

CSE 518A

# Crowdsourcing and Human Computation

Instructor: Chien-Ju Ho

# Plan for today

- Welcome and introduction
- What's the class about?
- Logistics

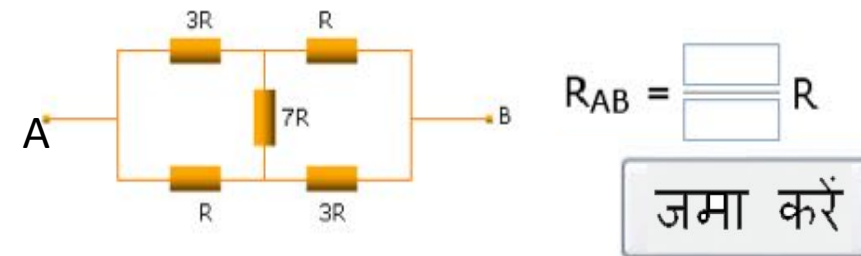
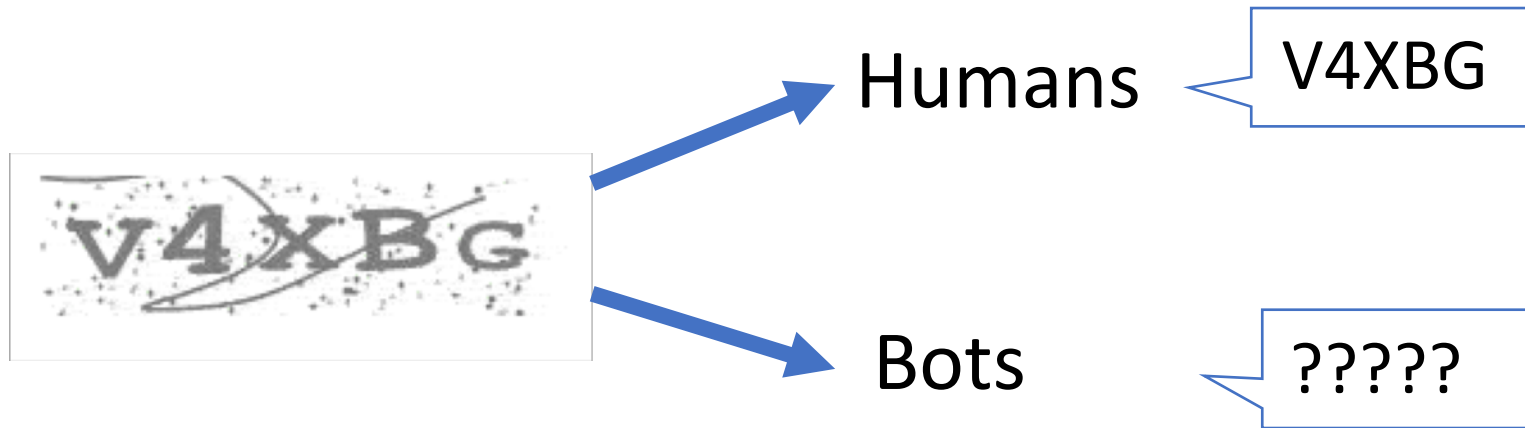
# About Me

- Joined WashU in Fall 2017.
- Research interests:
  - Design and analysis of human-in-the-loop systems
  - Crowdsourcing and human computation, machine learning, game theory, optimization, online behavioral social science, and human-computer interactions.

What is Crowdsourcing and Human Computation?

