Methods in AI Research
Chris' live lecture 2:
Scientific writing & Designing Responsible AI

Chris Janssen

c.p.janssen@uu.nl
www.cpjanssen.nl

Today's structure

1. Designing responsible AI
   a. Quiz questions that were harder for most → Seemed mostly OK
   b. Your questions
2. Group assignment: responsible AI
3. Scientific writing: Your Questions
4. Group assignment: writing
5. Remainder: available for 1-on-1 questions (I will stay on this call longer)

Let's make this a conversation..

Next week

- **New lab assignment**
  - **New Lecturers → will be assigned next week**
  - **Read assignment before class (after Sunday..)**
  - **Think about what you want to do**
  - **Keep it simple!**
- **Live lecture:**
  - **Bringing it all together**
  - **More info on exam**
  - **Quiz online to tell us what is less clear on connections or procedure → please fill out a bit earlier than usual**
  - **Preparatory assignment to follow today**

1A. Quiz questions Designing resp. AI

1B Your questions about Designing resp. AI

- **Procedural questions:**
  - (Dong's slide) Is this topic **exam material? → yes** (both my class and Dong's class)
  - Do we **need to read Lee's paper** as well or just learn their guidelines (principles) from the slides? → **no need to read**

1B Your questions about Designing resp. AI

- **Dong's content**
  - I am very **curious about current research about how to make make ML models more interpretable**, specifically regarding deep neural networks. I have seen a few approaches where people tried to find the reasoning behind individual layers by creating interpretable output as well as things like "deep dream", but I still don't know enough about it. It would be great to have more inspiration/input/guides to think about how to make our own future models understandable!
    - Dong will provide resources
    - Her course : Human-centered ML

1B Your questions about Designing resp. AI

- **Dong & Chris' content**
  - In the last few minutes of the last video, you mentioned that **'regulations are heavily debated'.** However, Chris mentioned in his video that **a regulation in the EU requires that AI should explain** its decision. If this regulation has passed, **which other regulations heavily debated?** → Details; relevant paper: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2903469

1B Your questions about Designing resp. AI

- **My content**
  - I find some of the **questions regarding** the **history of HCI** a little confusing. While the statement “according to the lecture” does limit the possible answers, there are nonetheless many more factors that could be mentioned too, e.g. the reduction of transistor size due to scientific progress in other fields + the increase of computational power that made many old ML algorithms feasible or economic stability combined with a culture of consumerism that benefited computer development (e.g. Nvidia, whose GPUs were primarily designed for games, but are now a vital part of many deep learning tasks (including cars and medical analysis)). Aren't these factors equally important?

1B Your questions about Designing resp. AI

- **My content**
  - More depth on the Lee et al as well as the Amershi et al principles please!
  - Could you maybe give an example of the principles applied to an automation which does not involve a safety-critical domain (such as automated vehicles/ automations)?
  - Next to all questions there is a statement that **we should be able to remember/list all of applicable principles/values/etc**.. **I find it difficult to just list them off the top of my head** because quite often they seem so obvious or so outside the scope I am thinking of that some multiple is. In the real world context I would likely use a reference when applying these principles for the first n times when applying them. Do you have any tips how to deal with that problem? Or do you want us to just blindly memorize them?

Re: last question. Go to:

https://aka.ms/aiguidelines

Overview Publications Groups

The Guidelines for Human-AI Interaction synthesize more than 20 years of thinking and research in human-AI interaction. Developed in a collaboration between Aether, Microsoft Research, and Office, the guidelines were validated through a rigorous, 4-step process described in the CHI 2019 paper, Guidelines for Human-AI Interaction. They recommend best practices for how AI systems should behave upon initial interaction, during regular interaction, when they're inevitably wrong, and over time.

We hope you can use these Guidelines for Human-AI Interaction throughout your design process as you evaluate existing ideas, brainstorm new ones, and collaborate with the multiple disciplines involved in creating AI.

