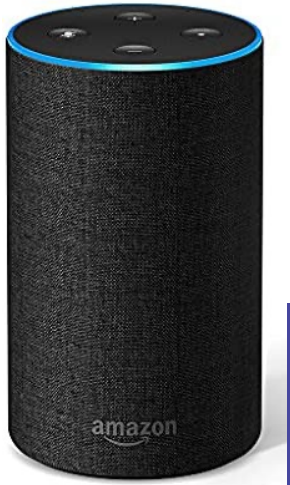
Interacting with AI

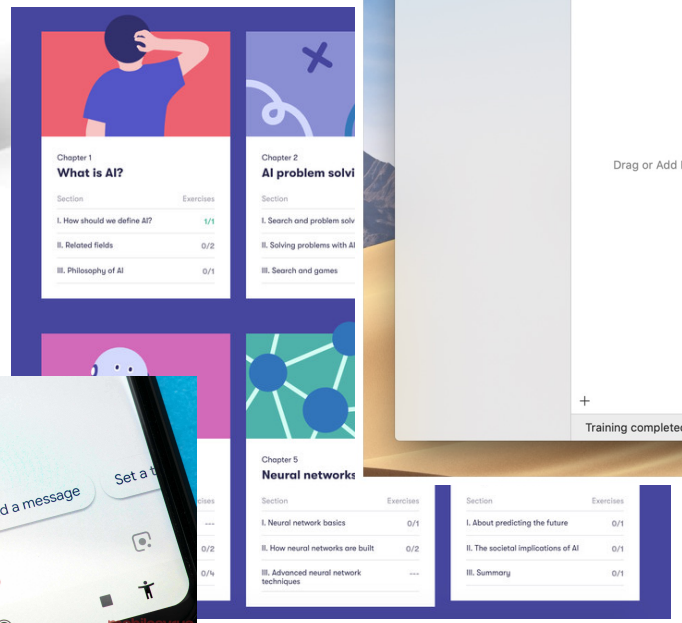
Human Computer Interaction

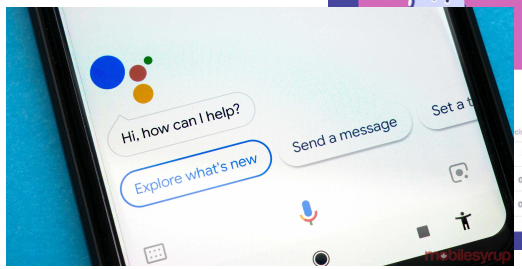
Luigi De Russis, Fulvio Corno

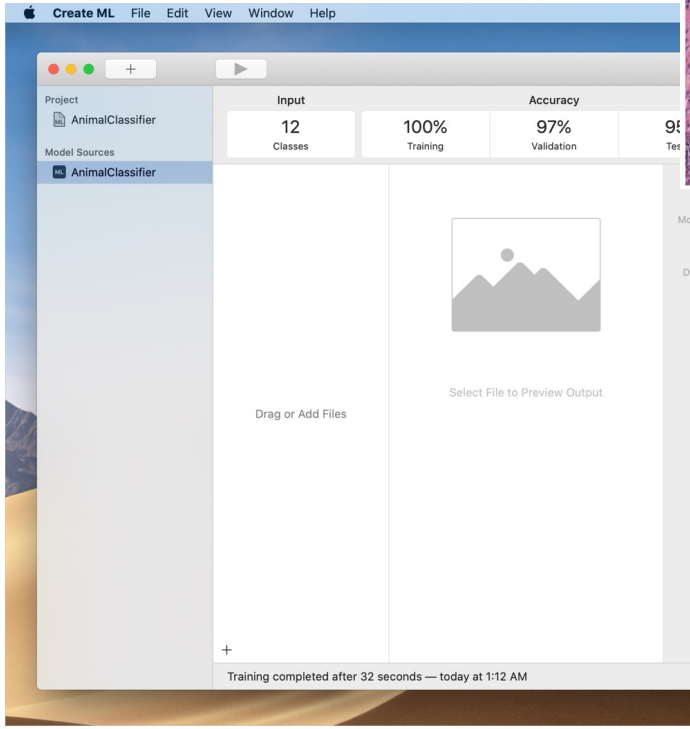
Academic Year 2021/2022

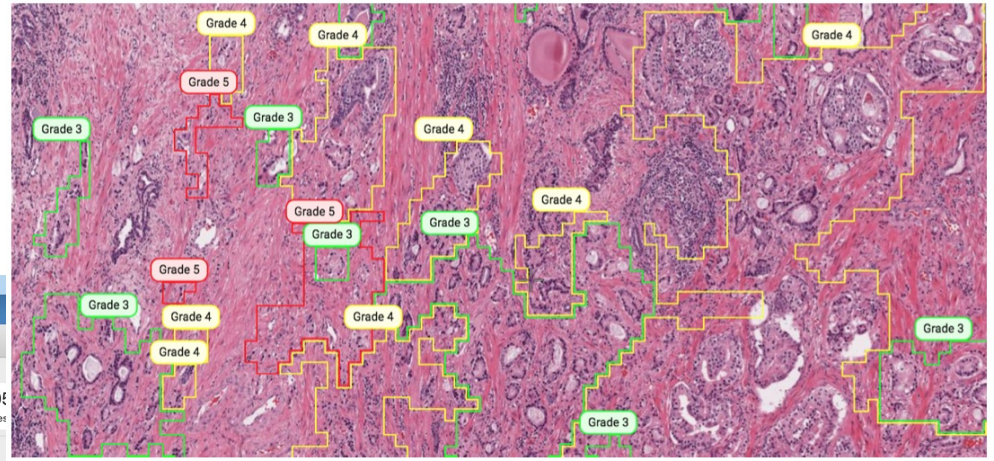
