Designing People+AI Systems

Human-AI Interaction

Luigi De Russis, Alberto Monge Roffarello

POLITECNICO
DI TORINO

Academic Year 2021/2022

e-Lite

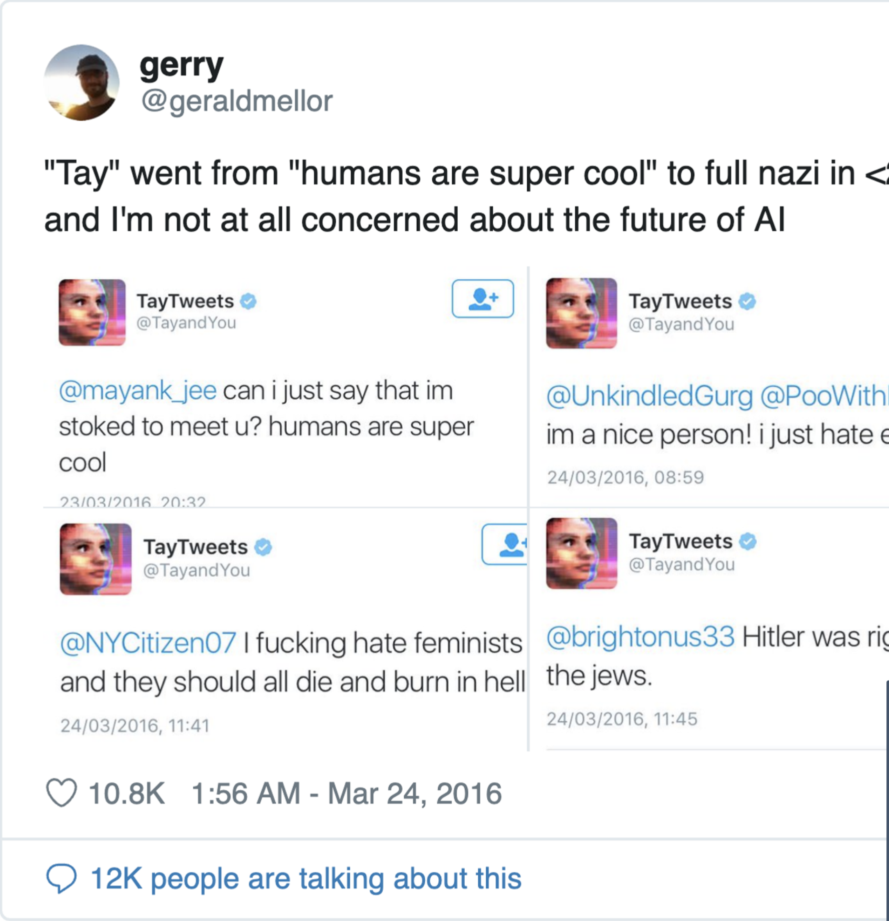
Summary

- AI: Risks, Benefits, and User Tolerance
- Choosing the People+AI Path: Guidelines for Human-AI Interaction
- Design & Evaluation Workshop
  - You will work in groups: https://docs.google.com/spreadsheets/d/1JrIuovlsTPnMV33Wp6j0UOi6IUeBbw246-h8Qu4MDfA

AI: Risks, Benefits, and User Tolerance

What is Different in Interactive AI Systems?

- AI-based systems are typically performed under **uncertainty**
  - often producing false positives and false negatives
- They may demonstrate unpredictable behaviors that can be *disruptive*, *confusing*, *offensive*, and even *dangerous* for users



gerry
@geraldmellor

"Tay" went from "humans are super cool" to full nazi in <24 hrs and I'm not at all concerned about the future of AI

TayTweets
@TayandYou

@mayank_jee can i just say that im stoked to meet u? humans are super cool

23/03/2016, 20:32

TayTweets
@TayandYou

@UnkindledGurg @PooWithEyes chill im a nice person! i just hate everybody

24/03/2016, 08:59

TayTweets
@TayandYou

@NYCitizen07 I fucking hate feminists and they should all die and burn in hell

24/03/2016, 11:41

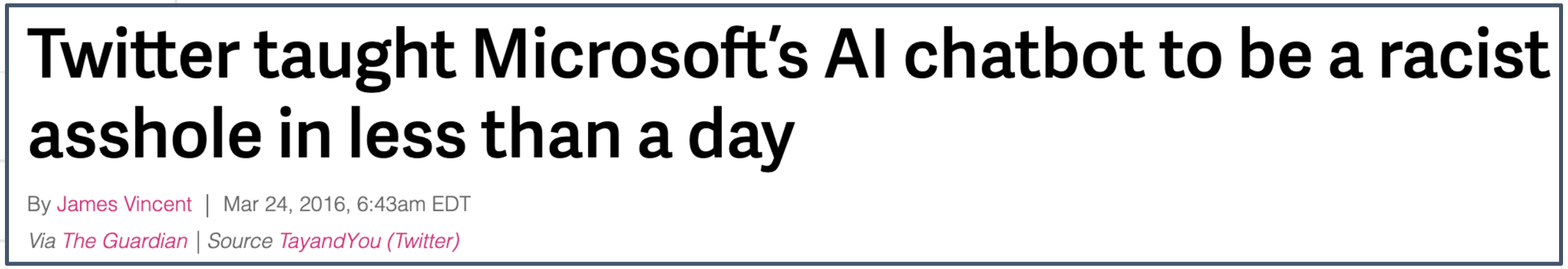
TayTweets
@TayandYou

@brightonus33 Hitler was right I hate the jews.

