# Toward Fairness in Al for People with Disabilities: A Research Roadmap

**Anhong Guo**<sup>1,2</sup>, Ece Kamar<sup>1</sup>, Jennifer Wortman Vaughan<sup>1</sup>, Hannah Wallach<sup>1</sup>, Meredith Ringel Morris<sup>1</sup>

Microsoft®
Research





<sup>&</sup>lt;sup>1</sup> Microsoft Research, Redmond, WA & New York, NY, USA

<sup>&</sup>lt;sup>2</sup> Human-Computer Interaction Institute, Carnegie Mellon University, Pittsburgh, PA, USA

# Fairness in AI for People with Disabilities

- Al has huge potential to impact the lives of people w/ disabilities
  - Speech recognition: caption videos for people who are deaf
  - Language prediction: augment communication for people w/ cognitive disabilities

# Fairness in AI for People with Disabilities

- All has huge potential to impact the lives of people w/ disabilities
  - Speech recognition: caption videos for people who are deaf
  - Language prediction: augment communication for people w/ cognitive disabilities
- However, Al systems may not work, or worse, discriminate/harm
  - If smart speakers do not recognize people with speech disabilities
  - If a chatbot learns to mimic someone with a disability
  - If self-driving cars do not recognize pedestrians using wheelchairs

1. Identify potential inclusion issues of AI systems

- 2. Test hypotheses to understand failure scenarios
- 3. Create benchmark datasets for replication and inclusion
- 4. Innovate new methods and techniques to mitigate bias

## 1. Identify potential inclusion issues of AI systems

- A. Categorization of AI capabilities
- Modalities: vision, audio, text, etc.
- Task:
  - Recognition: detection, identification, verification, analysis
  - Generation
- Integrative AI: combinations of the above

### 1. Identify potential inclusion issues of AI systems

#### B. Risk assessment of existing AI systems

- Computer vision: face, body, object, scene, text recognition
- Speech systems: speech recognition, generation, speaker analysis
- Text processing: text analysis
- Integrative AI: information retrieval, conversational agents

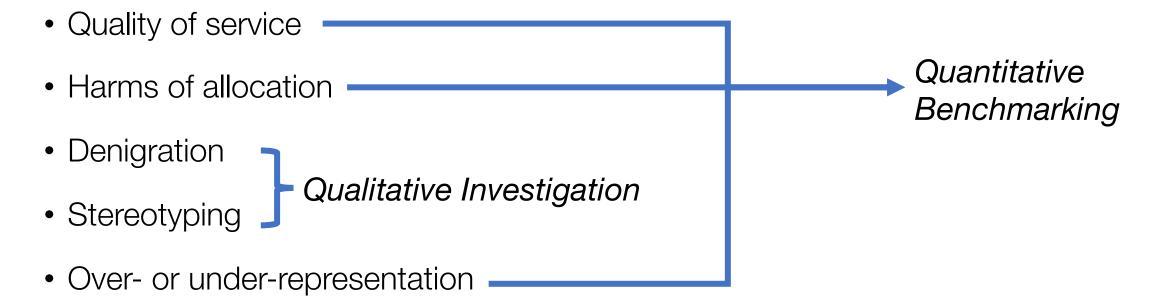
## 1. Identify potential inclusion issues of AI systems

#### C. General Al techniques

- Outlier detection: completion time to determine input legitimacy
- Aggregated metrics: Accuracy, F1, AUC, MSE
- Definition of objective functions
- Datasets: fail to capture use cases, lack representation of diverse groups

## 1. Identify potential inclusion issues of AI systems

#### D. Types of harm by unfair Al



- 2. Test hypotheses to understand failure scenarios
- 3. Create benchmark datasets for replication & inclusion

#### Ethical issues involved in data collection

- Is it acceptable to create such datasets by scraping existing online data?
  - How to preserve users' privacy, while ensures ground-truth labels?
  - Potential harms of aggregating data about disability?
- If curating data from scratch, how can we encourage contributions?
  - How to obtain consent for people with intellectual disabilities?

- 2. Test hypotheses to understand failure scenarios
- 3. Create benchmark datasets for replication & inclusion

#### Potential data collection approach

- First use online sources to perform exploratory analysis; Then use web data call asking people to contribute data
- Dataset should not be re-distributed due to ethical concerns; instead, use evaluation servers to support benchmarking by others

#### 4. Innovate new methods and techniques to mitigate bias

- Evaluate how much existing bias mitigation techniques work
- Design new modeling, bias mitigation, and error measurement techniques

## Thanks!

# Toward Fairness in AI for People with Disabilities: A Research Roadmap

Anhong Guo<sup>1,2</sup>, Ece Kamar<sup>1</sup>, Jennifer Wortman Vaughan<sup>1</sup>, Hanna Wallach<sup>1</sup>, Meredith Ringel Morris<sup>1</sup>

Microsoft Research, Redmond, WA & New York, NY, USA
 Human-Computer Interaction Institute, Carnegie Mellon University, Pittsburgh, PA, USA anhongg@cs.cmu.edu, {eckamar, jenn, wallach, merrie}@microsoft.com

#### Research Roadmap

aka.ms/Ala11y

- 1. Identify potential inclusion issues of AI systems
- 2. Test hypotheses to understand failure scenarios
- 3. Create benchmark datasets to support replication and inclusion
- 4. Innovate new methods and techniques to mitigate bias

## Anhong Guo, CMU HCII

https://guoanhong.com anhongg@cs.cmu.edu