# CAPTCHA

**C**ompletely **A**utomated **P**ublic **T**uring test to tell **C**omputers and **H**umans **A**part



Roughly 200 million CAPTCHAs are typed every day\*

10s of human time per CAPTCHA

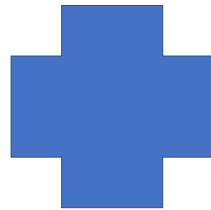
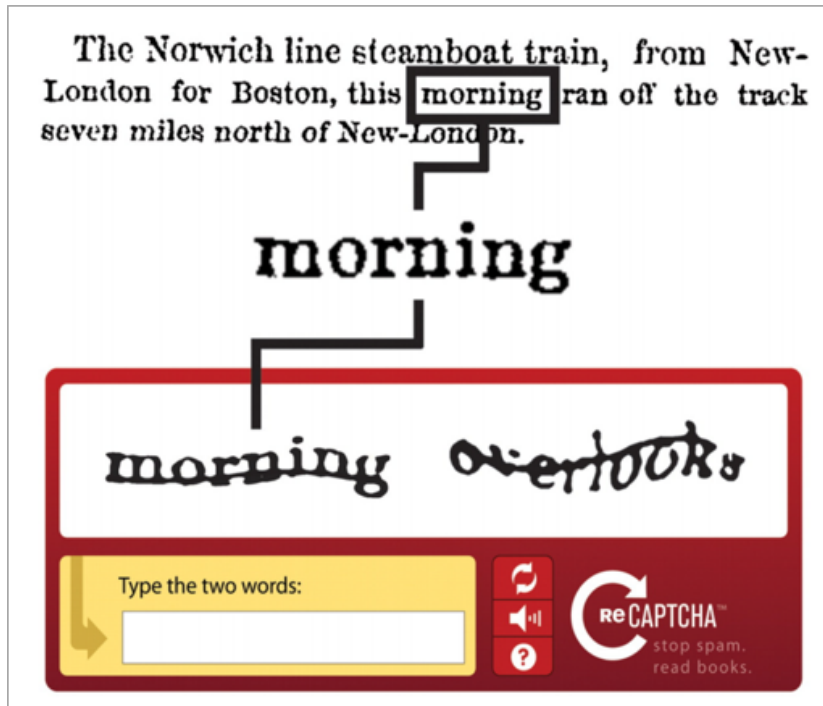
Can we utilize this wasted human computation power?

# What tasks are humans solving in CAPTCHAs?

- Optical Character Recognition (OCR)

**This aged portion of society were distinguished from**

Can we utilize CAPTCHAs to help solve OCR tasks?



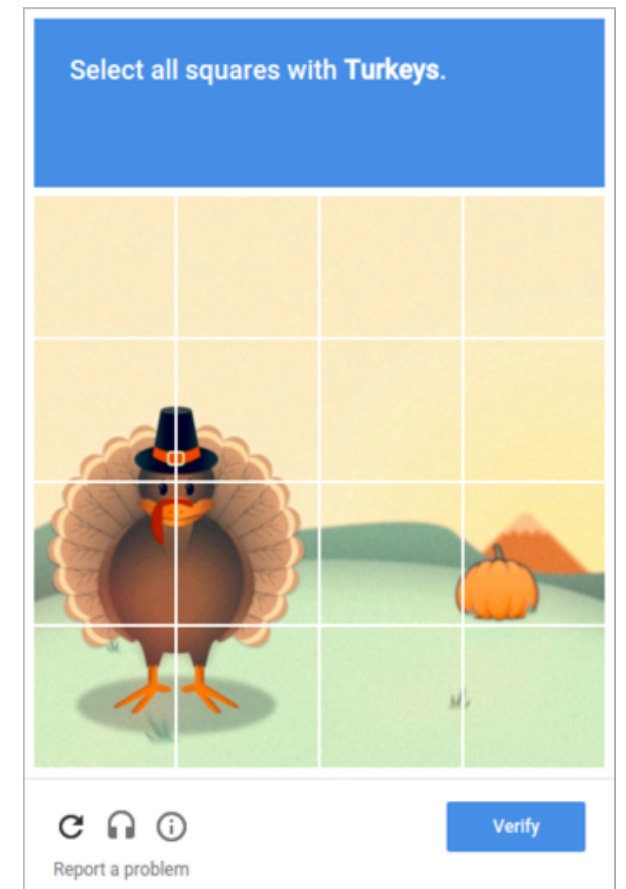
Word 1: an OCR task to solve  
Word 2: tell apart humans and bots

“reCAPTCHA has completely digitized the archives of The New York Times and books from Google Books, as of 2011”



# More than OCR

- Google acquired reCAPTCHA in 2009.