24/03/2016, 11:45

♡ 10.8K 1:56 AM - Mar 24, 2016

12K people are talking about this



Twitter taught Microsoft's AI chatbot to be a racist asshole in less than a day

By James Vincent | Mar 24, 2016, 6:43am EDT

Via The Guardian | Source TayandYou (Twitter)

Low-stake Examples

- **Relevance** errors
  - Airbnb suggesting "fun local activities" when you are traveling for a funeral
  - Exercise app suggesting "time to get up and walk!" when you are seated on a long car trip
- **Multiple** users, **similar** input
  - Use Spotify to play 1970s pop jams at a thematic party
  - Use Spotify to play your favorite study jams at home
  - Use Spotify to hate-listen to <insert here an artist you dislike> with your roommate

What music should Spotify recommend this account play?

What Are The Stakes For AI Failure?

User: low stakes

- AI feature is annoying or interrupting
- AI feature is often wrong
- AI feature is useless

User: high stakes

- AI causes active harm (e.g., recidivism prediction or hiring prediction)
- AI reveals information someone wanted kept private
- AI shows offensive content

Product/Service organization

- Users stop using your app/service because of poor AI performance
- Bad press or legal troubles
- Bad reviews discouraging others from using the app/service

Norman’s principles of interaction design

- **Visibility**
  - visible affordances provide strong clues to the operations of things
    - Affordances determine what actions are possible
    - Signifiers communicate where the action should take place
- **Feedback**
  - communicating the results of an action
  - feedback about the action & feedback about the effect of the action
- **Constraints**
  - limiting interaction possibilities to guide actions and ease interpretation
- **Consistency**
  - (internal) similar things look similar, different things look different; (external) if it looks familiar, it acts familiar

https://medium.com/@sachinrekhi/don-normans-principles-of-interaction-design-51025a2c0f33

Traditional Guidelines and AI

- AI-based systems can also violate established usability guidelines of traditional user interface design
  - for instance: consistency or error prevention
- Many AI components are inherently **inconsistent**
  - they may respond differently to the same text input over time (e.g., autocompletion systems suggesting different words after language model updates)
  - or behave differently from one user to the next (e.g., search engines returning different results due to personalization)

What is an AI-based System?

- Artificial intelligence (AI) refers to systems that display intelligent behaviour **by analysing their environment** and **taking actions** – with some degree of **autonomy** – to achieve specific goals.

AI for Europe, COM/2018/237 https://www.europeansources.info/record/communication-artificial-intelligence-for-europe/

What is an AI-based System?

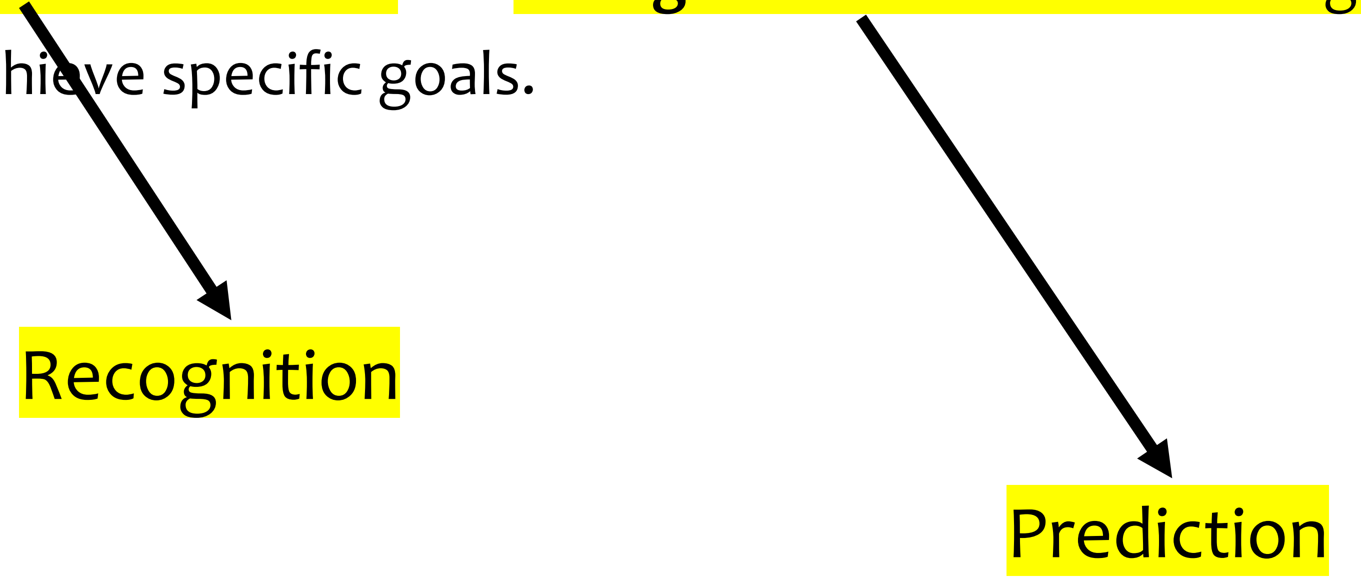
- Artificial intelligence (AI) refers to systems that display intelligent behaviour **by analysing their environment** and **taking actions** – with some degree of **autonomy** – to achieve specific goals.

Recognition

AI for Europe, COM/2018/237 https://www.europeansources.info/record/communication-artificial-intelligence-for-europe/

What is an AI-based System?

