Lecture 2:
Crowdsourcing: Background and Applications

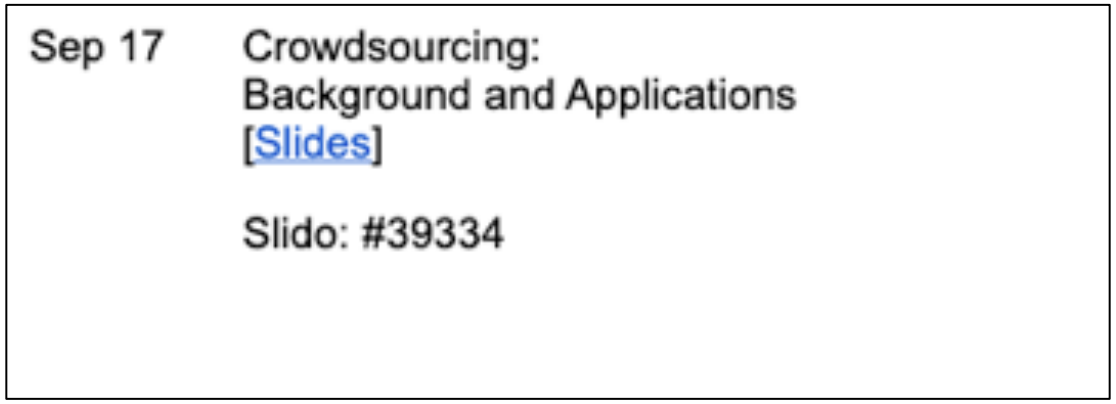
Questions: https://sli.do #39334

Instructor: Chien-Ju Ho

Logistics

Questions during lectures

- We plan to use Slido (http://sli.do) to collect questions during lectures
  - It applies the idea of crowdsourcing to rank questions
  - Please check the course page for the event code of each lecture



- Please also feel free to unmute yourself and ask questions directly

Piazza and Links to Reading Materials

- Sign up for Piazza
  - http://piazza.com/wustl/fall2020/cse518a

- Reading materials
  - Let me know if you find any of the links broken.
  - You might need VPN to get access to some of the files when outside of WashU.
  - Search paper titles in search engines should also lead you to the papers.

Assignment 1

- http://chienjuho.com/courses/cse518a/assignment1.pdf

- Submit via Gradescope.
- You should have already been added to the rosters on Gradescope.
- Due date:
  - October 2 (Friday)
  - **Do 3 tasks** AND **earn $0.25** as a crowd worker
  - MTurk is recommended
    - If you are denied registration for a new MTurk account, you can
      - Borrow accounts from your friends/classmates
      - Use other crowdsourcing platforms

Requirements of Assignment 1

Register a worker account in one of the crowdsourcing platforms (Amazon Mechanical Turk, Appen, clickWorkers, or microWorkers). Complete at least 3 different tasks and earn at least $0.25 (both conditions need to be met) on one of the crowdsourcing platforms. Keep track of the amount of time you spend in doing the tasks.

1. Provide screenshots to show that you have done the tasks (any reasonable screenshots will do). For example, you can show the earning summaries or the screenshots of the tasks.
2. Based on the amount of time you spend in doing the tasks and the amount of money you earn, what is your estimated hourly wage as a crowd worker?
3. Choose one of the tasks you have done and answer the following questions
   (a) Briefly describe the task.
   (b) Provide a guess of what the requester is using your data for.
   (c) Do you like the way the task is designed? In your opinion, what can the requester do to improve the task design? (You can answer this question from the perspectives of how to be more fair to workers, how to make the task easier to complete for workers, how to improve the efficiency of data collection, etc).
4. In your opinion, what can the platform (e.g., MTurk) do to improve the crowdsourcing process?

1. Proof that you have done it
2. Estimate hourly wage
3. How should the requester improve?
4. How should the platform improve?

Getting a worker account is probably the most challenging part of this assignment for many of you.

Please start early. Let me know if there are any issues.

There will be no extensions if the issues are brought up right before the deadline.