People

Saleema Amershi
Principal Researcher

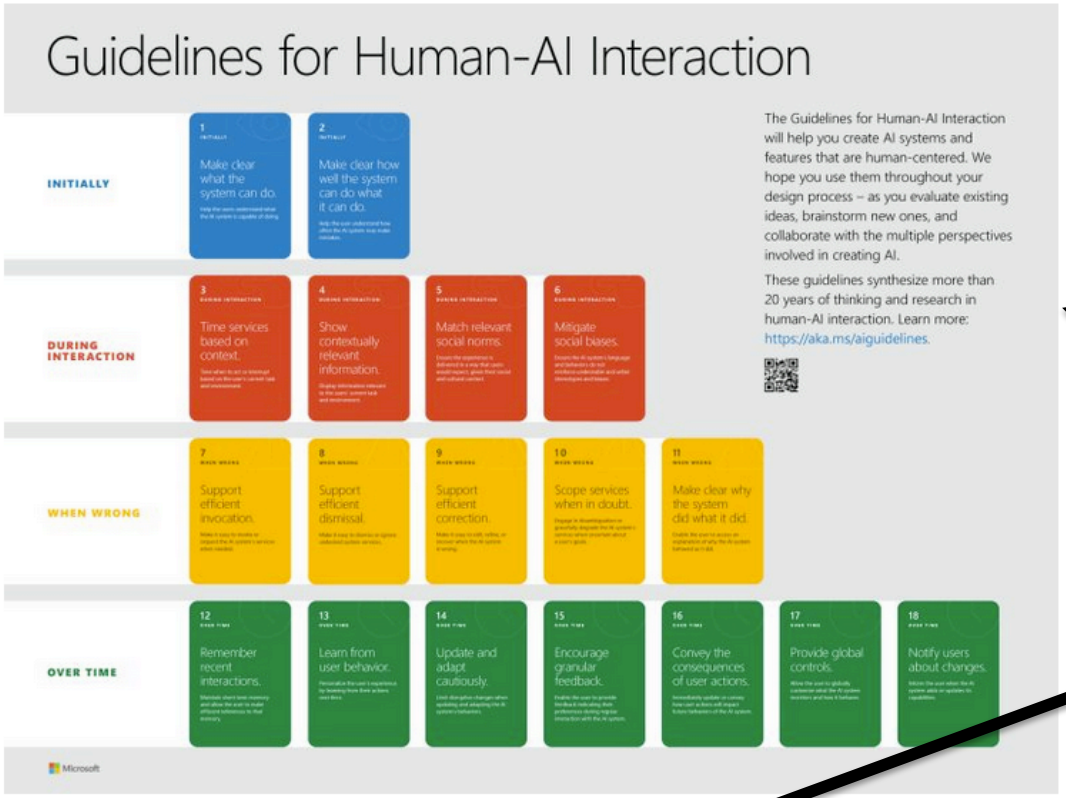
Paul Bennett
Partner Research Manager

Penny Collisson
Principal Design Research Manager
Microsoft

Derek DeBellis
Data Scientist II
Office Planning & Research

Re: last question. Go to:

https://aka.ms/aiguidelines



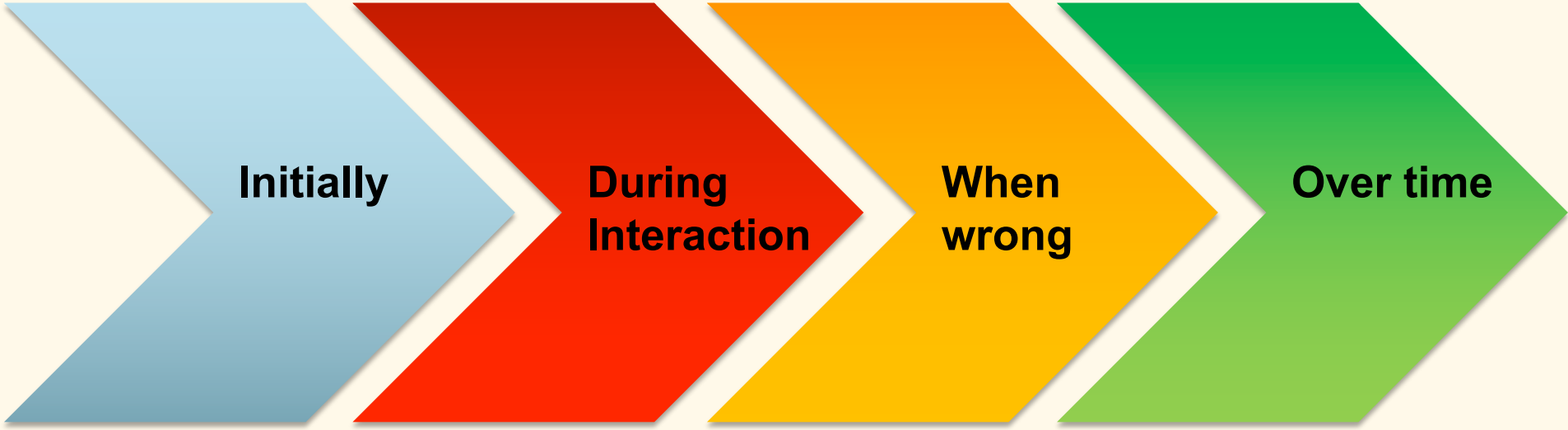
- Printable poster (PDF)
- Poster image (JPG)
- Printable cards – English (PDF)
- Printable cards – Chinese (PDF)
- Interactive cards with examples of the guidelines in practice

If you struggle with content:
Download poster and printable card deck for examples

(also even more material available)

Core message: 4 “top-level categories”

- **Think about design in all phases of product R&D + employment**



Group assignment

- **Before:**
  - **Think of an AI-infused technology that you use every day.**
    - write down the **application** you are thinking about
    - write down **2 guidelines that come from different "top-level categories"** (i.e., from "initially", "during interaction", "when wrong", or "over time") that you think apply to your application/setting.
    - write down for those guidelines whether they are **being adhered to or being violated** in your application and why
- **In group discussion:**
  - **Present ideas to each other and explain why the principles apply**
  - **(if time allows) come up with a 3rd principle from a category**
  - **Be prepared to report back 1 example**
- **Return to main room in XXX Minutes**

