Guest Lecture

# Human-Centered AI Design: Method, Case Study, and Lessons

**Haiyi Zhu**

Assistant Professor

Carnegie Mellon University

# AI affects many facets of human life & society

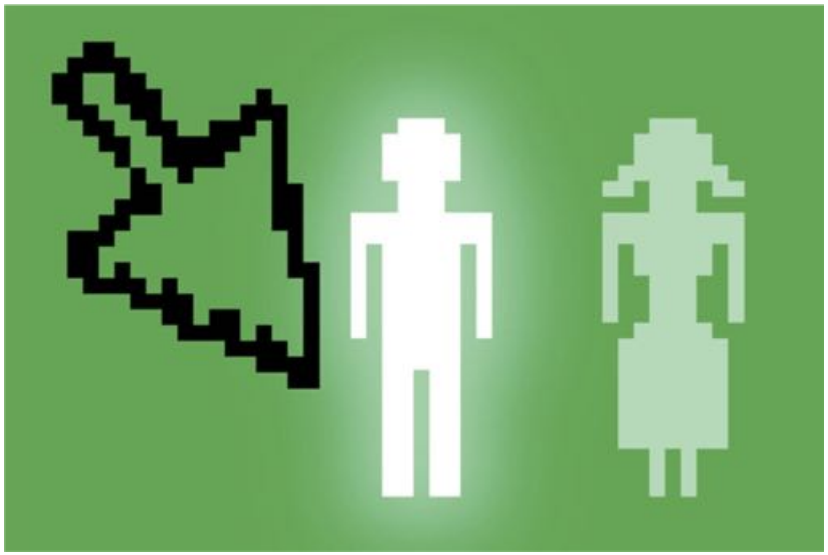
**Twitter taught Microsoft's AI chatbot to be a racist asshole in less than a day**

# AI affects many facets of human life & society

## *Amazon Pushes Facial Recognition to Police. Critics See Surveillance Risk.*

# Unintended biases

**African-American** defendants are substantially more likely than **White-American** defendants to be incorrectly classified as high risk by the algorithms.



Google Ads displayed more high paying jobs for **male users** than for **female users**.

Artificial Intelligence Isn't Killing Jobs; It's Killing Business Models - Forbes

AI Doesn't Eliminate Jobs, It Creates Them - Forbes

AI and robots will destroy fewer jobs than previously feared, says new ...
- The Verge

Will Artificial Intelligence Kill 90% Of Jobs? - Wall Street Survivor Blog

Why AI could destroy more jobs than it creates, and how to save them ...
- TechRepublic

Automation could kill 73 million U.S. jobs by 2030 - USA Today

# My personal experience with learning AI

Spring 2010: Third week of class.

**Theorem.** Let $\mathcal{H}$ be given, and let $d = \mathrm{VC}(\mathcal{H})$. Then with probability at least $1-\delta$, we have that for all $h \in \mathcal{H}$,

$$|\varepsilon(h) - \hat{\varepsilon}(h)| \leq O\left(\sqrt{\frac{d}{m}\log\frac{m}{d} + \frac{1}{m}\log\frac{1}{\delta}}\right).$$

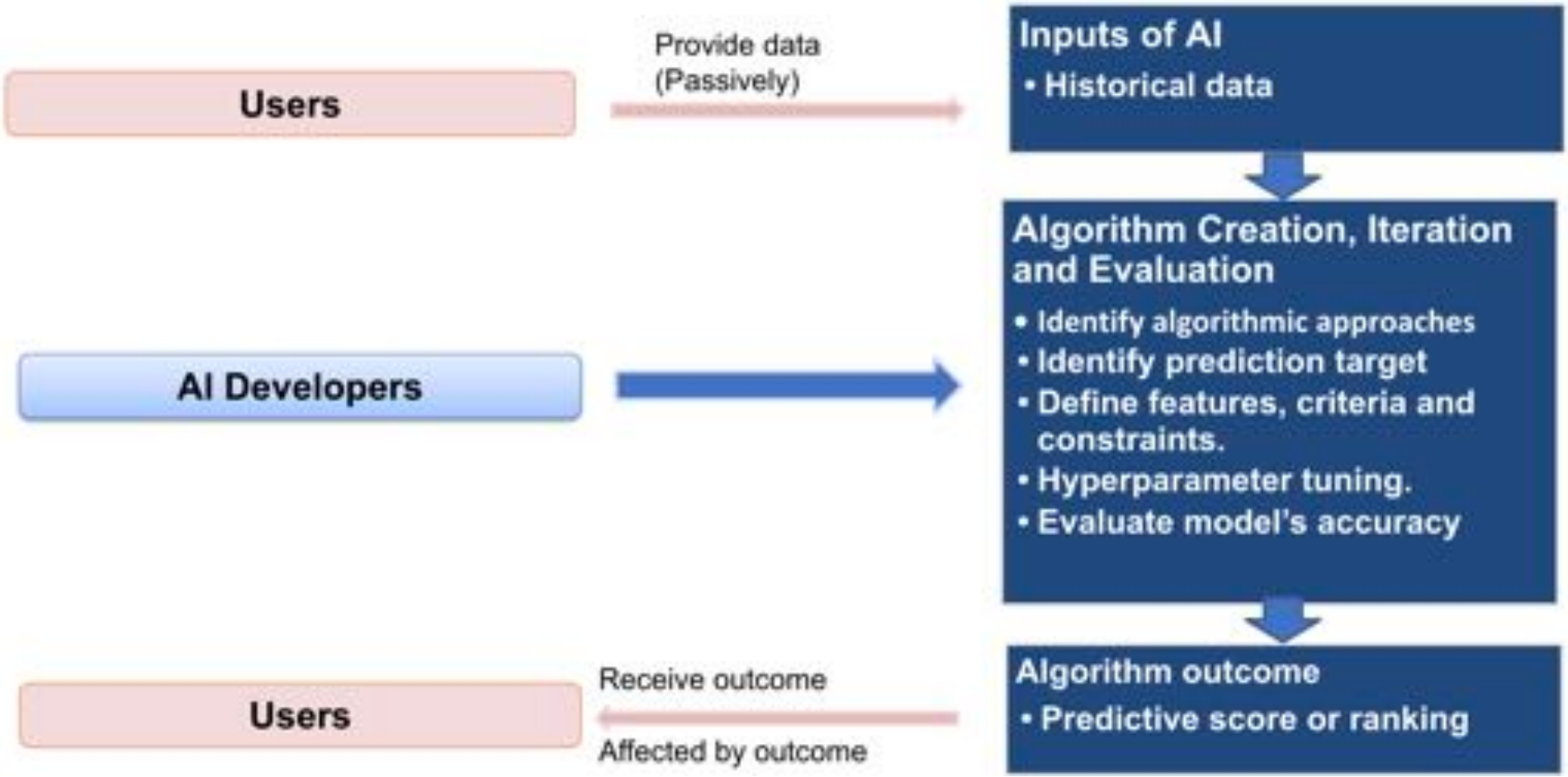
Thus, with probability at least $1-\delta$, we also have that:

$$\varepsilon(\hat{h}) \leq \varepsilon(h^*) + O\left(\sqrt{\frac{d}{m}\log\frac{m}{d} + \frac{1}{m}\log\frac{1}{\delta}}\right).$$

# Problems

- Most AI/ML development are disconnected from real-world problems
- Most AI/ML development consider “user-interfaces” or human impact as an afterthought; and focus narrowly on algorithms

