</th>
</tr>
</thead>
<tbody>
<tr>
<td>How might it affect a user if you <strong>apply</strong> this guideline?</td>
<td>How might it affect a user if you <strong>do not apply</strong> this guideline?</td>
</tr>
</tbody>
</table>

Type in your answer below

**Estimated impact on the user?**

Select High, Medium or Low

The HAX Workbook
Step 3: Draft implementation requirements
Fill in for each **high impact** guideline from the previous step

Brainstorm requirements for implementing this guideline.

**Include UI, AI, Data and Engineering requirements.**

Estimated resource commitment?

Select S, M, L, or XL
### The HAX Workbook

#### STEP 4: Prioritize
Select guidelines to implement

**Priority for guideline implementation?**

*Select from P0-P3*

<table>
<thead>
<tr>
<th>GUIDELINE</th>
<th>EXAMPLES</th>
<th>IS THIS GUIDELINE RELEVANT?</th>
<th>APPLYING THIS GUIDELINE MAY IMPACT A USER BY</th>
<th>NOT APPLYING THIS GUIDELINE MAY IMPACT A USER BY</th>
<th>ESTIMATED IMPACT ON THE USER?</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Make clear what the system can do. Set the right expectations about the AI system’s capabilities.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Make clear how well the system can do what it can do. Set the right expectations about how often the AI system might be mistaken.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Time services based on contact. Decide when to time or interrupt based on the user’s current task and environment.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Creating a voice-based virtual assistant that can call people

STEP 1.
Is this guideline relevant?
Yes, if the assistant is unsure whom to call, asking for clarification is less costly to a user than calling the wrong person.

Step 2.
How might applying this guideline (or not) impact users?
A person may become frustrated or embarrassed if the assistant calls someone unintended.

Step 3.
How might this guideline be implemented?
Ensure the AI model can compute its own uncertainty. When the AI is uncertain, disambiguate the user’s request by asking them to clarify their intent (e.g., “Do you mean John P. or John F.?“)

Step 4.
Should this guideline be prioritized?
Yes, the user benefits outweigh resource costs.
Using the HAX Workbook

When...

you have an idea for a new user-facing AI-based feature and you are starting to define requirements.

OR

you have an existing user-facing AI feature or prototype that you want to evolve or improve.

Don’t do this alone!

Involve PMs, User Researchers, Designers, Data Scientists and Engineers from your team or partner teams because implementing the guidelines can impact a system’s UI, data, and AI models.

Engage with users and stakeholders when using the workflow to better understand their needs and priorities.
Teams who have used the Workbook say

PM

This is a team alignment tool. We work on the client side, just trying to speak the same language as the intelligence side—having a framework to do that around [is] super helpful.

Engineer

I found a lot of points I hadn’t considered before, so it was helpful for me, I would use it as a checklist, when I develop a new feature to see if it abides by the Guidelines.

PM

As PMs, we need to frame things to bosses and other disciplines, and we need to get funding. If we can say, “We all agree that Guideline x is something we need to do,” that would be helpful for that pitch.

Designer

The guidelines serve as existing research evidence for my design decisions.
Best practices for how AI systems should behave during human-AI interaction

A tool to guide teams through planning and implementing human-AI interaction best practices

A set of flexible and reusable solutions to recurring human-AI interaction problems

A tool for generating scenarios to test based on likely human-AI interaction failures

More tools coming soon!

https://aka.ms/haxtoolkit
Why design patterns?
33 design patterns developed for 8 guidelines

G1-A: Introductory blurb
G1-B: Use explanation patterns
G1-C: Expose system controls
G1-D: Demonstrate possible system

G2-A: Match the communicated performance with the system’s – Language
G2-B: Match the communicated performance with the system’s – Numbers
G2-C: Report system

G9-A: Switch classification decisions
G9-B: Rich and detailed edits
G9-C: Undo automated actions
G9-D: Use G15
G9-E: Batch-editing

G10-A: Disambiguate before acting
G10-B: Avoid cold starts before eliciting user preferences
G10-C: Fall back to other strategies

G11-A: Local explanations
G11-B: Global explanations
G11-C: Present properties of system outputs
G11-D: Map system input attributes to ...

G14-A: Comprehensive updates
G14-B: Immediate, partial, non-disruptive updates

G15-A: Encourage explicit feedback on individual system outputs
G15-B: Request explicit feedback on selected system outputs
G15-C: Report inappropriate content ...

G16-A: Feedforward: Convey the consequences before user takes action
G16-B: Convey consequences immediately after user takes action
G16-C: Reconfirm consequences of past ...
Creating a voice-based virtual assistant that can call people

Patterns implementing Guideline 10:
G10-A: Disambiguate before acting
G10-B: Avoid cold starts by eliciting user preferences
G10-C: Fall back to other strategies
G10-A: Disambiguate before acting

Problem:
The AI system is uncertain of user intent and what further actions to take.

Solution:
Elicit clarification from the user before taking action to resolve the system’s uncertainty.

Use When:
• The system is able to compute its own uncertainty.
• The system is able to generate a list of probable alternate interpretations but is uncertain about which is correct.

How:
• Identify multiple options to show the user.
• Determine the uncertainty threshold that triggers disambiguation. When system confidence is extremely low, consider suppressing the feature altogether.
• To elicit clarification, ask a clarifying question or prompt the user to select from one or more probable options.
• …
Scope services when in doubt.

Engage in disambiguation or gracefully degrade the AI system’s services when uncertain about a user’s goals.

G10-A: Disambiguate before acting
Scope services when in doubt.

Engage in disambiguation or gracefully degrade the AI system’s services when uncertain about a user’s goals.

Do you mean call Satya N. or Satya K.?
Best practices for how AI systems should behave during human-AI interaction

A tool to guide teams through planning and implementing human-AI interaction best practices

A set of flexible and reusable solutions to recurring human-AI interaction problems

A tool for generating scenarios to test based on likely human-AI interaction failures

More tools coming soon!

https://aka.ms/haxtoolkit
Teams rarely tested AI experiences

“We focus more on hero kind of scenarios”

“There’s not as much of a discipline around going through and evaluating or auditing the error states, or... how you’re going to handle those errors”

“Its so hard to know until the bad things happen”
“This will standardize the error case design”

– PM

“[The Playbook will] force a consistent bar across different teams in different organizations.”

– PM

“We all feel like we’re sometimes speaking different languages, and [the Playbook]...puts everybody on the same level”

– Designer
Best practices for how AI systems should behave during human-AI interaction

A tool to guide teams through planning and implementing human-AI interaction best practices

A set of flexible and reusable solutions to recurring human-AI interaction problems

A tool for generating scenarios to test based on likely human-AI interaction failures

More tools coming soon!

https://aka.ms/haxtoolkit
Takeaways

Responsible AI is human-centered AI

Use the HAX Toolkit to build human-centered AI

Work with us!

https://aka.ms/haxtoolkit
Create human-centered AI with the Human-AI eXperience (HAX) Toolkit

Saleema Amershi, Ph.D.
Mihaela Vorvoreanu, Ph.D.