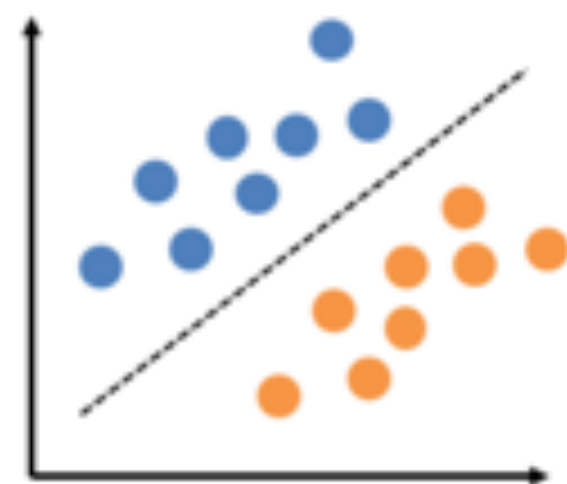




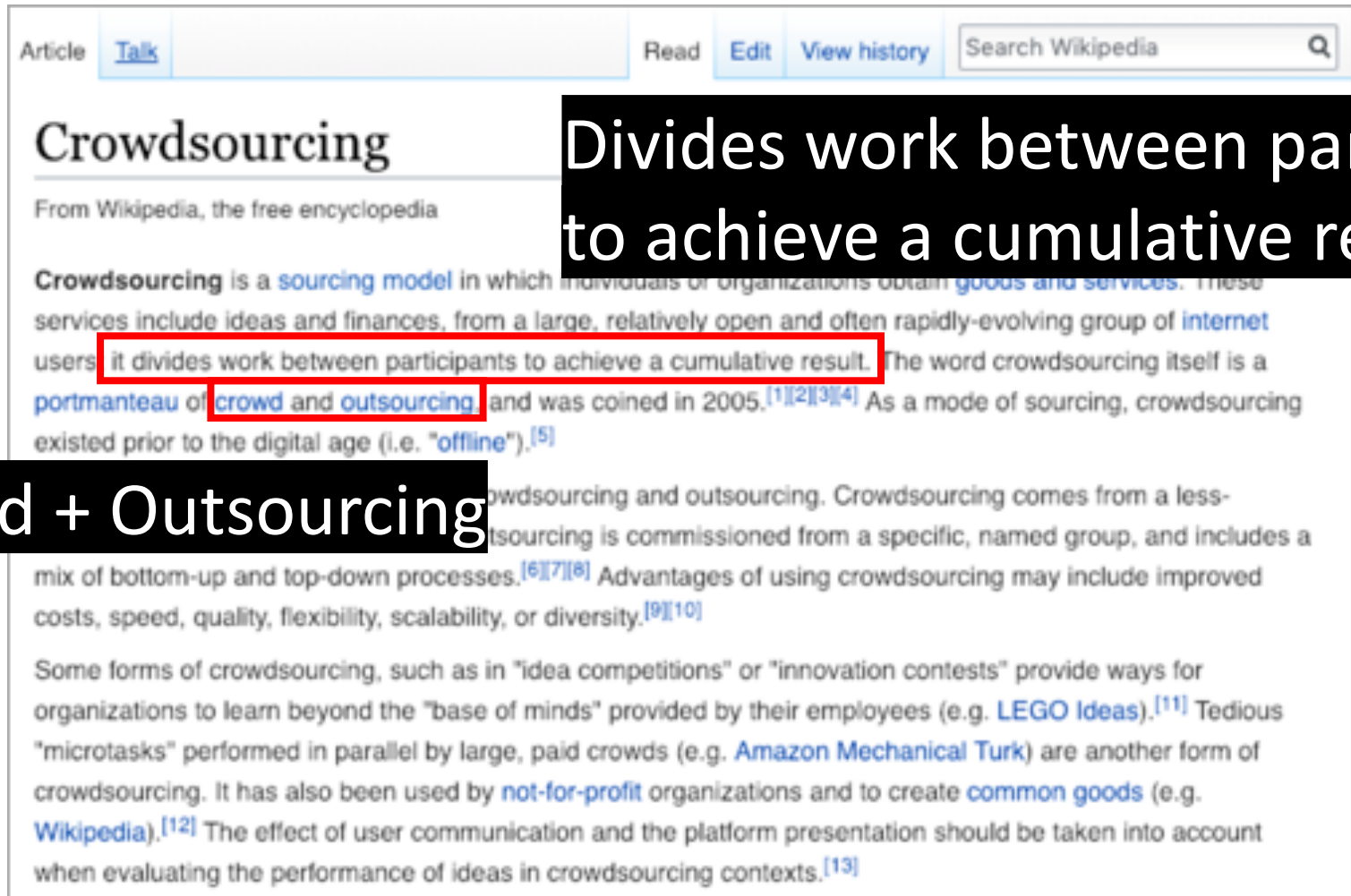
*Training Data*



*Hard Tasks*



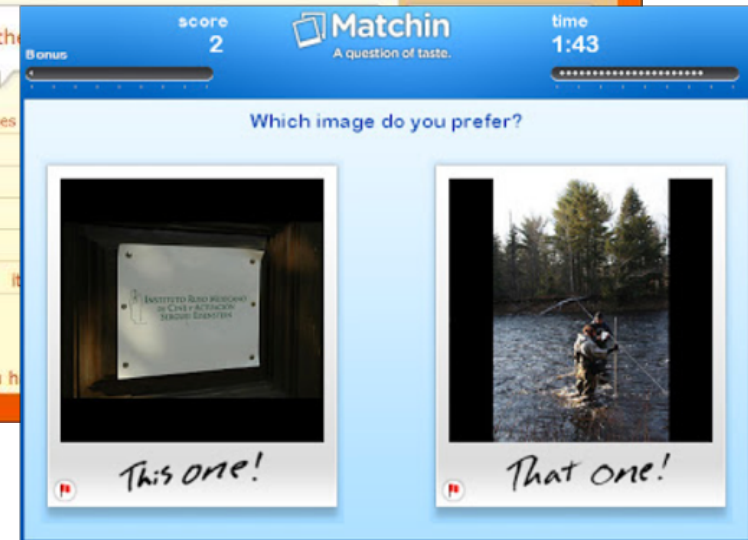