# Traditional development pipeline

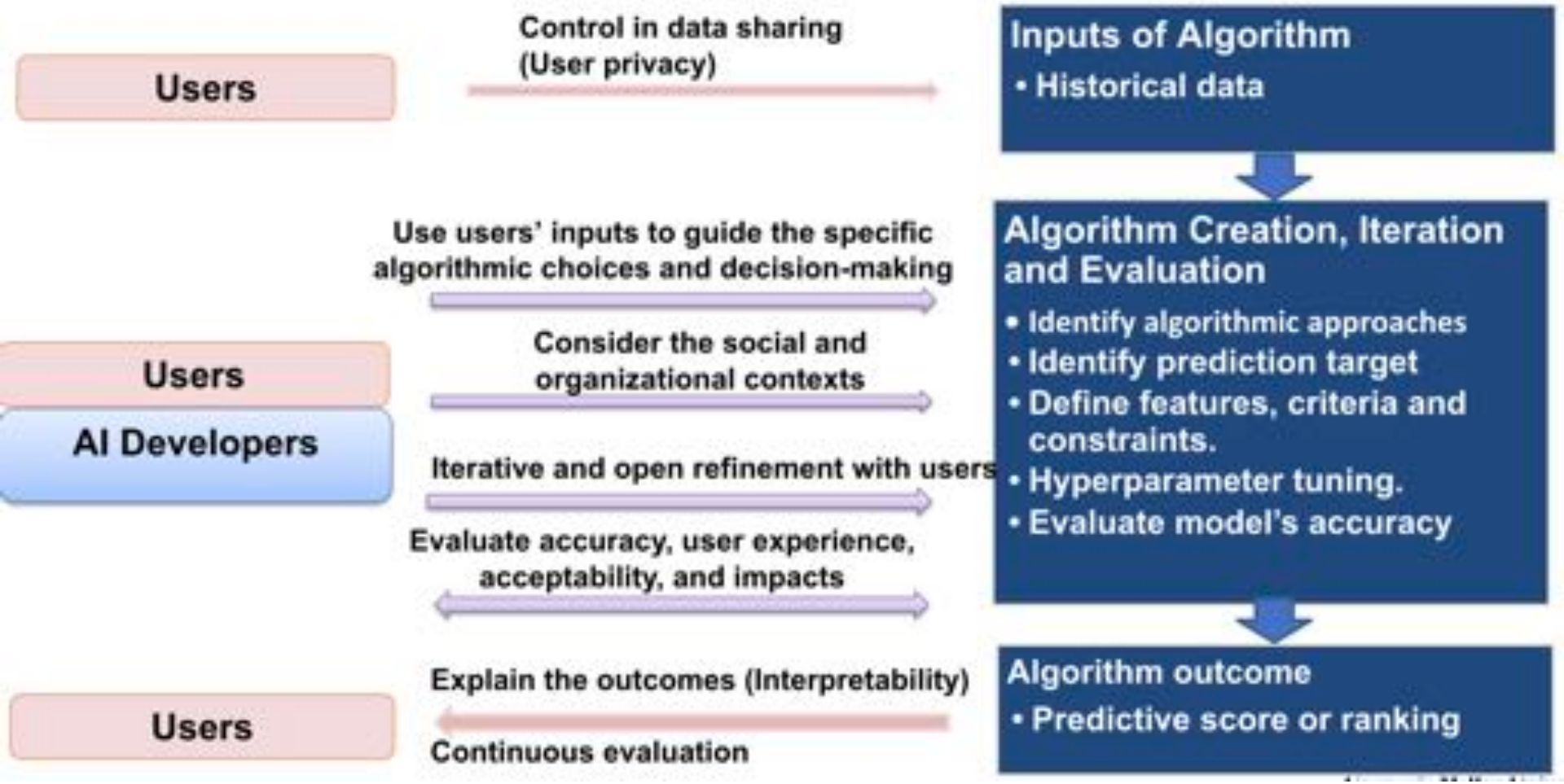



Haiyi Zhu, From Discovery to Design

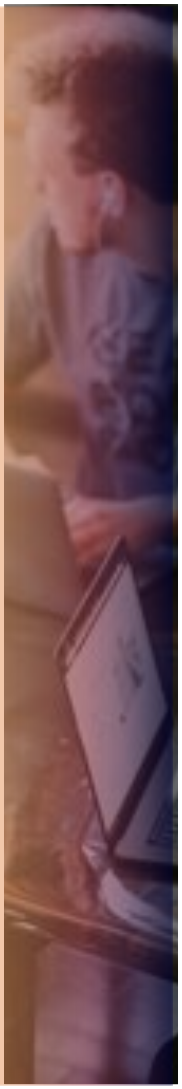
# Human-Centered Approach

- Consider users throughout the process of AI design and development in a principled and comprehensive manner.
- Incorporates users' considerations and constraints into early stages of algorithm design.
- Evaluate not just the model's accuracy (which still is important), but also acceptability, user experience, and the societal impact of the AI system

# Human-Centered Approach




Haiyi Zhu, From Discovery to Design



# Today's Talk

**1 Introduction**

**2 Case Study: Create AI to Support Recruitment in Massive-Scale Online Production**

**3 Ongoing Work**

**4 Recommender System Discussion**

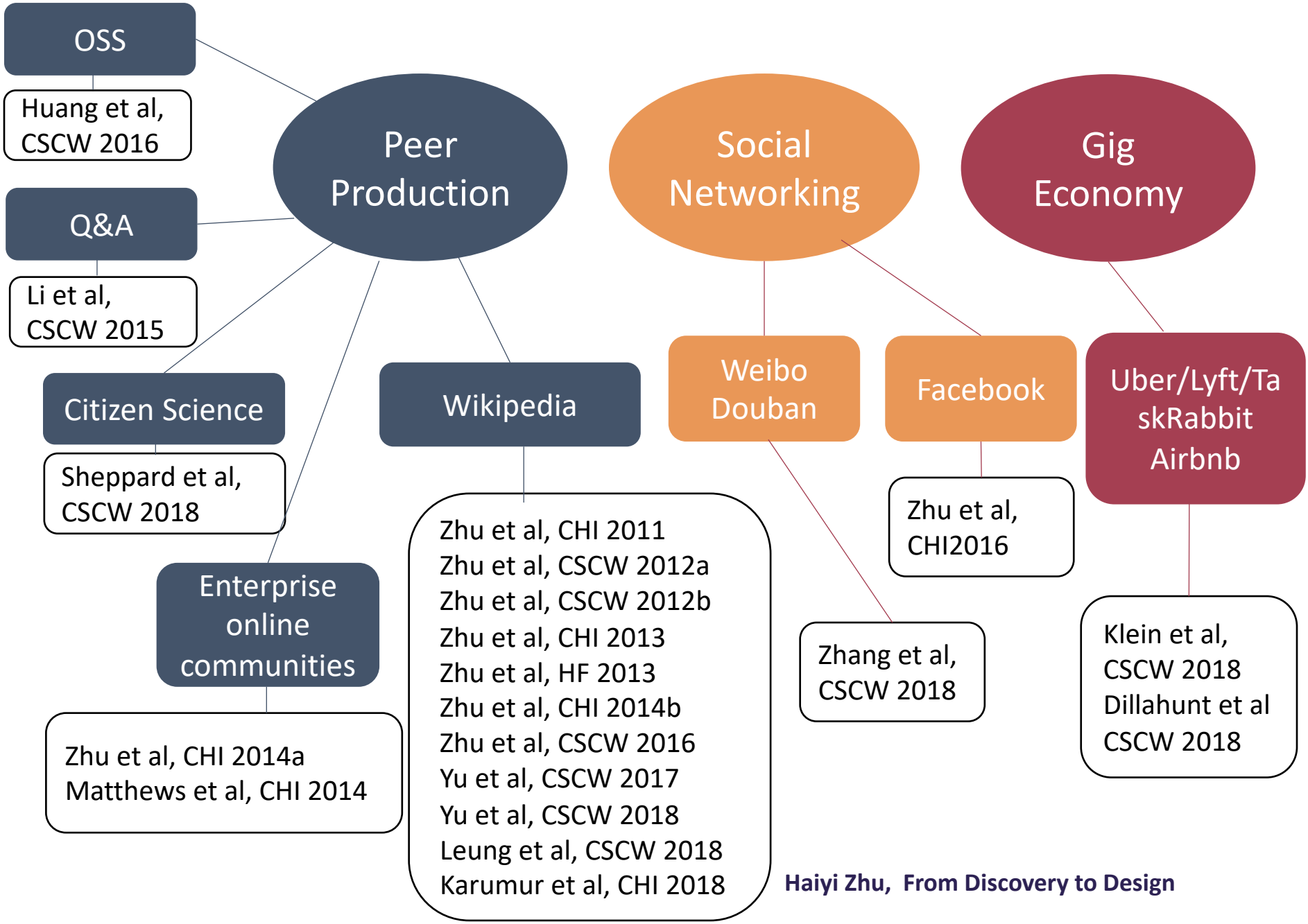
# My Research Trajectory

## Discovery

Generate new understandings of massive-scale collaboration and coordination on these platforms

## Design

Create novel AI technologies or improve existing AI technologies to support these platforms


OSS
Huang et al, CSCW 2016
Q&A
Li et al, CSCW 2015
Citizen Science
Sheppard et al, CSCW 2018
Enterprise online communities
Zhu et al, CHI 2014a
Matthews et al, CHI 2014
Peer Production
Wikipedia
Zhu et al, CHI 2011
Zhu et al, CSCW 2012a
Zhu et al, CSCW 2012b
Zhu et al, CHI 2013
Zhu et al, HF 2013
Zhu et al, CHI 2014b
Zhu et al, CSCW 2016
Yu et al, CSCW 2017
Yu et al, CSCW 2018
Leung et al, CSCW 2018
Karumur et al, CHI 2018
Social Networking
Weibo Douban
Zhang et al, CSCW 2018
Facebook
Zhu et al, CHI2016
Gig Economy
Uber/Lyft/TaskRabbit Airbnb
Klein et al, CSCW 2018
Dillahunt et al CSCW 2018

# Wikipedia: The Largest Collaborative Project in Human History

# Managing Volunteers in Wikipedia

# Functions of Groups in Online Production (Zhu, Kraut & Kittur, CSCW 2012)

# Understanding Member Turnover in WikiProjects (Yu, Wang, Lin, Ren, Terveen & Zhu, CSCW 2018)

- Studied 1,054 WikiProjects over 14 years
- WikiProjects are in general subject to high turnover
- The positive effects of one newcomer joining were larger than the negative effects of one old-timer leaving
- Decline due to shortage of newcomers

# Identify Suitable Newcomers
(Yu, Ren, Terveen & Zhu, CSCW 2017)

- Studied 79,704 project members
- Used pre-joining connections to predict their post-joining productivity and retention
- Two types of pre-joining connections
  - Interest match
  - Personal connections

# Takeaways

- We know WikiProjects (groups within Wikipedia) are important.
- We know WikiProjects need newcomers.
- We know who will thrive.