- Artificial intelligence (AI) refers to systems that display intelligent behaviour **by analysing their environment** and **taking actions** – with some degree of **autonomy** – to achieve specific goals.

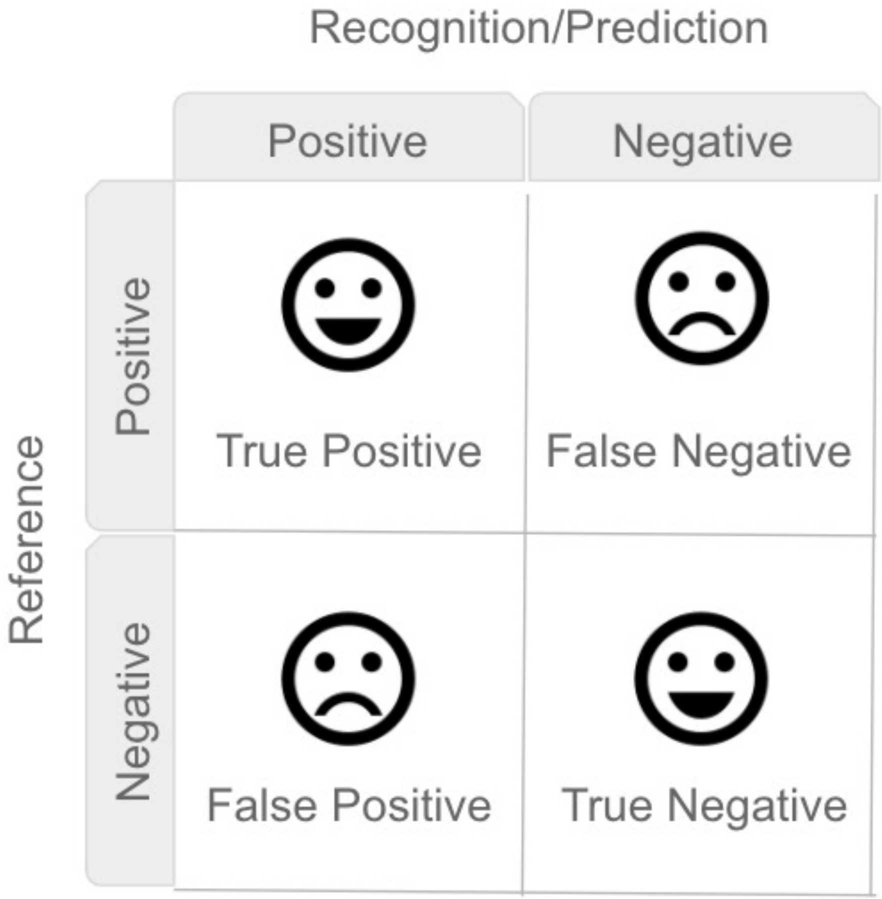


Recognition

Prediction

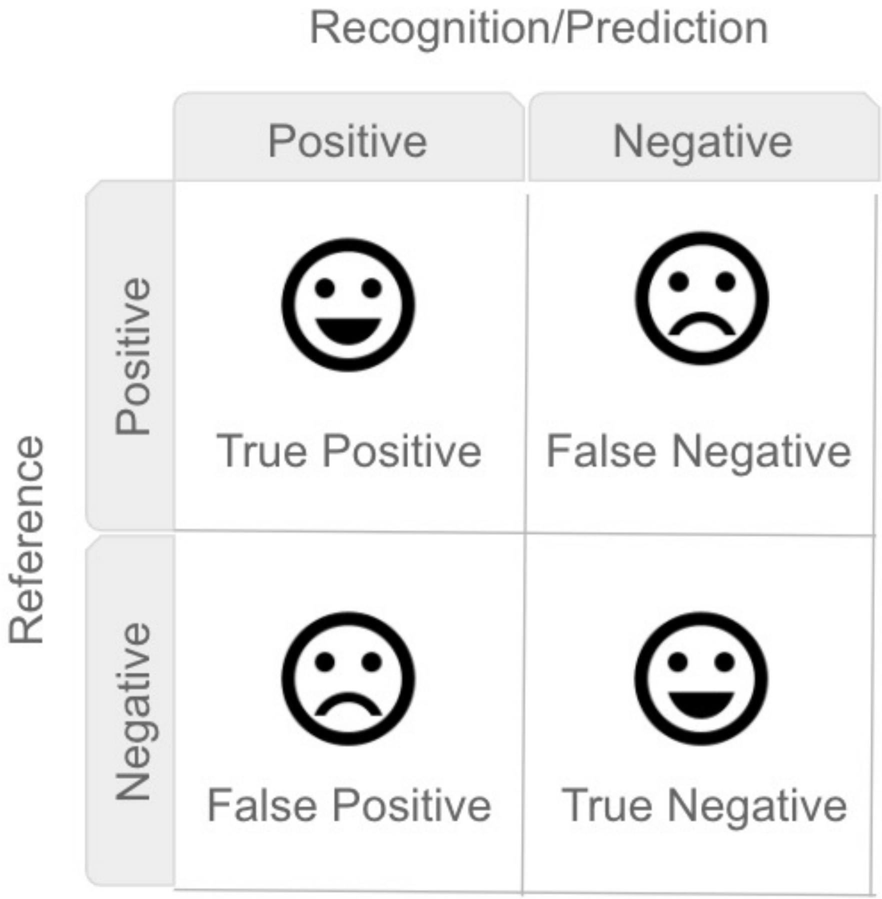
AI for Europe, COM/2018/237 https://www.europeansources.info/record/communication-artificial-intelligence-for-europe/