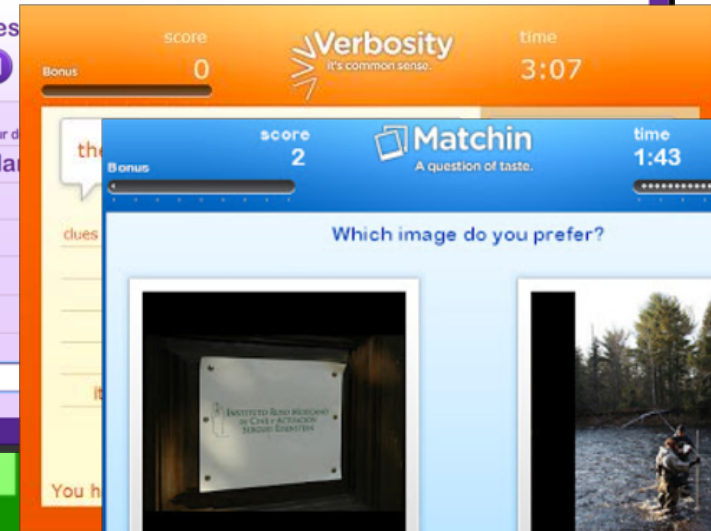
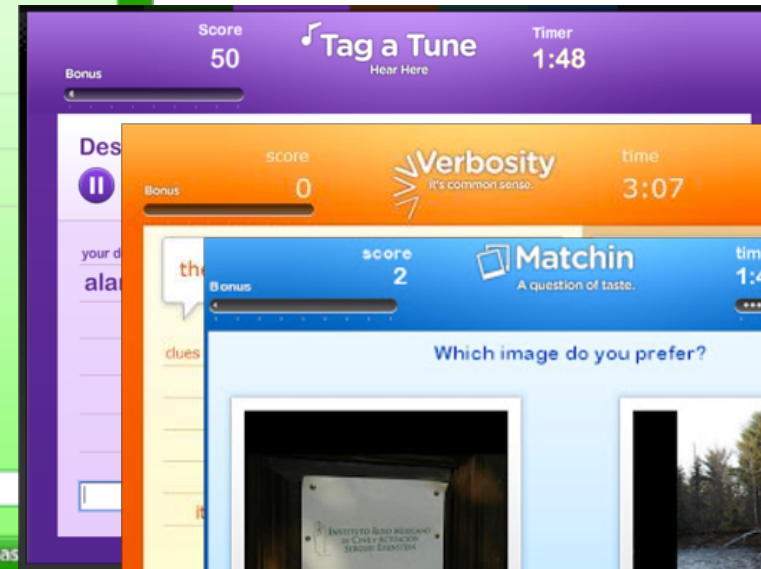
Are there other success example of  
crowdsourcing and human computation?