# Scale of the Problem

As of July 2017, English Wikipedia alone has

- ~ **500** active WikiProjects covering ~ **3.6 million** articles
- ~ **2.9 million** editors with more than 10 edits
- ~ **38,000** new editors registering on a monthly basis

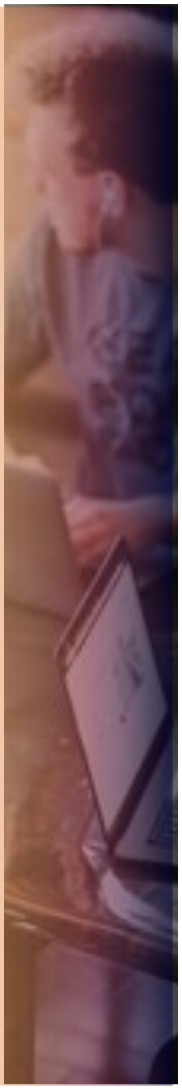
Can we design an Intelligent System to Help WikiProjects Automatically Identify and Recruit Suitable New Members?

## Discovery

**Understand large online platforms**

## Design

**Use the findings to inform the design of AI tools**



# Today's Talk

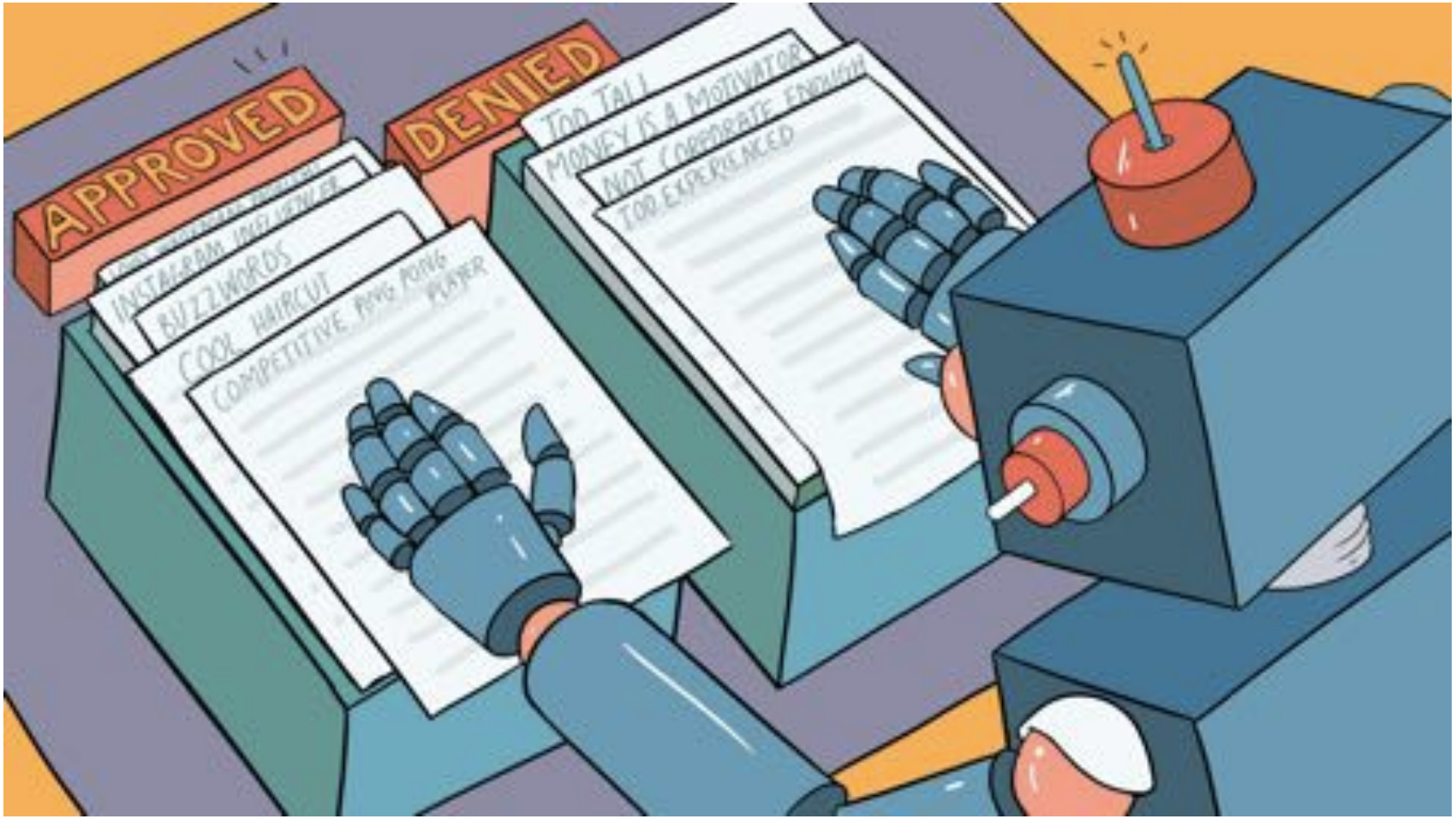
**1 Introduction**

**2 Case Study: Create AI to Support Recruitment in Massive-Scale Online Production**

**3 Ongoing Work**

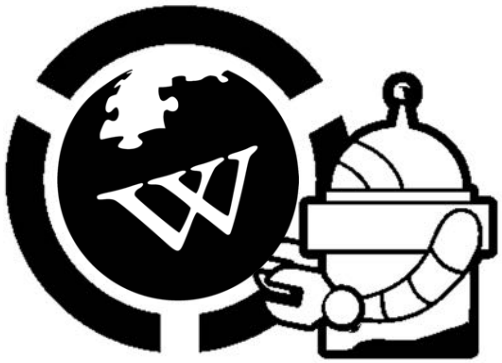
**4 Recommender System Discussion**

# Recruitment Algorithm




https://news.crunchbase.com/news/recruitment-startups-realize-with-the-help-of-ai-less-is-more/

# Challenges: Community Acceptance to AI technologies



AI tools failed because they are insensitive to contributors' motivations and community values (Halfaker et al. 2012)

# Challenges: Community's Low Acceptance of "Intervention" Research

# Challenges: Community's Low Acceptance of "Intervention" Research

3 Community

- 3.1 Wikipedia is not an anarchy or forum for free speech
- 3.2 Wikipedia is not a democracy
- 3.3 Wikipedia is not a bureaucracy
- 3.4 Wikipedia is not a laboratory
- 3.5 Wikipedia is not a battleground
- 3.6 Wikipedia is not compulsory

# Challenges: Community's Low Acceptance of "Intervention" Research

"If this proposal was an earnest effort to improve Wikipedia, the researchers would have worked with the community from the start to design a study that was **consistent with our needs and values**."

