

# DESIGNING AI FOR PEOPLE - INSTRUCTOR GUIDE

Written by Evan Peck ([evan.peck@colorado.edu](mailto:evan.peck@colorado.edu))

## OVERVIEW

The goal of this activity is to engage with the social and ethical implications of designing AI systems for people (Human-AI Interaction). While this could easily fill a course, this activity's objective is to *start* building an understanding of the consequences of careless AI design. It can be completed in as short as an hour.

## BACKGROUND

Given the potential harms of AI, researchers at Microsoft codified 150 AI-related design recommendations and distilled them into 18 general design guidelines. Those guidelines were later validated through iteration with 49 design practitioners on 20 different applications. While no set of guidelines is necessarily authoritative, applying Microsoft's guidelines to familiar applications gives students an opportunity to not only wrestle with some of the design challenges in creating Human-AI systems, but to also consider the effectiveness of the guidelines themselves. The high-level structure of this activity is:

Task	Time
Pass out Human-AI Guidelines (from Microsoft) & Intro	10 min
Break students into small groups - groups are assigned an application	5 min
Each individual student is assigned a task	10 min
Groups come to consensus about criteria and assessment	10 min
Return to full group - groups report	15 min
Compare with expert assessments / reflect on	10 min

*Note: the times listed above are doable, but probably close to the minimum time you should spend on each section.*

## LINKS:

- [Link to Human-AI card PDF](#) - distribute a copy to each student group
- [Microsoft Research Paper that was published on the guidelines](#) - use to compare student responses with experts



Overview of the 18 Human-AI guidelines created by Microsoft.

## INSTRUCTIONS

### 1. PASS OUT HUMAN-AI GUIDELINES CARDS TO STUDENTS

[The cards include examples and further explanation.](#) For the following reasons, resist the temptation to explain guidelines in any depth:

- The assignment is designed to peer instruction in order to communicate the application and purpose of each design criteria.
- At the end of the assignment, students will critique the guidelines themselves - *how easy will this be for developers to pick up and apply to my context?* This question is more meaningful if students are forced to use the criteria without heavy scaffolding

### 2. BREAK STUDENTS INTO SMALL GROUPS BASED ON APPLICATION AREA.

Choose applications that many students have likely engaged with in some capacity. Depending on the application, I prefer the set of:

1. Music Recommenders (example application: Spotify)
2. Autocomplete (example application: Text Messaging)
3. Social Networks (example application: Instagram)
4. Voice Assistants (example applications: Alexa, Google, Siri)
5. Navigation (example applications: Google Maps, Apple Maps)

## 6. Web Search (example applications: Google, Bing, etc)

More application areas can be found [in the research paper](#). Depending on the size of the group, form groups of 3-5 students that are assigned to evaluate one application area. Pass out Human-AI Guidelines to each group.

*Tip: For larger classes, assign 2 groups to each application area. This will allow for interesting discussion between the groups who assessed the same application.*

## 3. ASSIGN INDIVIDUAL ROLES TO EACH TEAM MEMBER - INDIVIDUAL WORK

Each group should equally designate team members for assessment of 3-6 guidelines.

Instructions for the individual. Their job is to:

1. Understand and be able to explain what each guideline means and how to apply it.
2. Apply their individual guidelines to the assigned application: **RED** for a clear violation, **GREEN** for a clear application, and **BLUE** for *does not apply*. Students should include a 1-sentence explanation with each assessment.

*Tip: I find that this is most effective when done in shared spaces. I use a shared virtual space - a Google Spreadsheet seen below. A public version can be viewed at: <https://bit.ly/hai-activity>*

		GUIDELINE	DESCRIPTION	EXAMPLE	(1) Music Recommenders	(2) Autocomplete	(3) Social Networks
INITIALLY	1	Make clear what the system can do	Help the user understand what the AI system is capable of doing.	Mark red with a note for a clear violation or violation			
	2	Make clear how well the system can do what it can do.	Help the user understand how often the AI system may make mistakes.	Mark green with a note for a clear application or application			
DURING INTERACTION	3	Time services based on context.	Time when to act or interrupt based on the user's current task and environment.				
	4	Show contextually relevant information.	Display information relevant to the user's current task and environment.	Use blue for does not apply			
	5	Match relevant social norms.	Ensure the experience is delivered in a way that users would expect, given their social and cultural context.				
	6	Mitigate social biases.	Ensure the AI system's language and behaviors do not reinforce undesirable and unfair stereotypes and biases.				
WHEN WRONG	7	Support efficient invocation.	Make it easy to invoke or request the AI system's services when needed.				
	8	Support efficient dismissal.	Make it easy to dismiss or ignore undesired AI system services.				
	9	Support efficient correction.	Make it easy to edit, refine, or recover when the AI system is wrong.				
	10	Scope services when	Engage in disambiguation or gracefully degrade the AI				