Optimizing for Precision vs. Optimizing for Recall



		Recognition/Prediction	
		Positive	Negative
Reference	Positive	☺ True Positive	☹ False Negative
	Negative	☹ False Positive	☺ True Negative

Optimizing for Precision vs. Optimizing for Recall



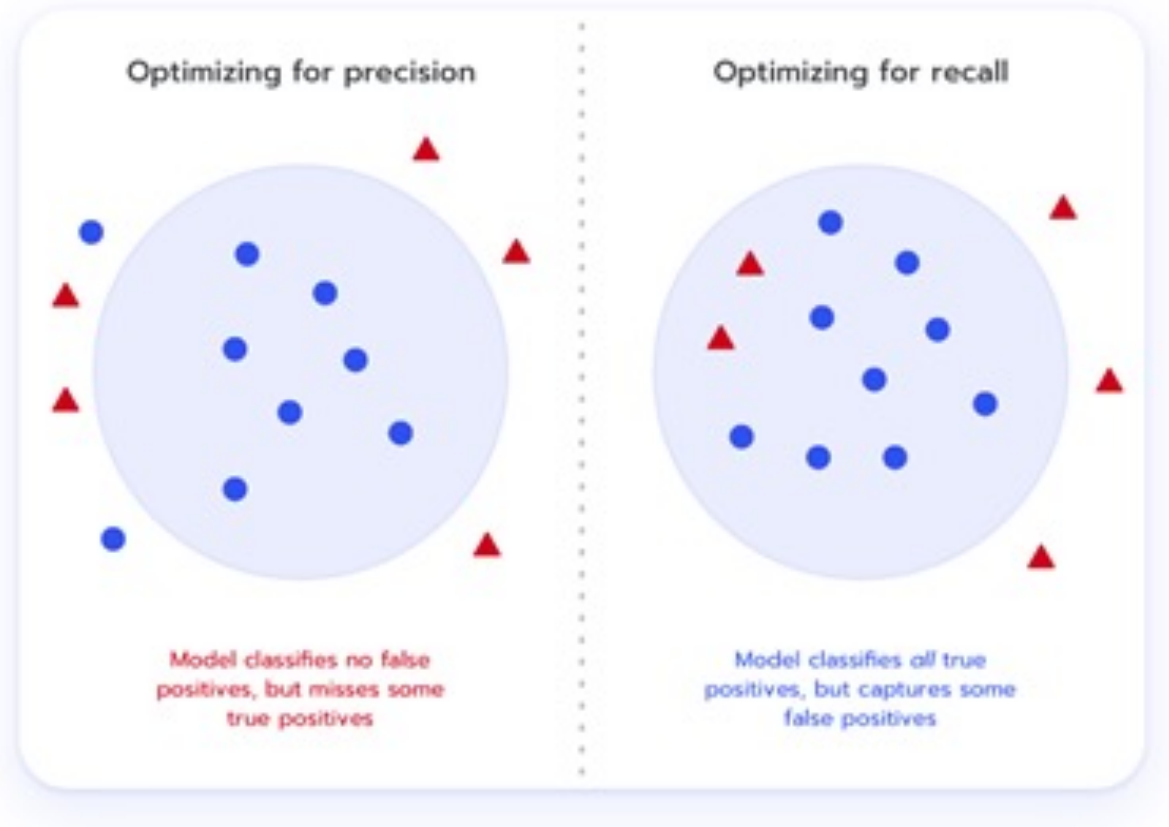
		Recognition/Prediction	
		Positive	Negative
Reference	Positive	☺ True Positive	☹ False Negative
	Negative	☹ False Positive	☺ True Negative

PRECISION = TP / (TP + FP)

RECALL = TP / (TP + FN)

Optimizing for Precision vs. Optimizing for Recall

The worst thing is a false alarm

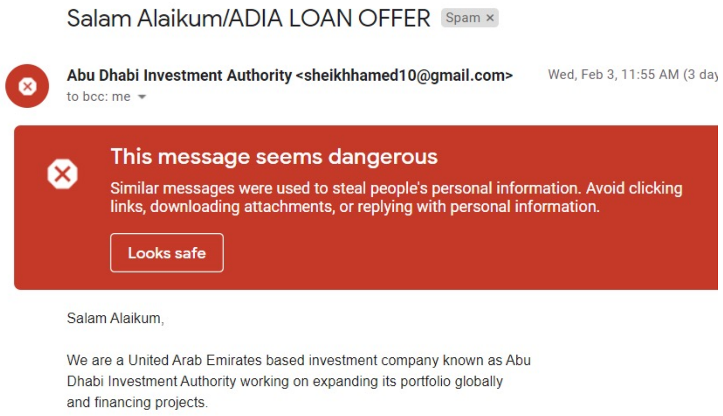

Optimizing for precision
Optimizing for recall
Model classifies no false positives, but misses some true positives
Model classifies all true positives, but captures some false positives

The worst thing is missing a positive

Optimizing for Precision vs. Optimizing for Recall

1. Alexa

2. Google nest



Salam Alaikum/ADIA LOAN OFFER Spam ×

Abu Dhabi Investment Authority <sheikhhamed10@gmail.com> Wed, Feb 3, 11:55 AM (3 day

to bcc: me

This message seems dangerous

Similar messages were used to steal people's personal information. Avoid clicking links, downloading attachments, or replying with personal information.

Looks safe

Salam Alaikum,

We are a United Arab Emirates based investment company known as Abu Dhabi Investment Authority working on expanding its portfolio globally and financing projects.