- Wikipedia Editor: dlthewave

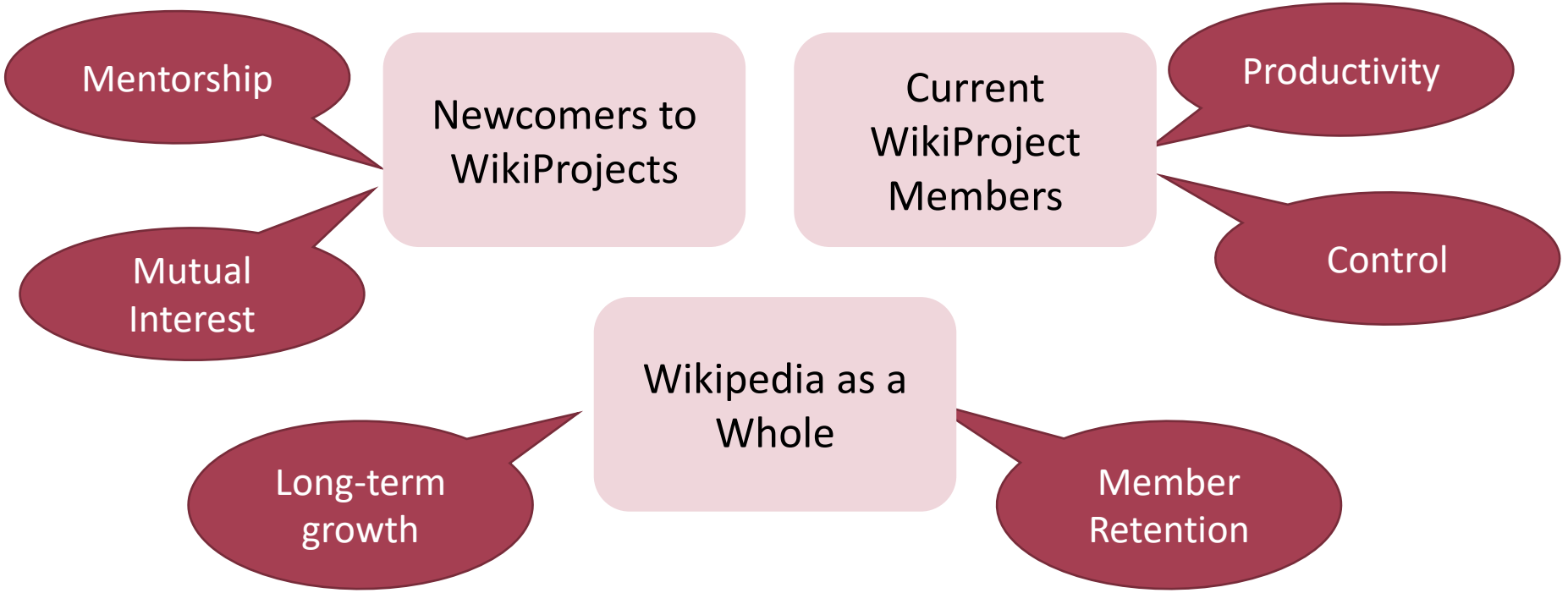
# Challenges: Multiple Stakeholders

Newcomers to WikiProjects

Current WikiProject Members

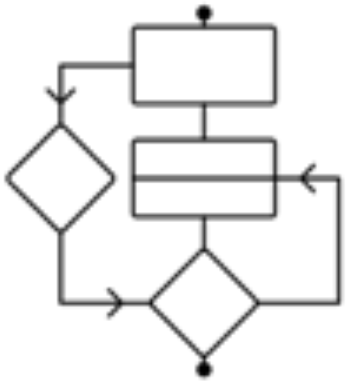
Wikipedia as a Whole

# Challenges: Different Needs and Values

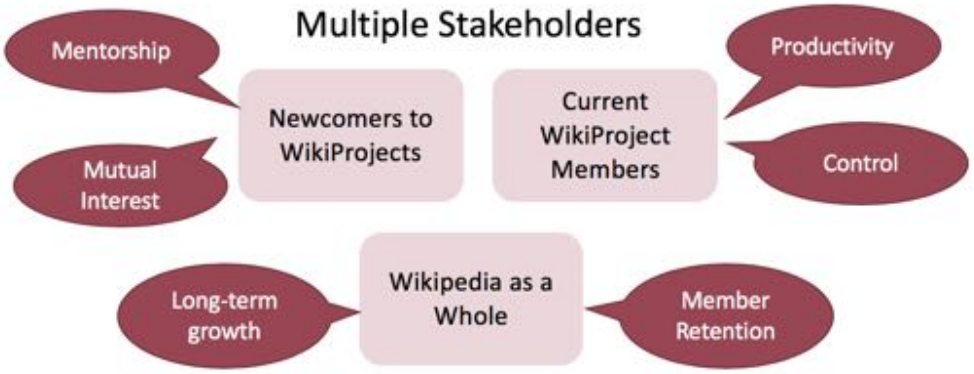


How can we design intelligent systems that are aligned with multiple stakeholders' needs and values?

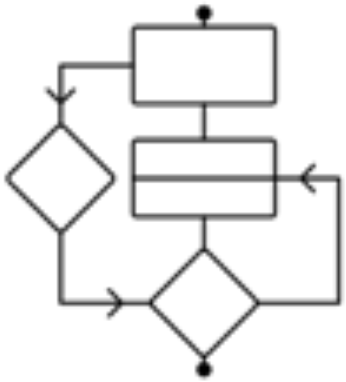
# Traditional Algorithm Design

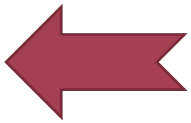


- Complicated Multi-Stakeholder Context
- Complex Human Needs and Values

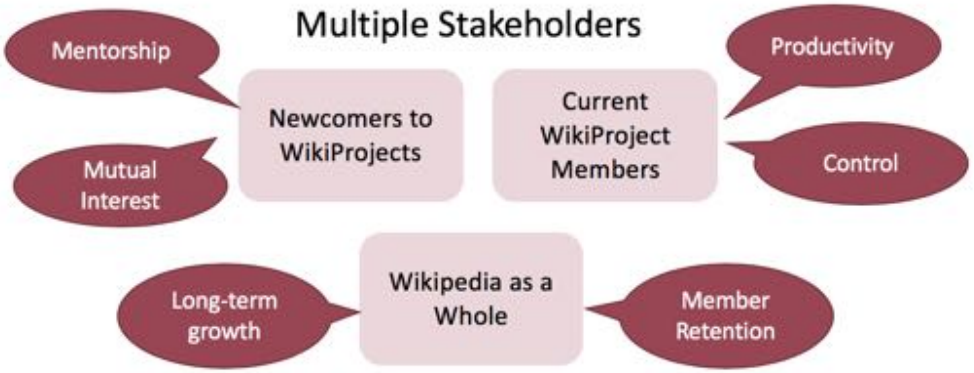


# Our Approach

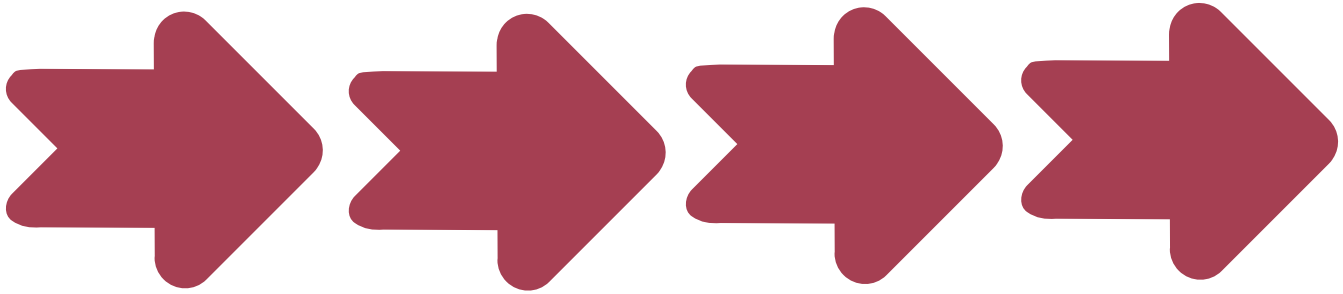




- Complicated Multi-Stakeholder Context
- Complex Human Needs and Values



# Our Approach: Value Sensitive Algorithm Design (Zhu, Yu, Halfaker & Terveen, CSCW 2018)



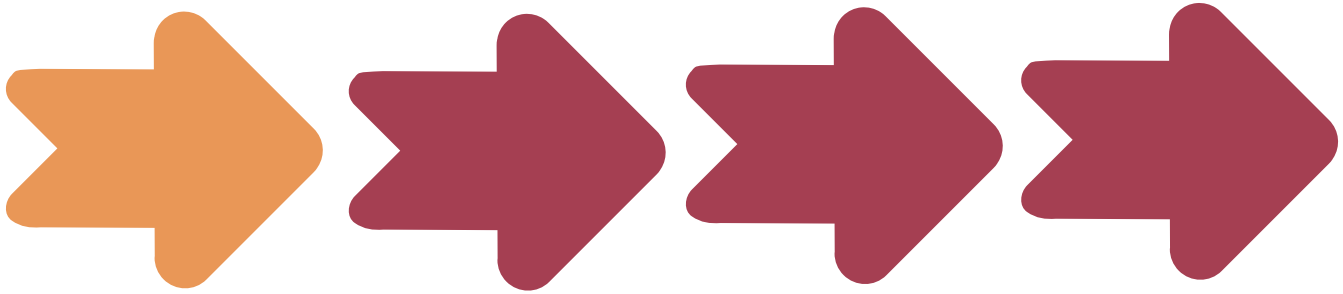
**Step 1**
Understand Stakeholder Values

**Step 2**
Design Initial Algorithm

**Step 3**
Pilot deploy with stakeholders

**Step 4**
Evaluate and Iteratively Refine

# Our Approach: Value Sensitive Algorithm Design
(Zhu, Yu, Halfaker & Terveen, CSCW 2018)



**Step 1**
Understand Stakeholder Values

**Step 2**
Design Initial Algorithm

**Step 3**
Pilot deploy with stakeholders

**Step 4**
Evaluate and Iteratively Refine

# Step 1. Understand Stakeholder Values

Newcomers to WikiProjects

Current WikiProject Members

Wikipedia as a Whole

**Define "values" broadly as "what a person or group of people consider important in life"**
(Borning and Muller 2012).

# Algorithm Design

| | |
|---|---|
| **Data preparation** | **“Who to include?”** |
| **Algorithmic Approach** | **“How to determine the fit?”** |
| **Output Presentation** | **“How to communicate the algorithms’ output”** |

# Stakeholder Values and Trade-offs

| | Major Choices | Newcomers to WikiProject | Current WikiProject members | Wikipedia as a whole |
|---|---|---|---|---|
| **Data preparation** | Experienced editors | ↑ Collaboration | ↑ Productivity | ↑ Member Retention |
| | Brand-new editors | ↑ Mentorship | ↓ Productivity | ↑ New member retention |
| **Algorithmic Approach** | Interest-based | ↑ Mutual Interest | ↑ Mutual Interest | |
| | Relationship-based | ↑ Personal connection | ↓ Productivity | |
| **Output Presentation** | Direct invite | | ↓ Control | |
| | Communicate with current members | | ↑ Control | |

# Value Sensitive Algorithm Design: Four-Step Approach

**Step 1**
Understand Stakeholder Values

**Step 2**
Design Initial Algorithm

**Step 3**
Pilot deploy with stakeholders

**Step 4**
Evaluate and Iteratively Refine

# Data Preparation - “Who to Include?”: Experienced or Brand-New Editors?

Current Project Members

Recruit Brand-New Editors

Wikipedia as a Whole

Design decision:

- Include both **brand-new** and **experienced** editors.
- Rank the two types of editors separately.