Divides work between participants to achieve a cumulative result

Crowd + Outsourcing

Utilize the crowd of humans to help solve tasks that computers/AIs cannot solve yet.



HEALTHY LIVING 09/19/2011 03:37 pm ET | Updated Nov 19, 2011

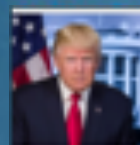
## Gamers Decode AIDS Protein That Stumped Researchers For 15 Years In Just 3 Weeks

von Ahn and Dabbish. "Labeling Images with a Computer Game". 2004.



# How often will Trump tweet this week?

PREDICT IT NOW!



**39 or fewer**

39.POTUSTWEETS.020717

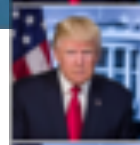
1¢ ↓ 1¢

2¢

1¢

99¢

98¢



**40 - 44**

40.POTUSTWEETS.020717

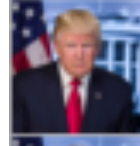
2¢ ↓ 5¢

3¢

2¢

98¢

97¢



**45 - 49**

45.POTUSTWEETS.020717

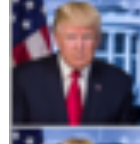
12¢ ↓ 11¢

14¢

12¢

88¢

86¢



**50 - 54**

50.POTUSTWEETS.020717

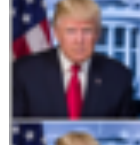
19¢ ↓ 7¢

21¢

19¢

81¢

79¢



**55 - 59**

55.POTUSTWEETS.020717

24¢ ↑ 1¢

25¢

23¢

77¢

75¢



**60 - 64**

60.POTUSTWEETS.020717

26¢ ↑ 12¢

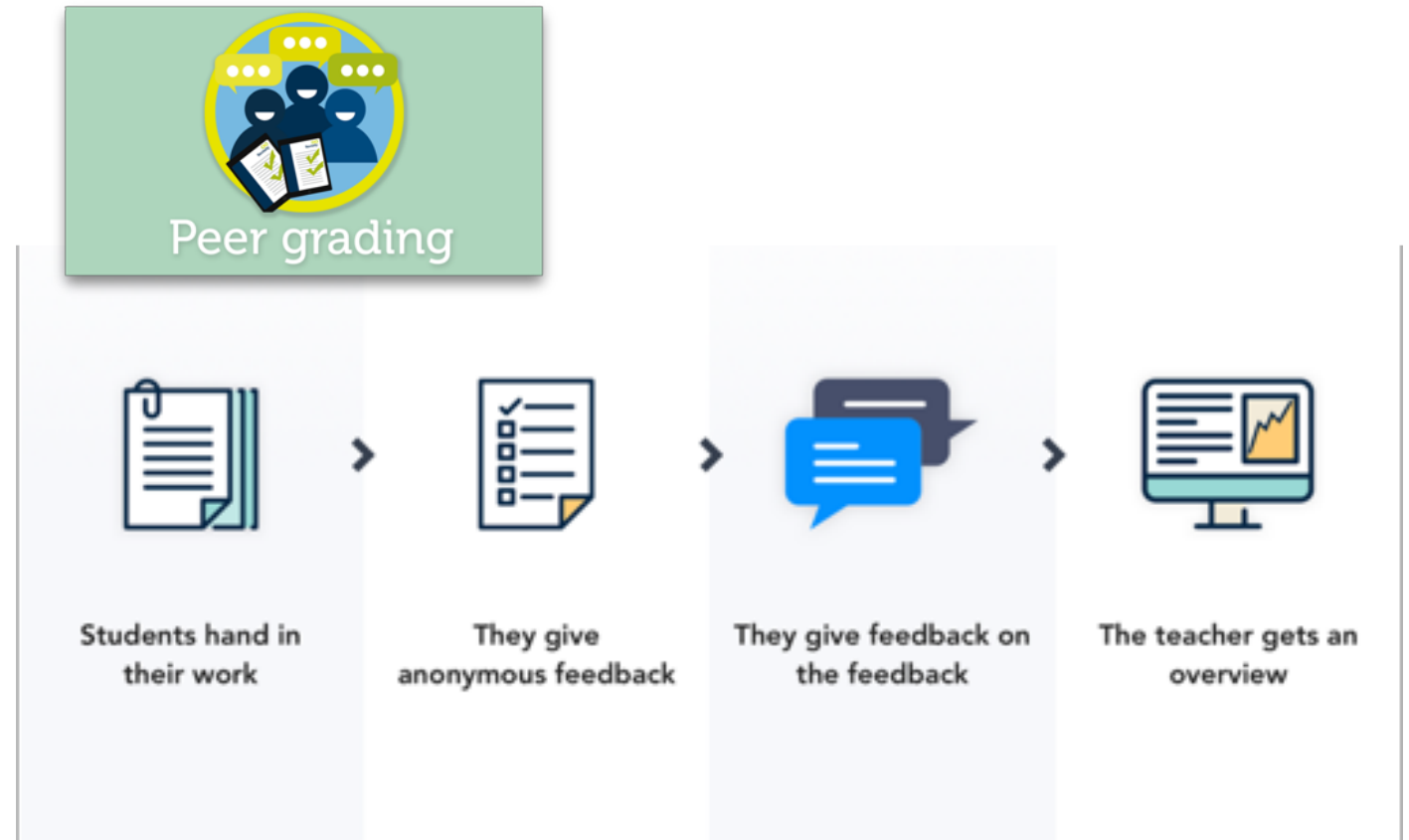
25¢

23¢

77¢

75¢

# Applying crowdsourcing to education



In fact, a lot more...





# General-Purpose Platform: Crowdsourcing Markets



## HIT Groups (1-20 of 1318)

Show Details Hide Details Items Per Page: 20

| Requester     | Title   | HITs   | Reward | Created | Actions                 |                                   |
|---------------|---|--------|--------|---------|-------------------------|-----------------------------------|
| Megan         | Categorization  | 45,696 | \$0.01 | 1h ago  | <a href="#">Preview</a> | <a href="#">Quality</a>           |
| Berch Mturk   | Kitchen Appliance Classification                          | 14,958 |        |         |                         |                                   |
| Kendra Dodson | Find email address and first/last name of Office Manag... | 9,327  |        |         |                         | <a href="#">Work</a>              |
|               | Find email address and first/last name of Office Manag... | 8,677  | \$0.11 | 1d ago  | <a href="#">Preview</a> | <a href="#">Accept &amp; Work</a> |