Discussion of your examples

My example

Initially

During Interaction

When wrong

Over time

My example

Initially | During Interaction | When wrong | Over time

Initially: 1 Make clear what system can do

My example

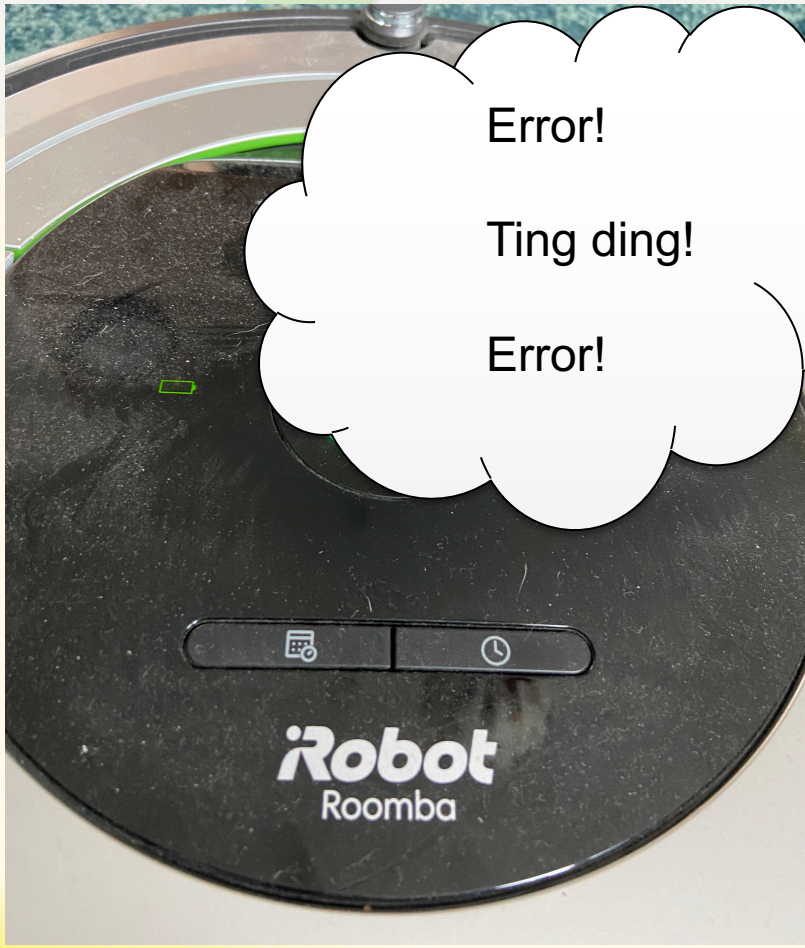
Initially | During Interaction | When wrong | Over time

During interaction:

4. Show contextually relevant information

A. Function that runs highlight

B. Error info



Match relevant social norms”
or “Mitigate social bias”:
(slightly) Female (robotic) voice?

My example

Initially | During Interaction | When wrong | Over time

When wrong

9. Support efficient correction

My example

Initially | During Interaction | When wrong | Over time

Over time

None seem to apply..

Little learning by my robot, no software updates… (different for different robots)

Today's structure

1. Designing responsible AI
   a. Quiz questions that were harder for most → Seemed mostly OK
   b. Your questions
2. Group assignment: responsible AI
3. **Scientific writing: Your Questions**
4. **Group assignment: writing**
5. **Remainder: available for 1-on-1 questions (I will stay on this call longer)**

Let's make this a conversation..

3 Your questions about writing

- **Process:**
  - I use **Grammarly** for more than 5 years, and it is really helpful. Tell others in the live lecture that there is a **Google Chrome extension** (then it even checks your spelling in emails you are writing)
  - I wish we had something like this **before starting work on our thesis**
  - **Can you recommend** specific well-written articles or perhaps journals **for meta-reading**?
    - Personal preference and depending on field, but..
    - See appendix A and C of ”assignment 2”

3 Your questions about writing

- **Process:**
  - You said that everyone is responsible for the group work [..] makes sense for poor writing [..]. However, if someone in a group would **plagiarize** their section it is much harder to find out for the group members. If this were to happen, **would the group members also be responsible for plagiarism**. I don't expect this to be a relevant issue for any group, but it would be nice to know what happens should this situation occur.
  - Are there any specific meta-writing guidelines we should be following from the faculty when writing our assignments?

3 Your questions about writing

- **Content**

  “Last week during the 10 minutes of discussion in groups I noticed that many different students held different beliefs about what a hypothesis should look like for example. I feel like there's different ways to form different aspects of papers in different faculties/schools/levels (for example a high school chemistry paper will look different from an AI master paper). How long can a hypothesis be? ”

3 Your questions about writing

- **Content:**
  - Aren’t preferences for title also a bit personal?
  - I found the topic of topic sentences quite interesting. How broadly is this being used in the scientific community? Do you, for example, use it for every paragraph? Or just sometimes, when the 'article needs it’.