# Algorithmic Approach – "How to determine the fit": Interest-based or Relationship-based?

Current Project Members

Relationship-based

Newcomers

Design decision:

- Implement both interest-based and relationship-based algorithms.

# Algorithms

Four Different Ranking Algorithms

- Interest-based algorithms:
  - Rule-based algorithm
  - Category-based algorithm
- Relationship-based algorithms:
  - Bond-based algorithm
  - Co-edit-based algorithm

# "How to communicate the algorithms' outputs": Directly invite newcomers Or Communicate with current members first?

Directly invite newcomers

Current project members

Design decision:

- Create a user interface for presenting the algorithm outputs to current project members
- Present top recommendations from all four algorithms, separating experienced editors and brand-new editors.

# Interface

| Username | Why we recommend this editor | First Edit Date | Total Edits in ENWP | Editor Status | Invite | Survey |
|---|---|---|---|---|---|---|
| Hulk | Hulk made 77 out of their total 187 edits to articles within the scope of your project. | 2017-8-24 | 187 | Experienced Editor | invite | survey |
| Wolverine | Wolverine made 49 out of their total 87 edits to articles within the scope of your project. | 2017-7-23 | 87 | Newcomer | invite | survey |
| Iron Man | Iron Man edited articles similar to articles your project members edited. For example, Iron Man and you project member Superman edited 9 same articles in their most recent 200 edits. | 2017-7-26 | 191 | Experienced Editor | invite | survey |
| Thor | Thor edited articles similar to articles your project members edited. For example, Thor and you project member Captain America edited 8 same articles in their most recent 200 edits. | 2017-5-19 | 52 | Newcomer | invite | survey |

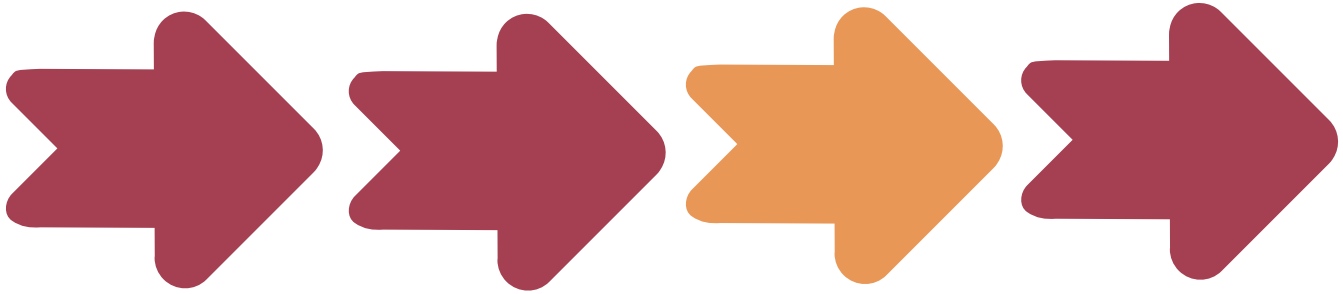
Recommended newcomer

Explanation

Invite link

Survey link

# Value Sensitive Algorithm Design: Four-Step Approach



**Step 1**
Understand Stakeholder Values

**Step 2**
Design Initial Algorithm

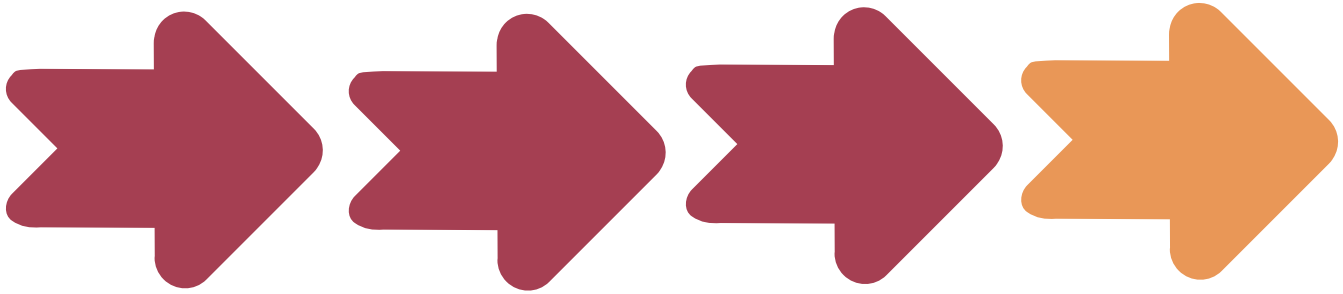
**Step 3**
Pilot deploy with stakeholders

**Step 4**
Evaluate and Iteratively Refine

# Deploy the Recruitment Tool

- Worked with the Wikimedia Foundation and with WikiProjects organizers to deploy our recruitment system
- Over a **six-month** period, we evaluated over **16,000** editors, and delivered **4** distinct batches of **385** recommendations to **18** WikiProjects.

# Value Sensitive Algorithm Design: Four-Step Approach



**Step 1**
Understand Stakeholder Values

**Step 2**
Design Initial Algorithm

**Step 3**
Pilot deploy with stakeholders

**Step 4**
Evaluate and Iteratively Refine

# Evaluation

**Algorithm Accuracy**

**Stakeholder Acceptance**

**Impacts on the Stakeholders**

# Accuracy Evaluation

**Algorithm Types**

| | Rule-based | Category-based | Bonds-based | Co-edit-based |
|---|---|---|---|---|
| Avg. Ratings (5 point scale) | 3.24 | 2.36 | 2.33 | 2.76 |
| Invitation Rates (Clickthrough) | 47% | 16% | 22% | 28% |

# Accuracy Evaluation

**Newcomer Types**

| | Experienced | Brand-new |
|---|---|---|
| Avg. Ratings | 2.85 | 2.88 |
| Invitation Rates | 34% | 32% |

# Evaluation

**Algorithm Accuracy** 

**Stakeholder Acceptance**

**Impacts on the Stakeholders**

# Acceptance Evaluation

**Quote from a current WP member**

***"This puts some science behind recommendations, and will be a great supplement to the current processes."***

# Acceptance Evaluation

**Message from a new member who was invited to join the WP.**

***"Thank you for reaching out to me and thank you for informing me about the WikiProject Africa … I appreciate it."***

# Acceptance Evaluation

**Feedback from the general Wikipedia community:**

Created a Signpost (Wikipedia's internal blog) to describe our project

# DISCUSS THIS STORY

+ Add a comment

THESE COMMENTS ARE AUTOMATICALLY TRANSCLUDED FROM THIS ARTICLE'S TALK PAGE. TO FOLLOW COMMENTS, ADD THE PAGE TO YOUR WATCHLIST. IF YOUR COMMENT HAS NOT APPEARED HERE, YOU CAN TRY PURGING THE CACHE.

- Very interesting. I would like to be notified when t… …er languages. -- [redacted] chat (talk) 08:52, 24 November 2017 (UTC)
- Amazing creation Bobo.03! It definitely is good in an age where 95% of valuable newbies get discouraged by veteran editors... J… …mber 2017 (UTC)
- The creator of this tool used it for Wi… …alth and it works extremely well. It also seems to identify a project's most active contributors. This is a good way to determine who are the editors that probably deserve a barnstar or two. Best Regards, [redacted] 15:25, 25 November 2017 (UTC)
- I'm really excited by the potential for routing editors towards active, subject-focused working groups. Sometimes W… …a can feel like a… …e. I think this project has the potential t… …to their effort. I'm really excited by the idea tha… …some struggling/inactive WikiProjects and h… contribs) 16:26, 25 November 2017 (UTC)

Very interesting.

Amazing creation.

I am really excited by the potential for routing editors towards active, subject focused working groups.