|               | Why is this review positive?                              | 7,965  | \$0.01 | 6d ago  | <a href="#">Preview</a> | <a href="#">Accept &amp; Work</a> |
|               | Why is this review negative?                              | 7,058  | \$0.01 | 6d ago  | <a href="#">Preview</a> | <a href="#">Accept &amp; Work</a> |
|               | Market Research Survey                                    | 6,680  | \$0.01 | 1h ago  | <a href="#">Preview</a> | <a href="#">Accept &amp; Work</a> |
| Kendra Dodson | Find email address and first/last name of owners or ge... | 4,511  | \$0.11 | 1d ago  | <a href="#">Preview</a> | <a href="#">Accept &amp; Work</a> |
| Geoffrey      | Classify Receipt  | 4,000  | \$0.02 | 2m ago  | <a href="#">Preview</a> | <a href="#">Accept &amp; Work</a> |

### Post Tasks:

- Audio transcription
- Image tagging
- Relevance evaluation
- Handwriting recognition
- Product information collection

Specify payments

# What is this course about?

- Study the design and analysis of crowdsourcing systems.
- Research questions:
  - Who are the people involved?
  - How to motivate people to work?
  - How to aggregate the noisy data contributed by multiple users?
  - How to optimize the workflow?
  - How to deal with practical challenges?
- Will cover research papers from a wide spectrum of research fields, including machine learning, economics, optimization, and human-computer interactions.

# Let's take a look at the course schedule

- <http://chienjuho.com/courses/cse518a/sp2019/>

# Grading

- Course Project: 50%
- Homework assignment: 15%
  - 2~3 homework assignments
- Paper reviews and class participation: 15%
- Paper presentation and leading of discussion: 20%

# Course Project

- The main component of the course.
- The project can be an **original research project** or an **extensive literature survey**.
  - You are strongly encouraged to start with a research project. You will have the chance to convert the project to literature review if things don't go well.
- Tentatively, you should work in groups of 2~3 students
  - Will announce the detailed guidelines next week after the class size is finalized.