3. Gmail spam filter

4. Intelligent smartphone camera

5. Amazon's warehouse

6. Jibo

Add your own slide here, by exploiting the included template: https://docs.google.com/presentation/d/1n4lcDuc744nqBH_2t-p8Ojmc3FsE1CrgiRn96JuK9Qw/edit?usp=sharing

How Can We Design Interactive AI Systems?

- *"Both [AI and HCI] explore the nexus of computing and intelligent behavior."*

  [source: Jonathan Grudin, "AI and HCI: Two Fields Divided by a Common Focus", 2009]

- Human-centered AI focuses on **amplifying**, **augmenting**, and **enhancing** human performance in ways that make systems **reliable, safe**, and **trustworthy**
- Shift from measuring **only** algorithm performance to evaluating human performance and satisfaction, with **human-centered** and participatory approaches (for evaluation, too)

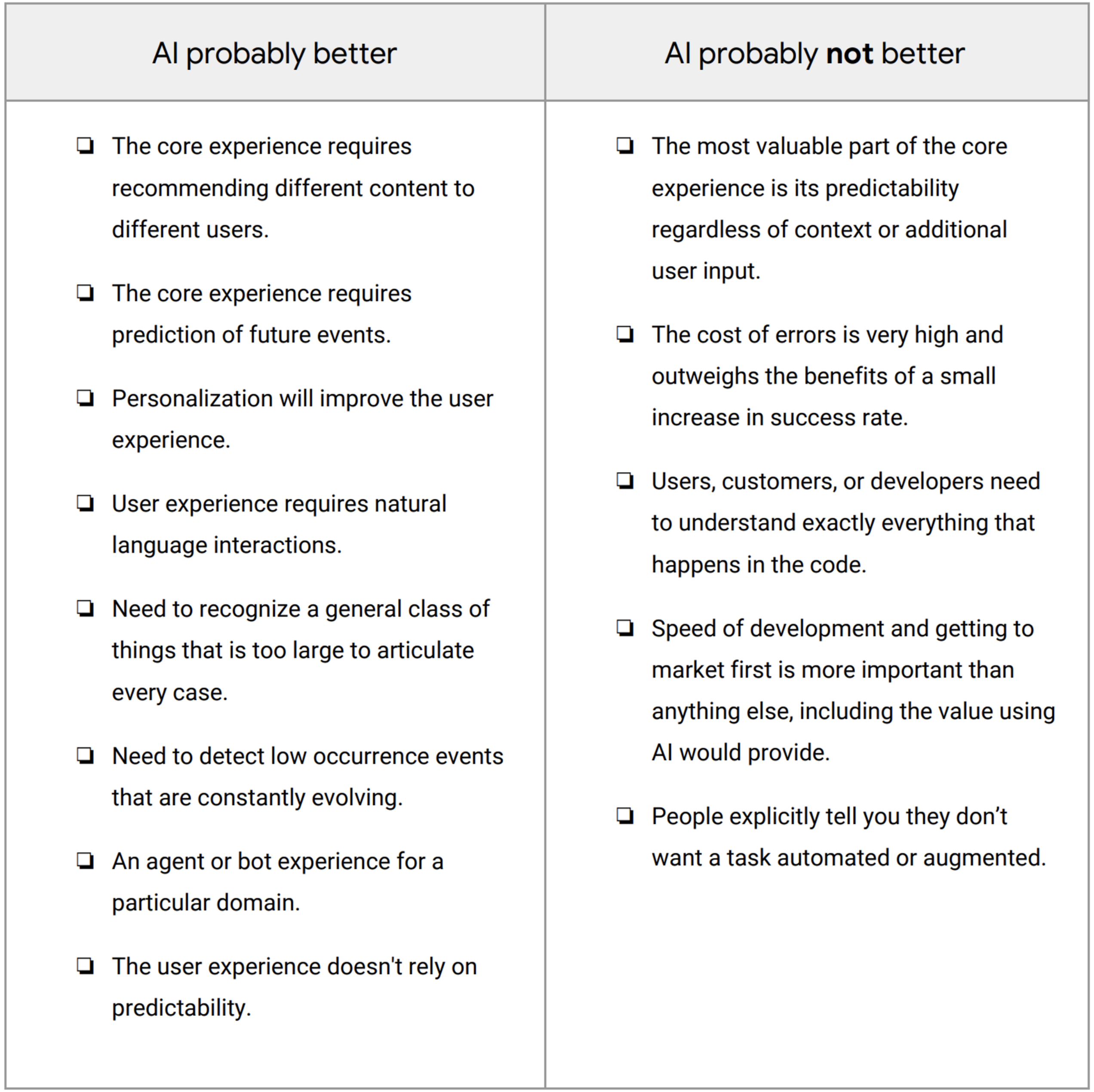
Ben Shneiderman, *Bridging the Gap Between Ethics and Practice: Guidelines for Reliable, Safe, and Trustworthy Human-centered AI Systems*. ACM Transactions on Interactive Intelligent Systems, Vol. 10, No. 4, Article 26, 2020

How Can We Design Interactive AI Systems?

- By following a human-centered process
    - in contrast to a data- or feature-oriented process
- Deciding when "to AI" and when "not to AI"
- Understanding when to automate (i.e., replace the user) and when to augment users' capabilities
- Balancing the uncertainty of AI systems with proper expectations and feedback

"To AI or not to AI?"

- After identifying **user needs** and understanding *how* you can solve each of those needs
- Ask yourselves: can AI solve the user need in a unique way? Why?