## Wikipedia is Not a Laboratory

**Bumping thread** for 30 days. : **Noyster** (talk), 10:13, 4 December 2017 (UTC)

Some of you have probably already seen this:

https://en.wikipedia.org/wiki/Wikipedia:Administrators%27_noticeboard/Incidentarchive965#Single_purpose_account_for_mass_adding_articles_by_a_num

I think that, although outrage is justified, much of the an
what is wrong with this sort of experiment. The real prob
laboratory, as a guinea pig, as an experiment. The purp
try to generate new knowledge by experimentation. We need a paragraph in this policy to that effect, which should conclude that editors who are using Wikipedia in any way for experimentation (as opposed to the summarizing of existing knowledge) may be banned. Some forms of experimental behavior, such as a breaching experiment in vandalism, are already forbidden, but other types of experiments are also not proper uses of Wikipedia or of the privilege of editing Wikipedia. I will draft a paragraph, but would appreciate any input in the next 24 to 72 hours.
02:39, 5 October 2017 (UTC)

I'd like to mention:

- Read-only experiments are difficult to detect and
probably very common and does not require any
like ORES, etc. Others may have controversial i
knowledge.
- There are a number of experiments done in
(not necessitating controversial
participation to hi

**"Editors who are using Wikipedia in any way for experimentation may be banned"**

**There are a number of experiments done in good faith to increase the quality of the project or the experience of editors, editor retention, etc. An example of a recent one is Bobo.03's WikiProject editor recommendations."**

# Evaluation

**Algorithm Accuracy** 

**Stakeholder Acceptance** 

**Impacts on the Stakeholders**

# Summary of findings

- Algorithms identified newcomers who were more suitable.
- Current members did additional investigation and selected the most promising newcomers.
- Current members provided additional help and mentorship to invited newcomers, which led to more contributions.

# Evaluation

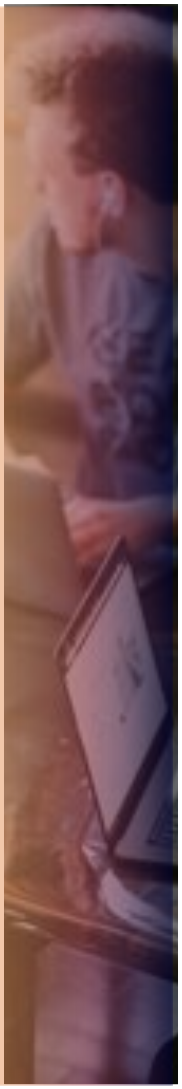
**Algorithm Accuracy** 

**Stakeholder Acceptance** 

**Impacts on the Stakeholders** 

# Contribution of this study
(Zhu, Yu, Halfaker & Terveen, CSCW 2018)

- Generated **knowledge** of stakeholder values regarding intelligent recruitment systems in Wikipedia.
- Created **four algorithms** for identifying suitable newcomers for WikiProjects.
- Designed and deployed a **working intelligent recruitment system** in the field. Evaluation showed that the system is well accepted by the community stakeholders.
- Demonstrated the potential of a **general design process** for incorporating stakeholders' needs and values into the creation of algorithms.



# Today's Talk

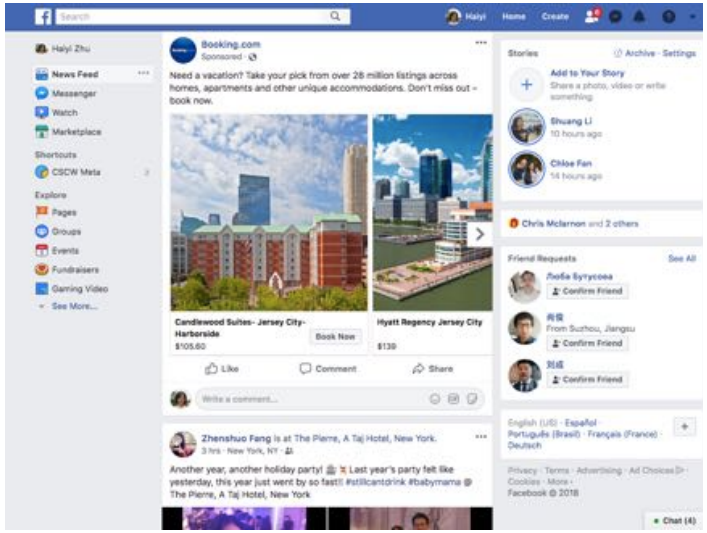
**1 Introduction**

**2 Case Study: Create AI to Support Recruitment in Massive-Scale Online Production**

**3 Ongoing Work**

**4 Recommender System Discussion**

# AI-based Management on Online Platforms






TaskRabbit

# AI-based Management in Society

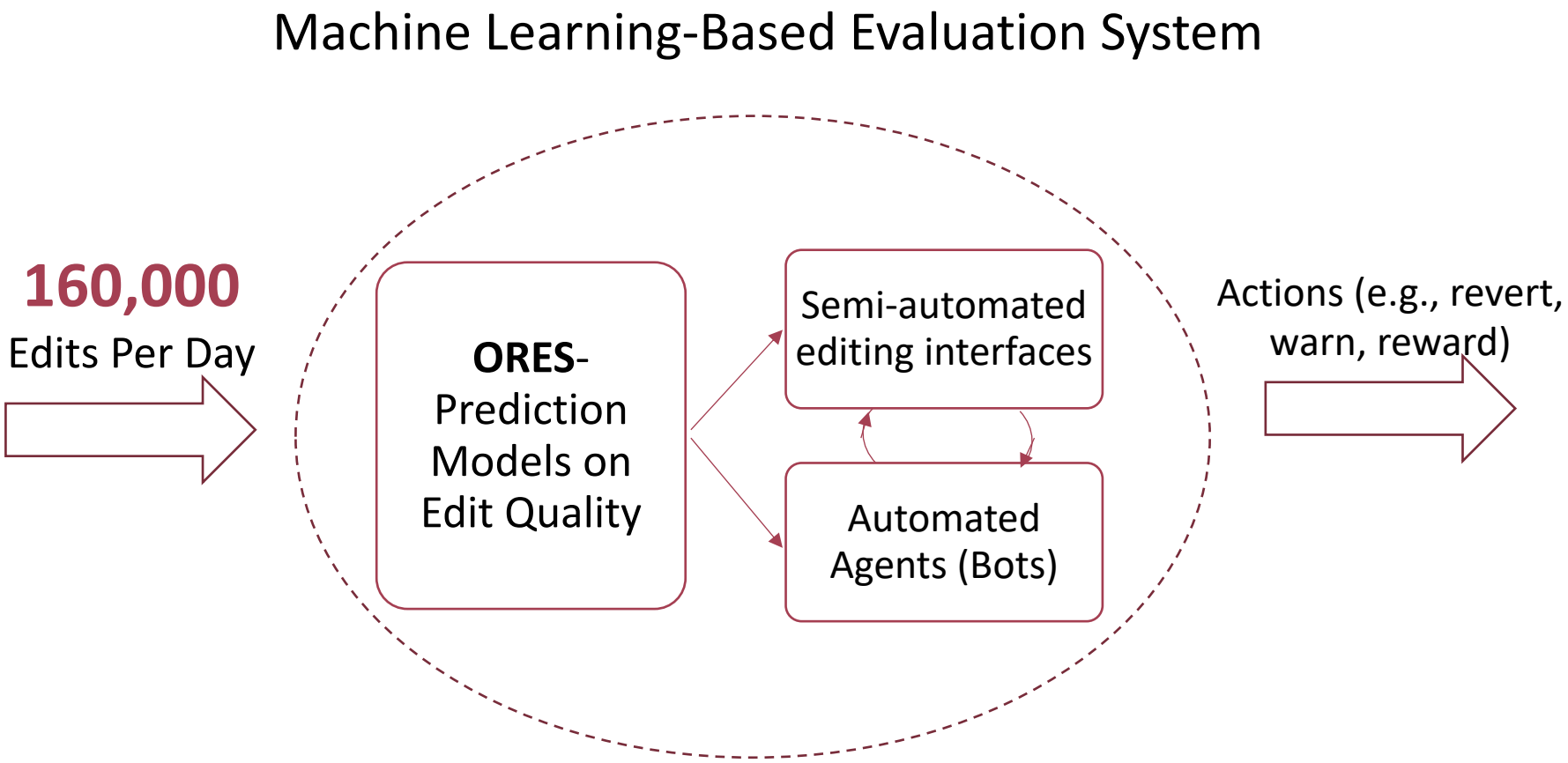
https://www.nsf.gov/news/special_reports/big_ideas/human_tech.jsp

# Ongoing Work: AI-Based Management

- **Improve the Machine Learning-based Work Evaluation System on Wikipedia**

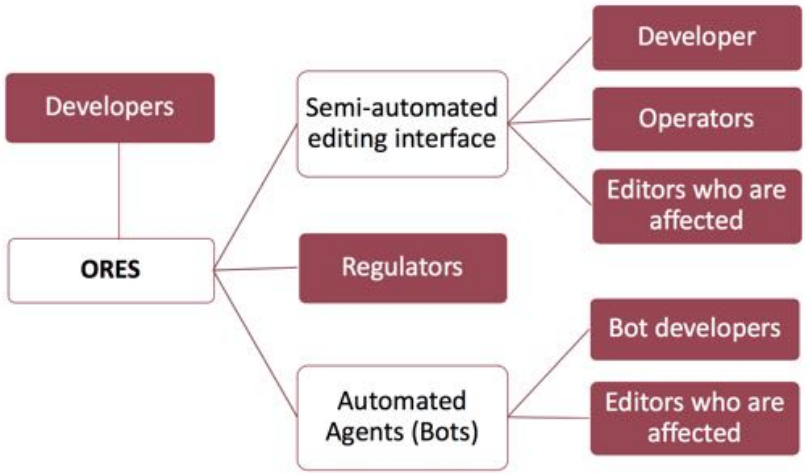
- **Empowering and Enhancing Gig Workers Through Building Intelligent Tools**