3 Your questions about writing

- **Audiences:**
  - How many different audiences should we try to satisfy. I know trying to satisfy everyone is bad but what about picking 3-4 targets?
  - Sometimes writing for multiple audiences with different skill levels at the same time. Take for instance the UU examiner AND the 2nd examiner (for instance the bachelor thesis (which wasn't necessarily focussed on publishing in a certain article). The 2nd examiner could be (as you said as well) from an entirely different field and therefore would require more/different explanations. Of course, if you are writing for a specific journal, or for your work, it's a clear case. But what if it's not? So here is my specific question: In the case that you identified more than one audience (and thus more than one skill level), should you then **'step down' to the lowest level of knowledge, or go somewhere in between?**

3 Your questions about writing

- **Bigger picture:**
  - Do you have any general tips for the introduction part and the discussion part? Mostly the structure of these two parts. I personally think these are the hardest.

  ===> after group assignment

Assignment

- Let's imagine you are conducting research on an intelligent tutoring system that is used by high school students. Answer these three questions briefly for your audience group:

  (1) What are their **knowledge and background**?
  (2) What are their **interests** (related to your research)?
  (3) What are their **expectations** of a report (in e.g., form, style, type of content that is covered, way that arguments are supported, etc)

- Now: discuss with each other what you have for each audience group. How do the two groups differ?

- Take XXXX minutes

Discussion assignment

- Students from groups 1, 3, 5, 7, 9: Engineers
  Students from groups 2, 4, 6, 8, 10: Journalist

- Students from groups 11, 13, 15, 17, 19: IT Consultancy firm
  Students from groups 12, 14, 16, 18, 20: Designers (e.g., interaction designers)

- Students from groups 21, 23, 25, 27, 29, 31, 33: Psychologist
  Students from groups 22, 24, 26, 28, 30, 32, 34: Staff of the ministry of education

3 Your questions about writing

- **Bigger picture:**
  - Do you have any general tips for the introduction part and the discussion part? Mostly the structure of these two parts. I personally think these are the hardest.

Tips: Use hourglass structure

Introduction

Middle part
(“the meat”, ”the specifics”)

General Discussion

Tips

- **What I expect to find in an introduction and general discussion**
  - **See assignment 2, section at end on “report”**

Introduction: goals

- **Overall goal: *Convince* (in a scientific way)**
  - **this is an interesting, relevant problem**
  - **that is tackled using an interesting, relevant method**
  - **the reader to continue reading (if this is relevant)**

Introduction: Heuristics

- **First 3 - 4 paragraph (hourglass model)**
  1: What is bigger problem?
  2: What is specific problem?
  3: What will you do specifically?
  4: Summary of what comes next
- **Contextualize problem with literature**
  - No need to explain every detail of every study
  - More detail on very relevant study. Especially for replication studies
- **(for empirical work) Finish with a clear research question or hypothesis**
  - Discuss how that relates to literature (why do you have that hypothesis?)
- **Use subheaders for important themes**

General discussion: goals

- **Overall goal: *Contextualize & interpret***
  - **What is main finding? Overarching pattern**
  - **Your findings in other literature**
  - **Your findings in light of practice/application**
  - **What limitations might limit ability to conclude (and future work might explore)**

General discussion: Heuristics

- **Have sections on:**
  - **Summary of result (in context of literature)**
  - **Implications (for theory and/or practice)**
  - **Limitations and future work**
  - **Conclusion** (side-note: I like that short and snappy!)

- **Do not over-claim or over-interpret results**

- **Demonstrate how your work moves science (and practice) forward: why, how, to what degree**

Next week

- Topics
  - New assignment for lab (read upfront)
  - No prerecorded lectures (catch up on other material)
  - Live lecture: integration of ideas so far
    - Ask questions using quiz (early deadline)
    - Do preparatory assignment

Today's structure

1. Designing responsible AI
   a. Quiz questions that were harder for most → Seemed mostly OK
   b. Your questions
2. Group assignment: responsible AI
3. Scientific writing: Your Questions
4. Group assignment: writing
5. Remainder: available for 1-on-1 questions (I will stay on this call longer)

Let's make this a conversation..