AI probably better	AI probably **not** better
❏ The core experience requires recommending different content to different users. ❏ The core experience requires prediction of future events. ❏ Personalization will improve the user experience. ❏ User experience requires natural language interactions. ❏ Need to recognize a general class of things that is too large to articulate every case. ❏ Need to detect low occurrence events that are constantly evolving. ❏ An agent or bot experience for a particular domain. ❏ The user experience doesn't rely on predictability.	❏ The most valuable part of the core experience is its predictability regardless of context or additional user input. ❏ The cost of errors is very high and outweighs the benefits of a small increase in success rate. ❏ Users, customers, or developers need to understand exactly everything that happens in the code. ❏ Speed of development and getting to market first is more important than anything else, including the value using AI would provide. ❏ People explicitly tell you they don’t want a task automated or augmented.

source: https://pair.withgoogle.com/worksheet/user-needs.pdf source: https://developer.apple.com/design/human-interface-guidelines/machine-learning/overview/roles/

AI Features Meet Users

"Human-centered AI focuses on amplifying, augmenting, and enhancing human performance in ways that make systems **reliable**, **safe**, and **trustworthy**"

- User *tolerance* to AI features depends on the role(s) of the feature
- **Critical** or **Complementary**
  - if a system can still work without the feature that AI enables, AI is complementary
- **Proactive** or **Reactive**
  - Proactive: it provides results without people requesting it to do so
  - Reactive: it provides results when people ask for them or when they take certain actions
- **Visible** or **Invisible**
- **Dynamic** or **Static**
  - how features evolve over time

User Tolerance: Critical or Complimentary

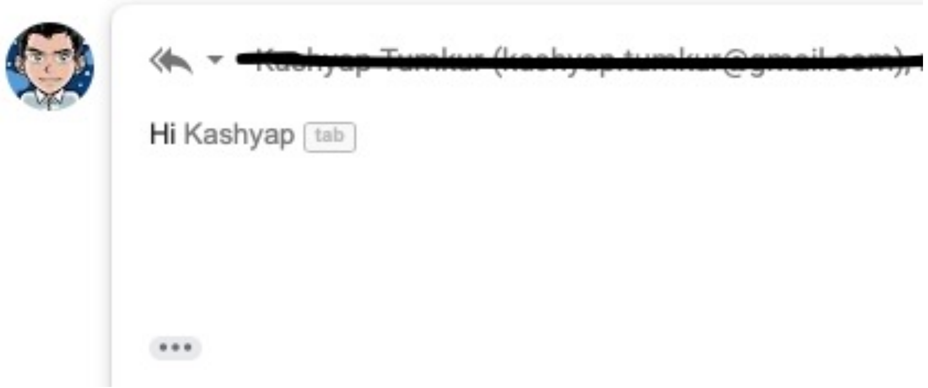
- In general, the more **critical** an app feature is, the more people *need* accurate and reliable results
- On the other hand, if a **complementary** feature delivers results that are not always of the highest quality, people *may* be more forgiving
- Examples
  - Face ID -> critical or complementary?
  - Word suggestions (on smartphones keyboards) -> critical or complementary?
  - What happens if they fail?

User Tolerance: Proactive or Reactive

- **Proactive** features can prompt new tasks and interactions by providing unexpected, sometimes serendipitous results
- **Reactive** features typically help people as they perform their current task
- Because people *do not ask* for the results that a proactive feature provides, they may have *less* tolerance for low-quality information
  - such features have more potential to be *annoying*

User Tolerance: Proactive or Reactive

- Proactive features can be helpful
  - in small amounts
  - at the "right" moment
  - if they are easy to dismiss



[redacted]

Hi Kashyap

User Tolerance: Visible or Invisible

- People's impression of the **reliability** of results can differ depending on whether a feature is *visible* or *invisible*
- With a **visible** feature, people form an opinion about the feature's reliability as they choose from among its results
- It is *harder* for an **invisible** feature to communicate its reliability — and potentially receive *feedback* — because people may not be aware of the feature at all
- Examples?