# Work Evaluation in Wikipedia

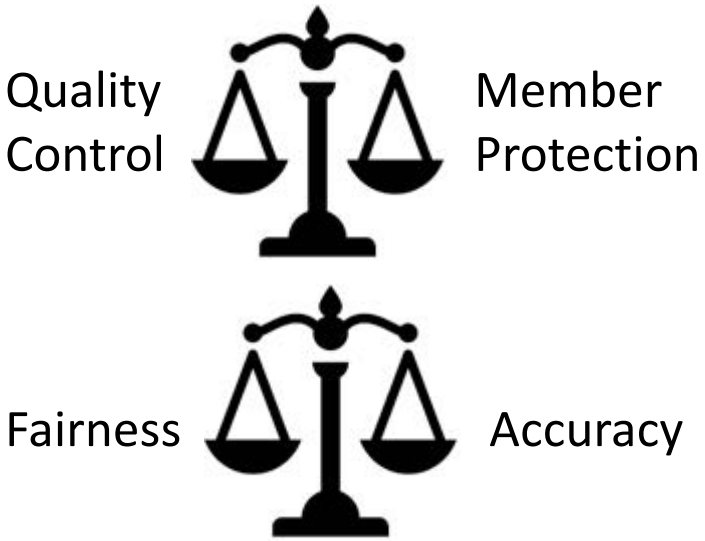


# Challenges of ML-based Evaluation System

Diverse work contexts and multi-stakeholder situations




Tensions between different stakeholder goals



How can we handle the inherent trade-offs between different stakeholder goals in a machine learning-based content moderation?

- Understand stakeholders' goals
- Design and evaluate novel techniques to capture, explain and negotiate the trade-offs between stakeholder goals

*Sponsored by NSF CHS Core Program
*Sponsored by NSF EAGER on AI and Society
*Partially sponsored by Facebook grant on mechanism design for social good
*Partner with Wikimedia Foundation


# Ongoing Work

- **Improve the Machine Learning-based Work Evaluation System on Wikipedia**

- **Empowering and Enhancing Gig Workers Through Building Intelligent Tools**

BI Business Insider

## Upwork finds the pandemic could be turning gig economy white-collar - Business Insider

An Upwork report found that this summer, 24% more people made a recent decision to enter the gig economy than in most years on record.

2 weeks ago

CNN

## The $185 million campaign to keep Uber and Lyft drivers as contractors in California

Sponsored by Uber, Lyft, DoorDash, Instacart and Postmates, California Proposition 22 would exempt drivers for their services from the rules of ...

1 day ago

# Empowering and Enhancing Workers Through Building A Community-Centered Gig Economy

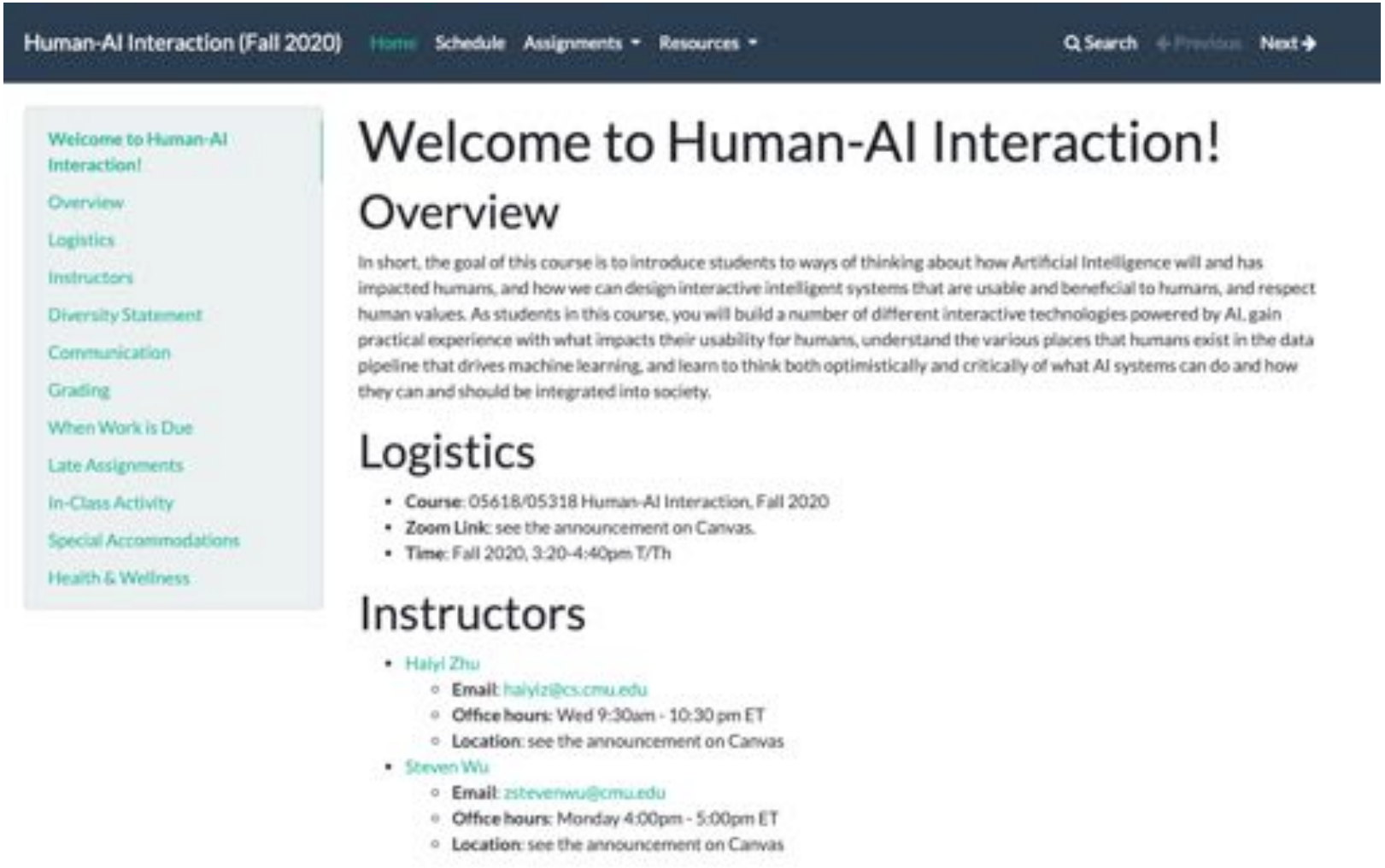
*Sponsored by NSF Smart and Connected Community Program

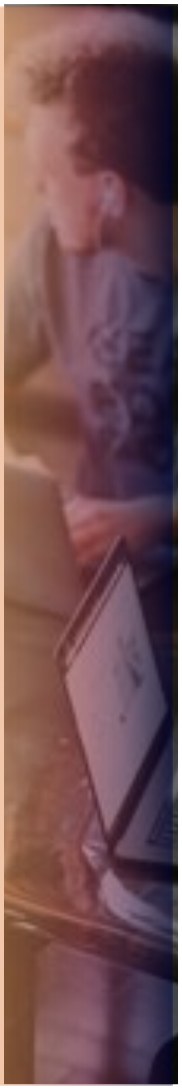
* Four institutions
Carnegie Mellon University - Haiyi Zhu and Steven Wu
University of Minnesota - Gord Burtch and Zhi-Li Zhang
University of Texas at Austin - Min Kyung Lee
Worcester Polytechnic Institute (WPI) - Yanhua Li