AI is everywhere!











AI is everywhere!

- When it "works", it's *great*!
- When it "fails", it does it *spectacularly*…
  - Tesla Smart Summon, https://www.youtube.com/watch?v=VbVoTK-lM00
  - Alexa, https://www.youtube.com/watch?v=QFpUN3kYTDA

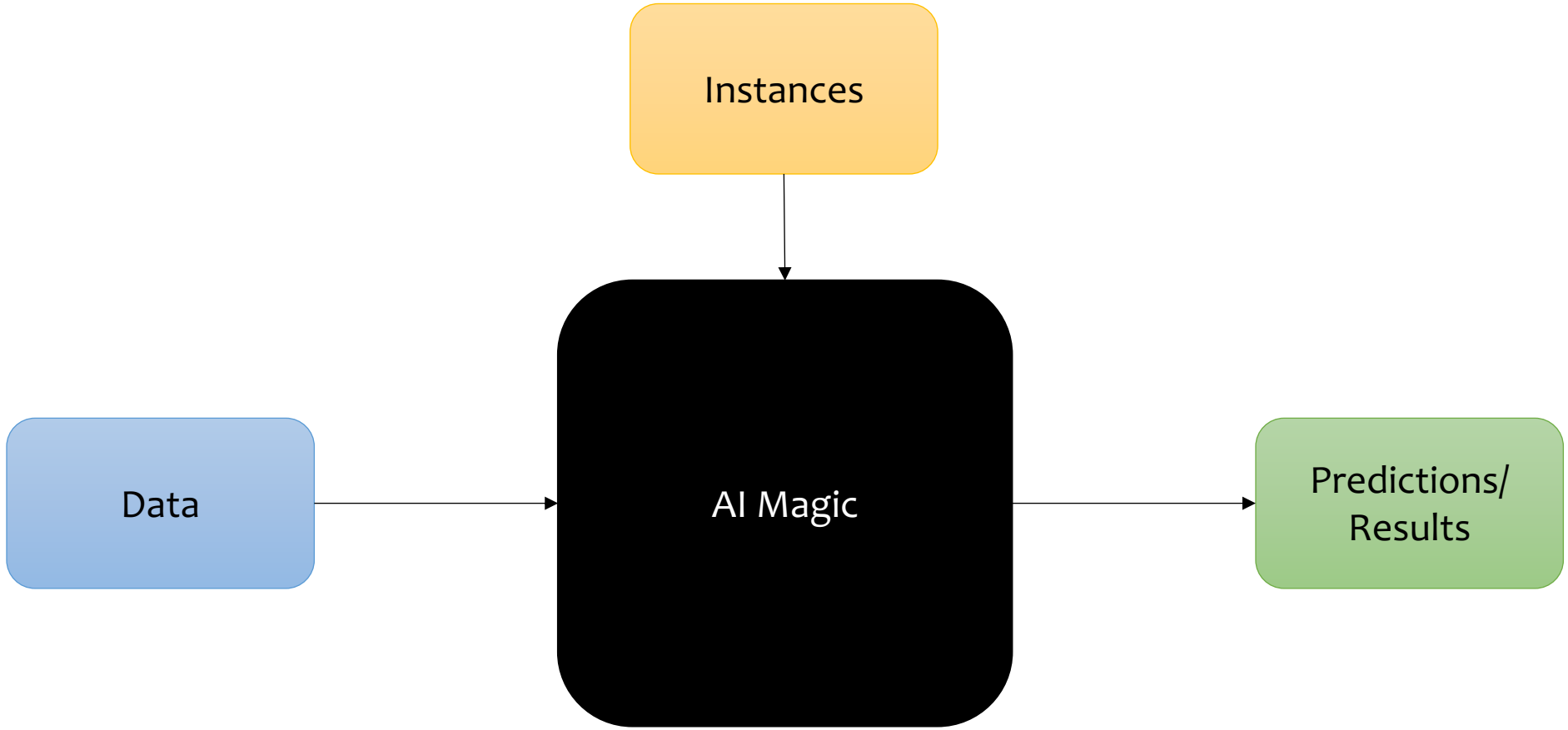
AI is everywhere!

