

Ask not *AI can* do, but what *AI should* do: Towards a framework of task delegability

Brian Lubars

brian.lubars@colorado.edu

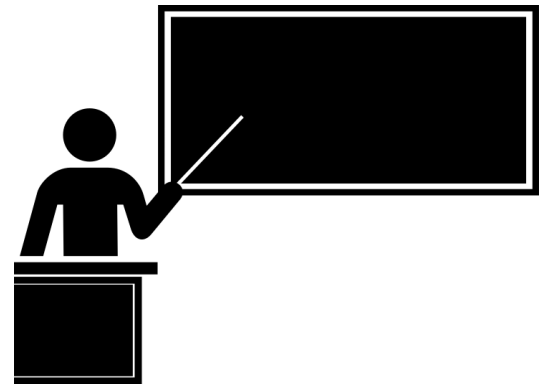
Chenhao Tan

chenhao.tan@colorado.edu


University of Colorado, Boulder

NeurIPS 2019

What can AI do?



What **can** AI do?



AI holds promise for addressing aspects of nearly all societal challenges.

AI applications have led to growing controversies

***In Wisconsin, a Backlash
Against Using Data to
Foretell Defendants' Futures***

AI applications have led to growing controversies

*In Wisconsin, a Backlash
Against Using Data to
Foretell Defendants' Futures*

*San Francisco Bans Facial
Recognition Technology*

**A facial recognition ban is
coming to the US, says an
AI policy advisor**

Bans in cities could spread to states, then up to Washington.

AI applications have led to growing controversies

*In Wisconsin, a Backlash
Against Using Data to
Foretell Defendants' Futures*

THE A.I. "GAYDAR" STUDY AND THE REAL
DANGERS OF BIG DATA

*San Francisco Bans Facial
Recognition Technology*

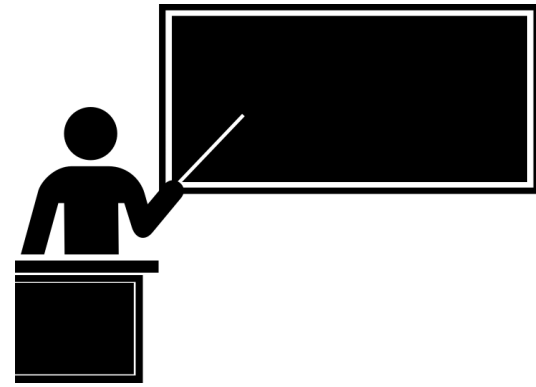
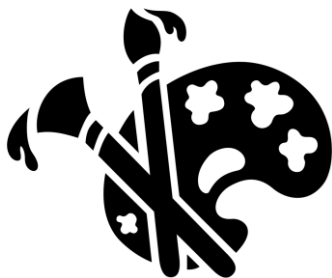
**A facial recognition ban is
coming to the US, says an
AI policy advisor**

Bans in cities could spread to states, then up to Washington.

What *can* AI do?



What *should* AI do?



What *can* AI do?



What *should* AI do?

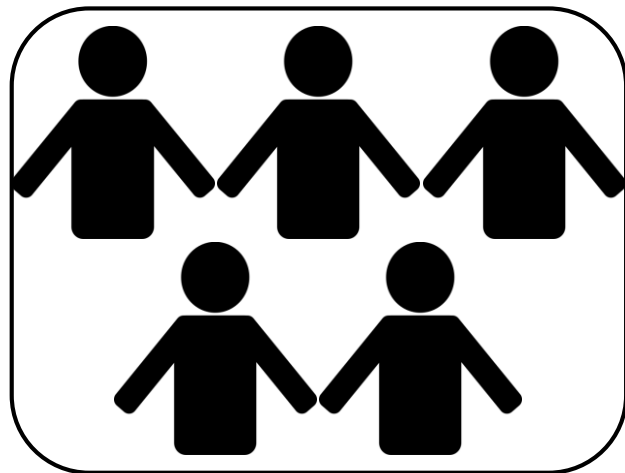
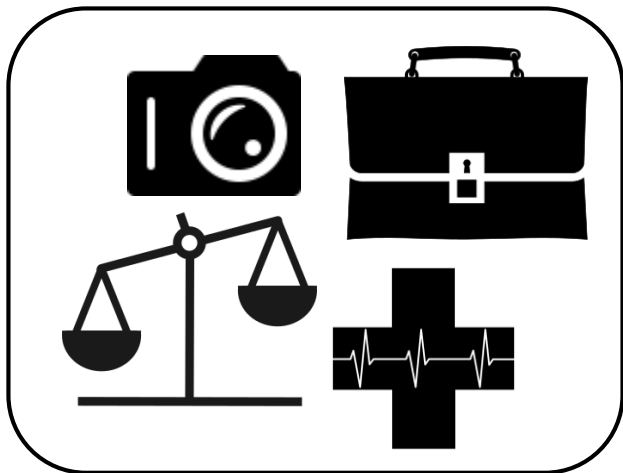