# 05618/05318 Human-AI Interaction

https://haiicmu.github.io/





# Today's Talk

**1 Introduction**

**2 Case Study: Create AI to Support Recruitment in Massive-Scale Online Production**

**3 Ongoing Work**

**4 Recommender System Discussion**

# Recommender systems

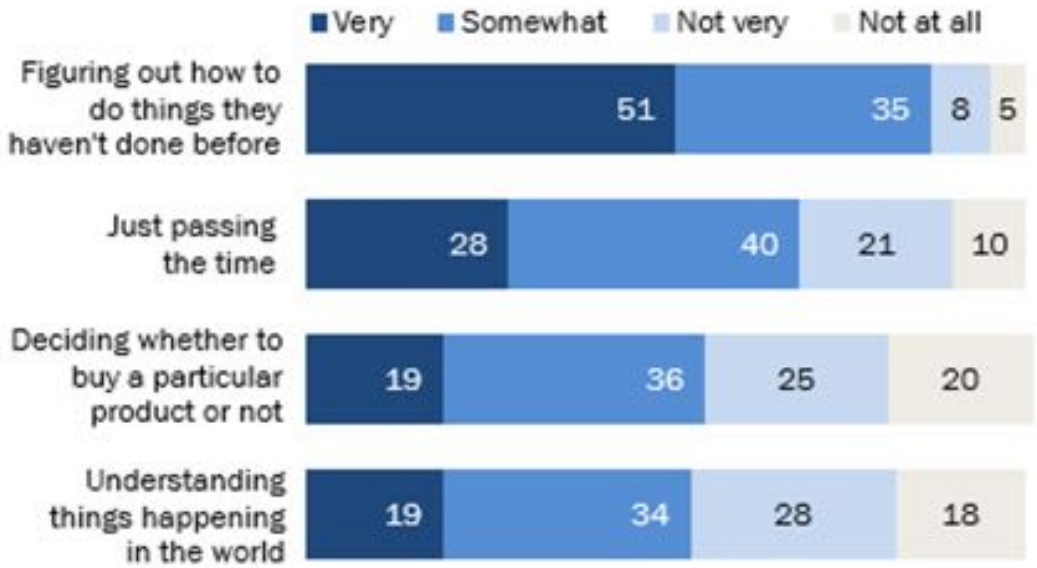
- Provide personalized recommendations sensitive to users' interests
- Aggregate and report other people's opinions and preferences

# Pew Research Report

https://www.pewresearch.org/internet/2018/11/07/many-turn-to-youtube-for-childrens-content-news-how-to-lessons/

## One-in-five YouTube users say it is very important for helping them understand things that are happening in the world

*% of **U.S. adults who use YouTube** who say the site is ___ important when it comes to ...*

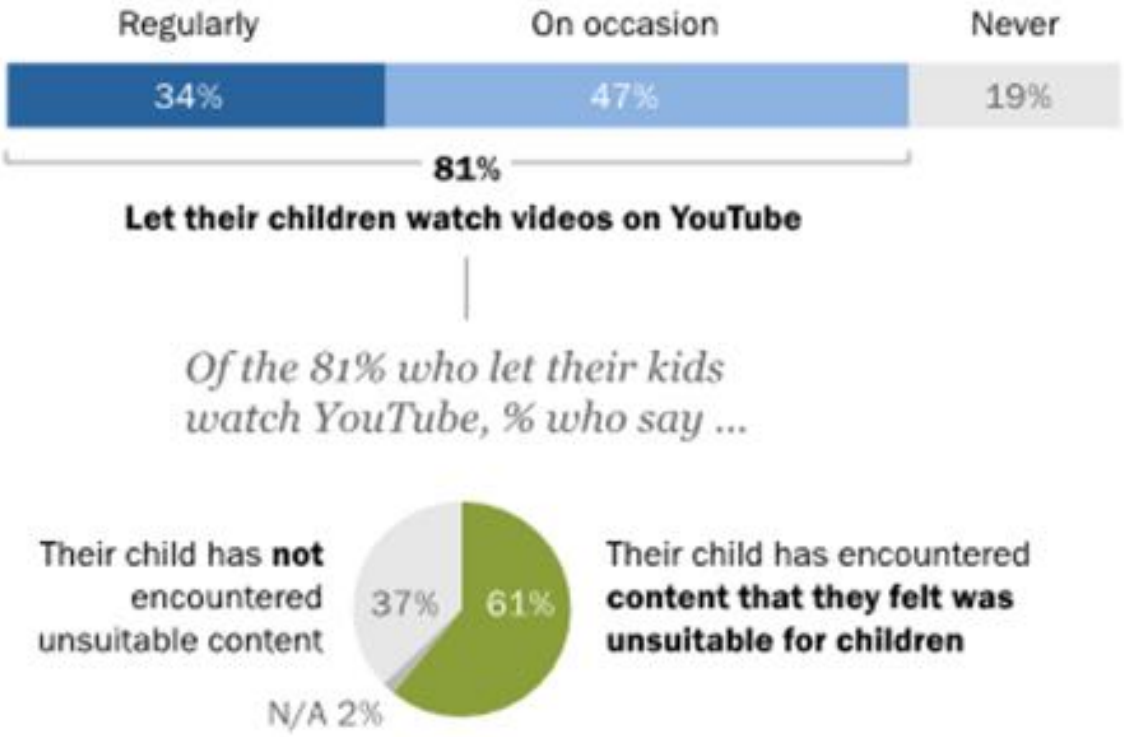



Note: Respondents who did not give an answer are not shown.
Source: Survey of U.S. adults conducted May 29-June 11, 2018.
"Many Turn to YouTube for Children's Content, News, How-To Lessons"

PEW RESEARCH CENTER

## Around one-third of parents of young children regularly let their child watch videos on YouTube

*% of U.S. parents with children age 11 or younger who say they let their child/children watch videos on YouTube ...*

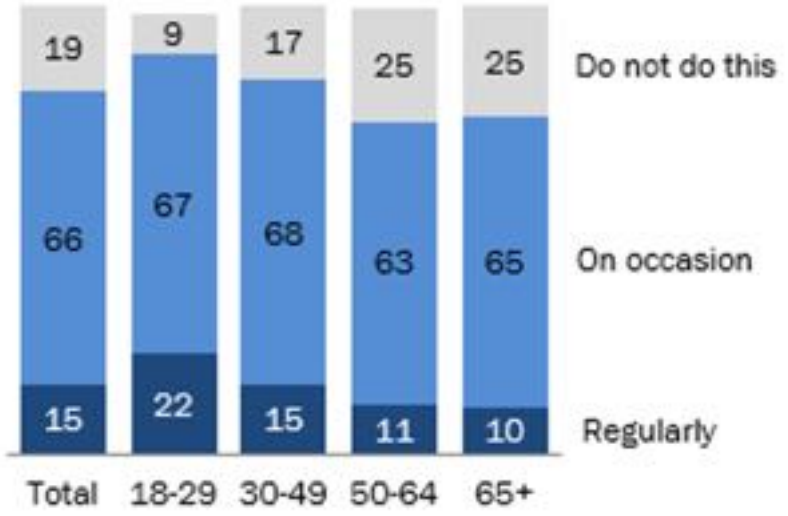



Source: Survey of U.S. adults conducted May 29-June 11, 2018.
"Many Turn to YouTube for Children's Content, News, How-To Lessons"

PEW RESEARCH CENTER

## Majority of YouTube users across a wide range of age groups watch recommended videos

*% of **U.S. adults who use YouTube** who say they watch the recommended videos that appear alongside the video they are currently watching ...*

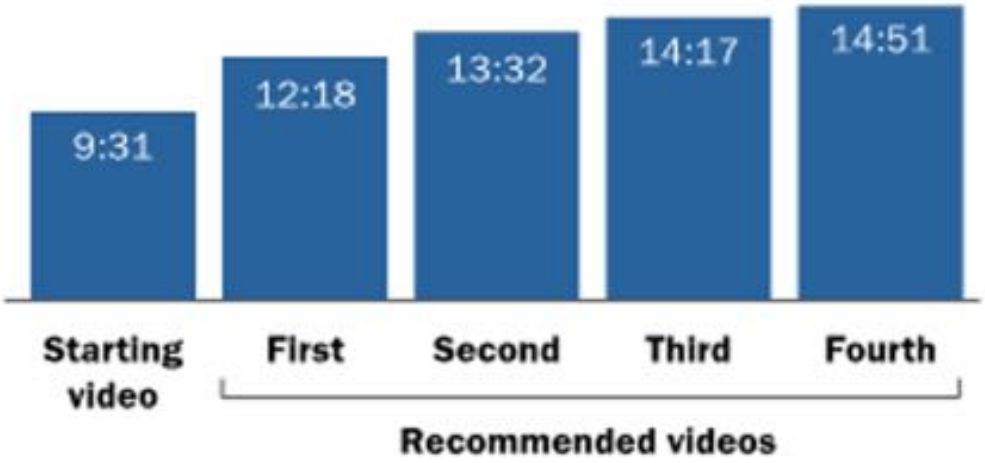



Note: Respondents who did not give an answer are not shown.
Source: Survey of U.S. adults conducted May 29-June 11, 2018.
"Many Turn to YouTube for Children's Content, News, How-To Lessons"

PEW RESEARCH CENTER

## YouTube recommendations point to progressively longer videos

*Average video length (min:sec)*




Source: Analysis of recommended videos collected via 174,117 five-step "random walks" beginning with videos posted to English-language YouTube channels with at least 250,000 subscribers, performed using the public YouTube API. Data collection performed July 18-Aug. 29, 2018.
"Many Turn to YouTube for Children's Content, News, How-To Lessons"

PEW RESEARCH CENTER

**Discussion:**

What factors (other than accuracy) should be taken into considerations in designing a video recommender system?

- What factors (other than accuracy) should be taken into considerations in designing a video recommender system for young children?

# THANK YOU



Haiyi Zhu


WWW.HAIYIZHU.COM


haiyiz@cs.cmu.edu