- … and/or it is ***very*** problematic, e.g.,
  - *"IBM boasted that its AI could 'outthink cancer.' Others say computer systems that read X-rays will make radiologists obsolete…"*
  - *"Systems developed in one hospital often flop when deployed in a different facility. Software used in the care of millions of Americans has been shown to discriminate against minorities. And AI systems sometimes learn to make predictions based on factors that have less to do with disease than the brand of MRI machine used, the time a blood test is taken or whether a patient was visited by a chaplain."*

  [source: https://www.scientificamerican.com/article/artificial-intelligence-is-rushing-into-patient-care-and-could-raise-risks/]

- Why?

A Possible Reason: The Typical Approach



Algorithms As The (Main) Answer?

- Algorithms are **not always** the "answer"
  - for instance: if you go to Netflix for the first time, what should it recommend you watch?
  - this is the *cold start problem*, and it is not really and fully solved
    - algorithmically speaking, at least

⇒ A **suitable** user interface is **critical** to overcome some limitations

  - Keeping people involved and considering them since the beginning is fundamental

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

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: high stakes

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

User: low stakes

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

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

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)

How Can We Design Interactive AI Systems?

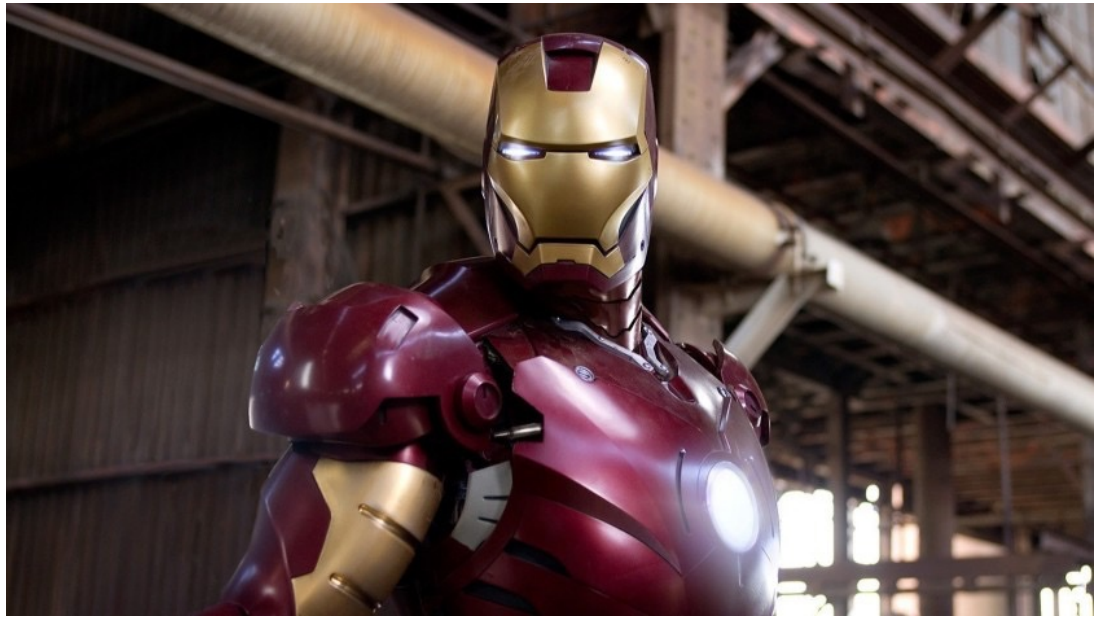
- *"Both [AI and HCI] explore the nexus of computing and intelligent behavior."*

[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

Automation vs. Augmentation?