Understanding human preferences of delegation to AI

1. Which tasks do people want automation/machine assistance on?
2. How much machine assistance?

Approach: ask people!

1. A framework for *task delegability* to AI.
2. A dataset of 100 tasks.
3. Survey to measure delegability and validate framework.



Task Delegability Framework

Motivation: why a person performs a task

Difficulty: the process of performing a task

Risk: the outcome of (failing) a task

Trust: the interaction between the person and AI

Delegability:



1) Human-only



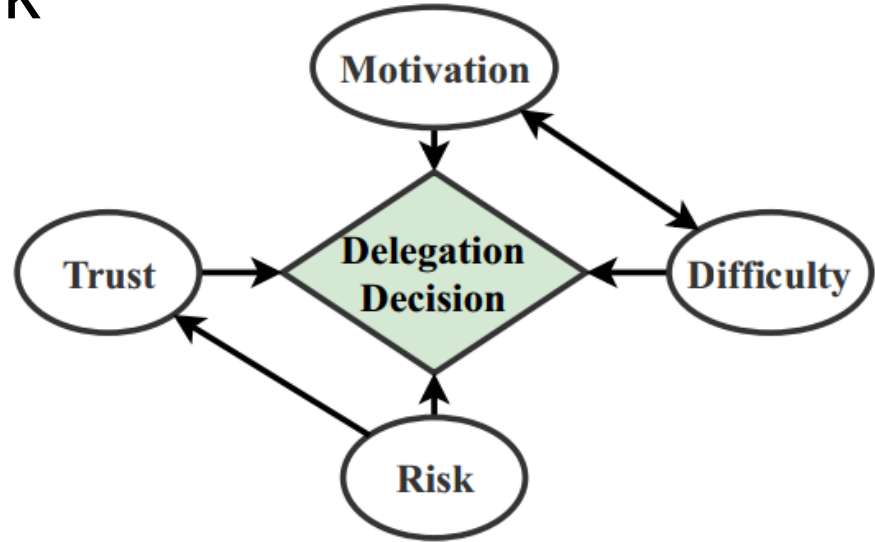
2) Machine-in-the-loop (human in control)



3) Human-in-the-loop (machine in control)