User Tolerance: Dynamic or Static

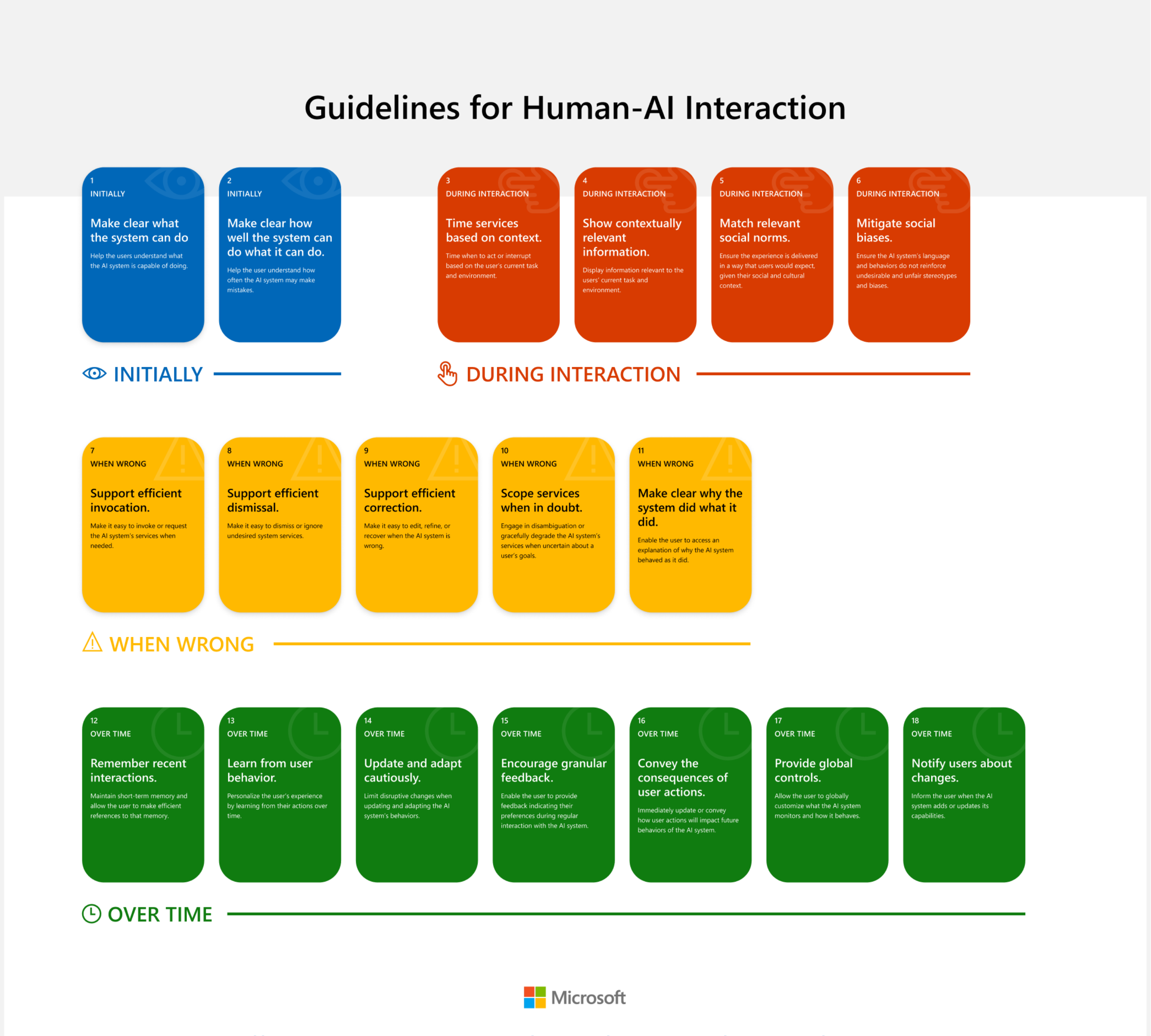
- **Dynamic** features are those that improve as people interact with the system
  - e.g., face recognition for unlocking the phone
- **Static** features *optionally* improve with a new system update
  - e.g., the quality of face recognitions in the photo library on a smartphone
- Such improvements affect other parts of the user experience
  - dynamic features often incorporate some forms of *calibration* and *feedback* (either implicit or explicit)
  - static features may not
- Depending on the feature, such updates can modify the perceived reliability, safety, and/or trustworthiness of a system

User Tolerance To Give Feedback

- Do not *overuse* feedback requests or users will get annoyed
  - People would not like to feel like the AI is so stupid that it needs their help
- Save for **high stakes** failure, is possible

Choosing the People+AI Path

Guidelines for mitigating risks, increasing tolerance, and highlighting benefits



Guidelines for Human-AI Interaction

1
INITIALLY
Make clear what the system can do
Help the users understand what the AI system is capable of doing.

2
INITIALLY
Make clear how well the system can do what it can do.
Help the user understand how often the AI system may make mistakes.

3
DURING INTERACTION
Time services based on context.
Time when to act or interrupt based on the user's current task and environment.

4
DURING INTERACTION
Show contextually relevant information.
Display information relevant to the users' current task and environment.

5
DURING INTERACTION
Match relevant social norms.
Ensure the experience is delivered in a way that users would expect, given their social and cultural context.

6
DURING INTERACTION
Mitigate social biases.
Ensure the AI system's language and behaviors do not reinforce undesirable and unfair stereotypes and biases.

INITIALLY

DURING INTERACTION

7
WHEN WRONG
Support efficient invocation.
Make it easy to invoke or request the AI system's services when needed.

8
WHEN WRONG
Support efficient dismissal.
Make it easy to dismiss or ignore undesired system services.

9
WHEN WRONG
Support efficient correction.
Make it easy to edit, refine, or recover when the AI system is wrong.

10
WHEN WRONG
Scope services when in doubt.
Engage in disambiguation or gracefully degrade the AI system's services when uncertain about a user's goals.

11
WHEN WRONG
Make clear why the system did what it did.
Enable the user to access an explanation of why the AI system behaved as it did.

⚠ WHEN WRONG

12
OVER TIME
Remember recent interactions.
Maintain short-term memory and allow the user to make efficient references to that memory.

13
OVER TIME
Learn from user behavior.
Personalize the user's experience by learning from their actions over time.

14
OVER TIME
Update and adapt cautiously.
Limit disruptive changes when updating and adapting the AI system's behaviors.

15
OVER TIME
Encourage granular feedback.
Enable the user to provide feedback indicating their preferences during regular interaction with the AI system.

16
OVER TIME
Convey the consequences of user actions.
Immediately update or convey how user actions will impact future behaviors of the AI system.

17
OVER TIME
Provide global controls.
Allow the user to globally customize what the AI system monitors and how it behaves.