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 most valuable part of the core experience is its predictability regardless of context or additional user input.
❑ The core experience requires prediction of future events.	❑ The cost of errors is very high and outweighs the benefits of a small increase in success rate.
❑ Personalization will improve the user experience.	❑ Users, customers, or developers need to understand exactly everything that happens in the code.
❑ User experience requires natural language interactions.	❑ Speed of development and getting to market first is more important than anything else, including the value using AI would provide.
❑ Need to recognize a general class of things that is too large to articulate every case.	❑ People explicitly tell you they don't want a task automated or augmented.
❑ 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.	

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

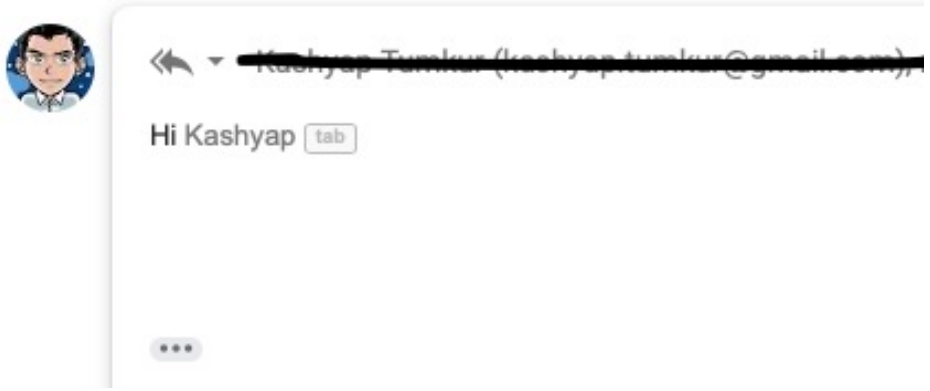
- In general, the more **critical** an app feature is, the more people *need* accurate and reliable results
- On the other hand, if a **complementary** feature delivers results that are not always of the highest quality, people *may* be more forgiving
- Examples
  - Face ID -> critical or complementary?
  - 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



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

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

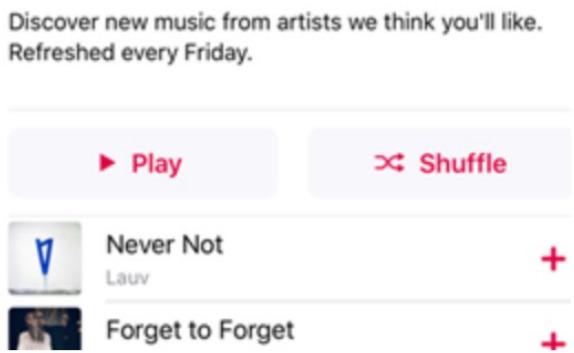
2

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



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

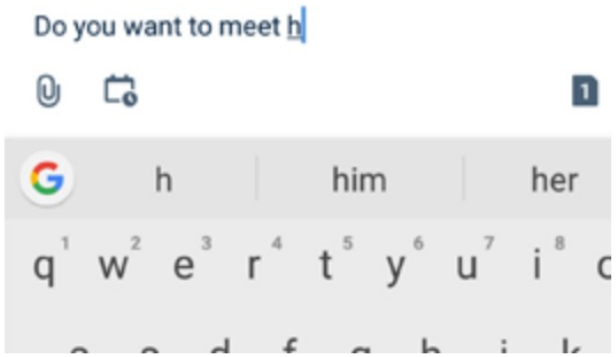
6

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



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

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

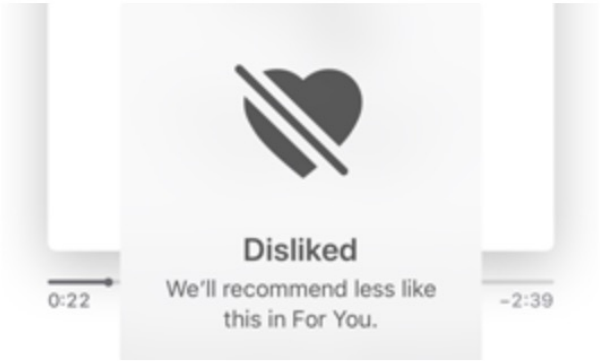
16

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



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/