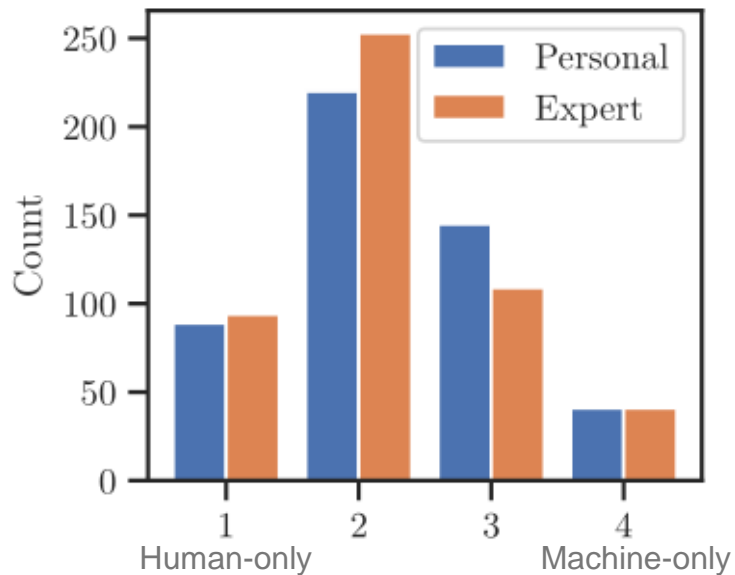


4) Machine only

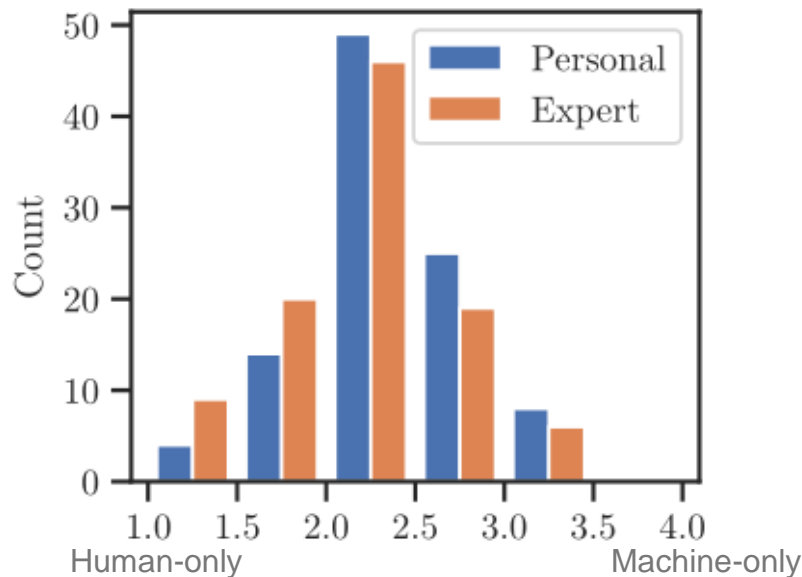


Results

- Most people prefer machine-in-the-loop designs (human in control).



(a) Individual survey responses



(b) Responses averaged by task

Results

- Most people prefer machine-in-the-loop designs
- Trust is the factor most highly correlated with delegability

Factor	Component	Pearson r
Trust	Machine ability	0.52
Trust	Value alignment	0.48
Trust	Interpretability	NS
Difficulty	Social skill requirements	-0.30
Difficulty	Creative skill requirements	-0.22

Results

- Most people prefer machine-in-the-loop designs
- Trust is the factor most highly correlated with delegability
 - Exception: interpretability

Factor	Component	Pearson r
Trust	Machine ability	0.52
Trust	Value alignment	0.48
Trust	Interpretability	NS
Difficulty	Social skill requirements	-0.30
Difficulty	Creative skill requirements	-0.22

Results

- Most people prefer machine-in-the-loop designs.
- Trust is the factor most highly correlated with delegability.
- Social & creative tasks are negatively correlated with delegability.

Factor	Component	Pearson r
Trust	Machine ability	0.52
Trust	Value alignment	0.48
Trust	Interpretability	NS
Difficulty	Social skill requirements	-0.30
Difficulty	Creative skill requirements	-0.22

Case study: medical domain

Task Description	Social skills required (Difficulty)	Doctor's ability (Difficulty)	Impact (Risk)	Machine ability (Trust)	Delegability
Medical Diagnosis: Flu	3.4	4.6	4.2	3	2.4
Medical diagnosis: cancer	2.6	3.6	4.8	2.4	2
Explaining treatment options: cancer	4.4	4.2	4.6	2.4	1.4

Three selected task results from our *expert* surveys.

Takeaways

- Understanding and tracking public preferences of delegation to AI: valuable source of information
 - Machine-in-the-loop designs are typically preferred.
 - Trust is most correlated with delegation preferences.
 - Interpretability is not strongly correlated, although people do find it important in some tasks.
- First steps towards a delegability framework

Thank you!

<https://delegability.github.io>