## 4. PEER INSTRUCTION + GROUP CONSENSUS

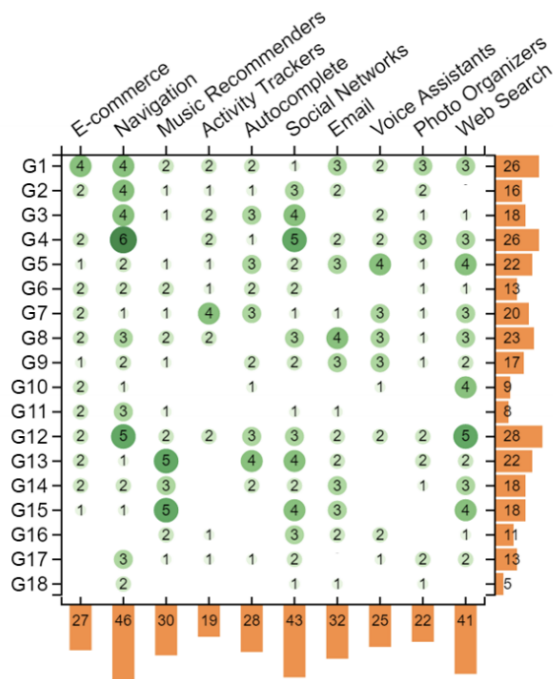
- Share with students that their group evaluations will be compared against those of experts (I find that this helps with
- *Peer instruction*: Each student reports back to their group (1) describing the criteria they focused on, and (2) how they used that criteria to evaluate their given application.
- Each group comes to a consensus about their evaluation of the

## 5. CLASS DISCUSSION

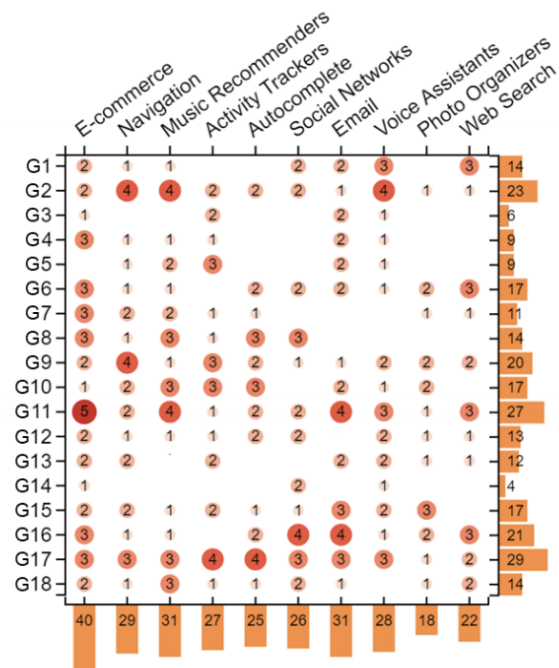
Bringing the class back together, discuss each application, and have the assigned groups report each (RED) clear violation.

## 6. COMPARE WITH EXPERTS

Once the class comes to consensus about their evaluations, reveal and compare their evaluations with those done by experts in [Amershi et al.](#)



(a) Counts of “clear application” or “application” responses.



(b) Counts of “clear violation” or “violation” responses.

Image from [research paper \(Amershi et al.\) that introduces and validates AI guidelines](#)

## POTENTIAL DISCUSSION AND/OR REFLECTION QUESTIONS:

- What is the reasoning behind the differences that we see between our student evaluations of AI products vs. expert evaluations of AI products?
- Even among experts, some dimensions appear to have significant disagreement (both violations and applications), how can you reconcile these differences?
- Imagine you are a developer who is receiving these guidelines for the first time (much like you). Which criteria would be most difficult to decide? What is it about the wording or design of the criteria that may have led to that ambiguity?
- Compare the criteria with [Google’s AI guidebook patterns](#). How are they different in the dimensions that they emphasize? How do you think that will inform or nudge design decisions in Google vs. Microsoft?

#### ALTERNATIVES TO THIS ACTIVITY:

*If you have more time and a larger class:* Use a group-oriented [jigsaw](#) approach. Assign 2 groups to each application. Those pairs of groups come together to debate and come to a consensus about how their evaluations differed (this might expose ambiguities in the practical application of some of the guidelines that are useful to draw out into conversation).

*If students are using AI in the class:* The activity could be reoriented towards student projects.

*If you want to focus on corporate responsibility:* While this activity leans on the guidelines created by Microsoft, it represents only *one* perspective from industry. An alternative twist to this activity could involve half the students performing the same evaluation using [Google's AI guidebook patterns](#). Evaluating the same applications from different perspectives will foster discussions of