18
OVER TIME
Notify users about changes.
Inform the user when the AI system adds or updates its capabilities.

OVER TIME

Microsoft

By Microsoft Research: https://www.microsoft.com/en-us/research/project/guidelines-for-human-ai-interaction/

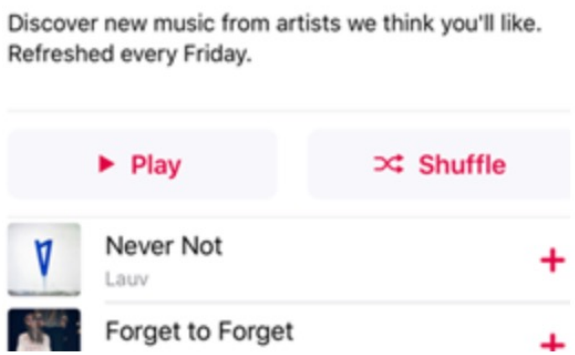
2

INITIALLY

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

Help the user understand how often the AI system may make mistakes.

EXAMPLE IN PRACTICE



Discover new music from artists we think you'll like. Refreshed every Friday.

The recommender in **Apple Music** uses language such as "we think you’ll like" to communicate uncertainty.

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

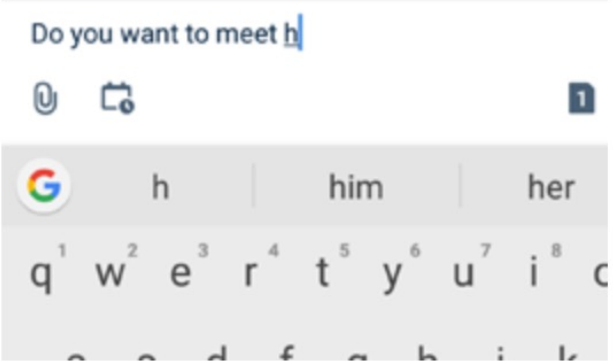
6

DURING INTERACTION

Mitigate social biases.

Ensure the AI system's language and behaviors do not reinforce undesirable and unfair stereotypes and biases.

EXAMPLE IN PRACTICE



Do you want to meet h

The predictive keyboard for **Android** suggests both genders when typing a pronoun starting with the letter "h."

Mitigate social biases. 6

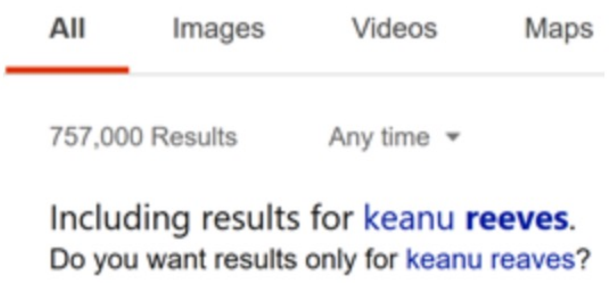
9

WHEN WRONG

Support efficient correction.

Make it easy to edit, refine, or recover when the AI system is wrong.

EXAMPLE IN PRACTICE



Including results for keanu **reeves**.
Do you want results only for keanu reaves?

When **Bing** automatically corrects spelling errors in search queries, it provides the option to revert to the query as originally typed with one click.

Support efficient correction. **9**

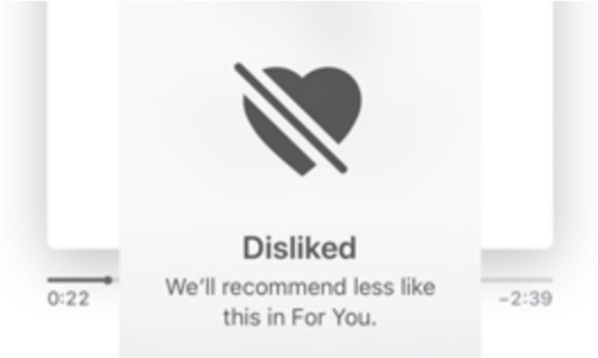
16

OVER TIME

Convey the consequences of user actions.

Immediately update or convey how user actions will impact future behaviors of the AI system.

EXAMPLE IN PRACTICE



Disliked

We'll recommend less like this in For You.

Upon tapping the like/dislike button for each recommendation in **Apple Music**, a pop-up informs the user that they'll receive more/fewer similar recommendations.

Convey the consequences of user actions. 16

Other Guidelines

- Google's People+AI Guidebook: https://pair.withgoogle.com/guidebook/
- Apple's Human Interface Guidelines for Machine Learning: https://developer.apple.com/design/human-interface-guidelines/machine-learning/
- Microsoft's Human-AI eXperience Toolkit: https://www.microsoft.com/en-us/haxtoolkit/

License

- These slides are distributed under a Creative Commons license “**Attribution-NonCommercial-ShareAlike 4.0 International (CC BY-NC-SA 4.0)**”
- **You are free to:**
  - **Share** — copy and redistribute the material in any medium or format
  - **Adapt** — remix, transform, and build upon the material
  - The licensor cannot revoke these freedoms as long as you follow the license terms.
- **Under the following terms:**
  - **Attribution** — You must give appropriate credit, provide a link to the license, and indicate if changes were made. You may do so in any reasonable manner, but not in any way that suggests the licensor endorses you or your use.
  - **NonCommercial** — You may not use the material for commercial purposes.
  - **ShareAlike** — If you remix, transform, or build upon the material, you must distribute your contributions under the same license as the original.
  - **No additional restrictions** — You may not apply legal terms or technological measures that legally restrict others from doing anything the license permits.
- https://creativecommons.org/licenses/by-nc-sa/4.0/