# Timeline (Tentative)

- Jan 29: Deciding the team members
- Feb 19: Project proposal
  - Brief description of the proposed project (1~2 paragraph)
  - Citing at least one paper that's relevant to your proposal
- Mar 3: Milestone 1
  - A brief literature review and the description of your plan (one page)
  - Last chance to change the topic of the project
- Apr 9: Milestone 2
  - Summary of your current progress (2 pages)
  - Last chance to convert the research project to (a more extensive) literature review
- Apr 23/25: In-class project presentations
- Apr 28: Project report due

# HCOMP 2019

- The seventh AAIL Conference on Human Computation and Crowdsourcing.
  - <https://www.humancomputation.com/>
- Full paper due: June 5, 2019
  - Could be a good target for submitting a paper.

# Paper Reviews

- Finish the required reading and submit a review (including a summary and answering additional questions).
  - Due by the midnight before each lecture.
  - Exception: the review for next lecture is due at noon the day of the lecture.

Jan 17    Crowdsourcing: Background and Applications

**Required**

[The Rise of Crowdsourcing](#). Howe. Wired. 2006.

[Submit Review](#)

(Due: Noon, Jan 17).

**Optional**

[Labeling Images with a Computer Game](#). von Ahn and Dabbish. CHI 2004.

[reCaptcha: Human-based Character Recognition via Web Security Measures](#).

von Ahn et al. Science. 2008.

[Predicting Protein Structures with a Multiplayer Online Game](#). Cooper et al.

Nature. 2010.



# Do I *really* need to read the whole paper?

- You are not expected to understand every detail or every proof in the papers that we cover.
- However, you should be ready to answer the following questions:
  - Summary the paper in 2~3 sentences.
    - What's the research question the paper is solving?
    - What's the proposed approach?
    - What are the results?
  - List 1~3 points you like about the paper.
  - Coming up with potential future directions based on this paper.
    - It might be hard in the beginning. Try to think about what assumptions they make, and whether you can relax some of those? Can you apply the method/approach of the paper in different domains/applications, etc?

# Paper Presentations and Leading of Discussion

- Take a look at the topics from Feb 7 to Apr 11.
  - You will be asked to sign up to present the paper(s) and lead the discussion for one of the lectures, possibly in groups of 2~3 students (again, more to come next week).
- Presenters:
  - Read the required paper and at least one optional paper for the assigned class
  - Discuss with me (one week before class) about the presentation and the reading questions
  - Give a presentation for 70~80 minutes and lead the discussion in class
- Non-presenters:
  - Submit reviews on time and engage in the discussion in class.

# Grading

- Course Project: 50%
- Homework assignment: 15%
  - 2~3 homework assignments
- Paper reviews and class participation: 15%
- Paper presentation and leading of discussion: 20%

# Take another look at the syllabus

- Collaboration policy
- Late day policy

# Next lecture:

- Read the required reading of next lecture and submit the review!

|        |  |  |   |
|--------|--|--|---|
| Jan 17 | Crowdsourcing: Background and Applications | <p><b>Required</b><br/><a href="#">The Rise of Crowdsourcing</a>. Howe. Wired. 2006.</p> <p><b>Optional</b><br/><a href="#">Labeling Images with a Computer Game</a>. von Ahn and Dabbish. CHI 2004.<br/><a href="#">reCaptcha: Human-based Character Recognition via Web Security Measures</a>.<br/>von Ahn et al. Science. 2008.<br/><a href="#">Predicting Protein Structures with a Multiplayer Online Game</a>. Cooper et al.<br/>Nature. 2010.</p> | <p><a href="#">Submit Review</a><br/>(Due: Noon, Jan 17).</p> |
|--------|--|--|---|

- Please submit the review on time so I can get a better sense of how many students will stay in the course.
- You **cannot** apply the late-day rule for this review.

# Another thing to do

- Register as a worker of Amazon Mechanical Turk and Figure Eight today
  - MTurk: <https://www.mturk.com/>
  - Figure Eight: <https://www.figure-eight.com/>
- We will ask you to be a worker yourself and complete tasks on crowdsourcing markets in Assignment 1
  - MTurk: After you successfully register, Amazon will send you invitation for working, which may takes several business days.
- Do it early, and let me know if there are any issues in the process.

Questions?