Logistics: Prerequisites

- Need to be comfortable with several math concepts and basic programming
  - Probability
  - linear algebra
  - calculus
  - basic concepts about computer science and ability to program

- If you are not sure, please take a peek at the papers from **Sep 24 – Oct 15**
  - You should at least be able to understand the formulation and main results
  - You will need to implement some algorithms in these papers

Logistics: Grading of Reviews and Assignments

- Reviews grading
  - If you show reasonable effort, you will get full points
  - Please use complete/full sentences to structure your responses
  - I do read through all the review responses
  - You can assume you get full points if not notified by me within 2 weeks of submission

- Typical reviews involve 4 questions
  - Summarize the paper in 2~3 sentences
  - List 1~3 points you like about the paper
  - Two paper-specific questions

- Homework assignment grading
  - Again, they are loosely graded; minor issues won't get point reductions
  - I do pick out mistakes/issues and deduct points

Logistics: Presentations and Leading of Discussion

- You need to present papers and lead discussions as a group
  - Need to do it once (or at most twice) this semester
  - The target number of groups is 10~12
  - The size of the group depends on the final class size
    - most likely 2 students per group
  - Talk to me a week before the lecture.
  - Strongly encouraged to include discussions in class.
- More details will be announced next week

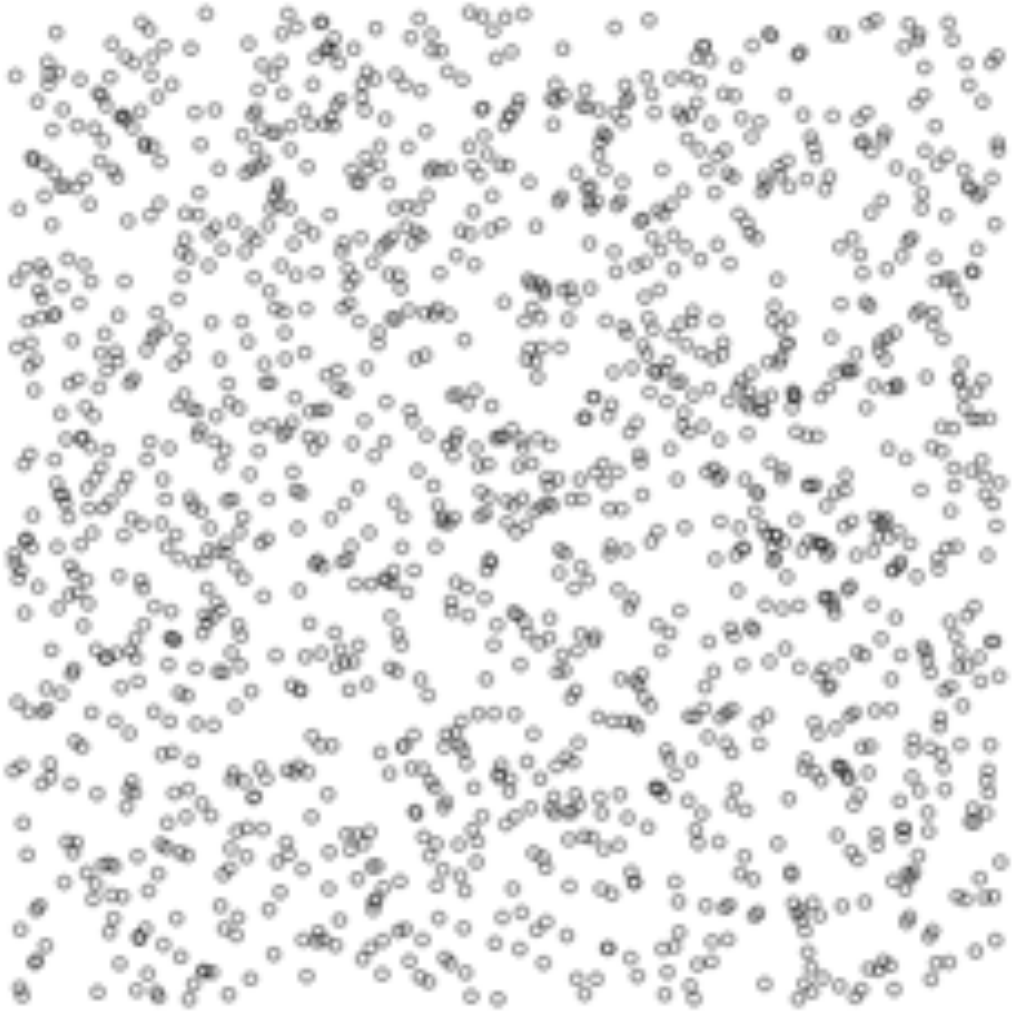
Logistics: Course Projects

- Project proposal (Tentative Due: Oct 9)
  - You can work on anything involving humans in the computation process (theory, empirical, application, literature survey)
  - I'll make the final call on whether it's relevant to the course and might suggest modifications.

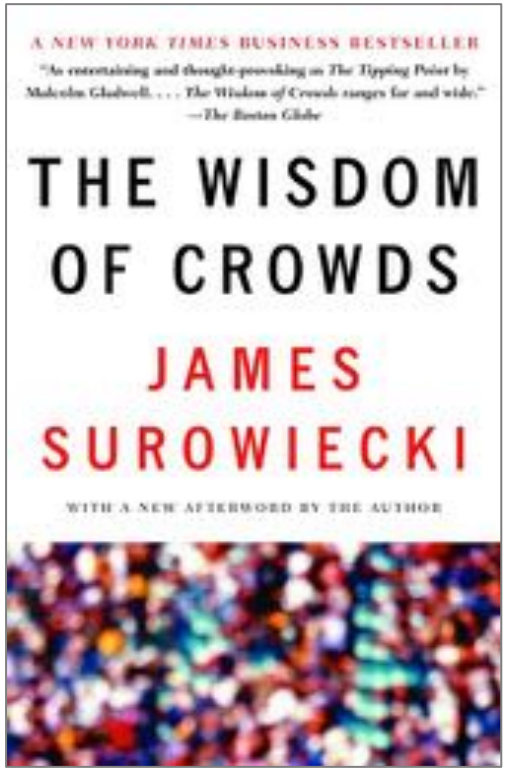
- How to choose topics
  - I'll generate a list of potential/past projects next week.
  - Look at future lectures to see what topics you might be excited about.
  - Feel free to schedule appointments with me to discuss your course projects.

Lecture Today

How many circles are in the image?

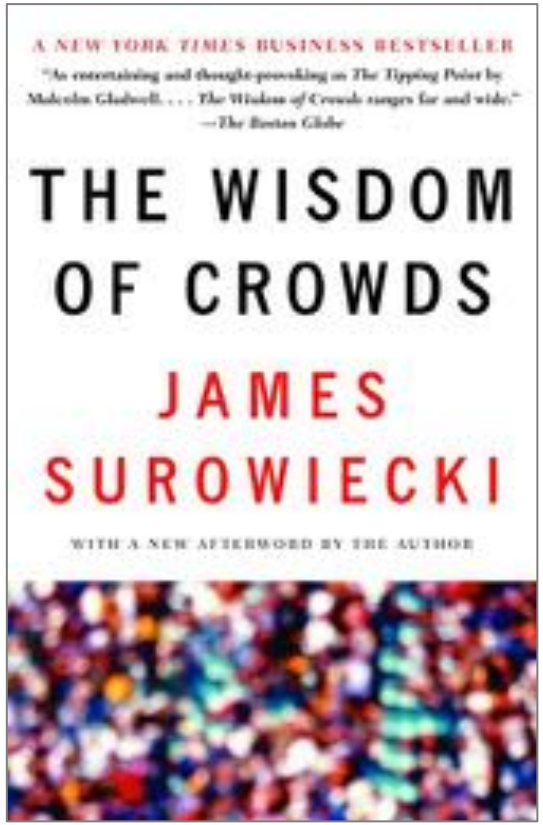


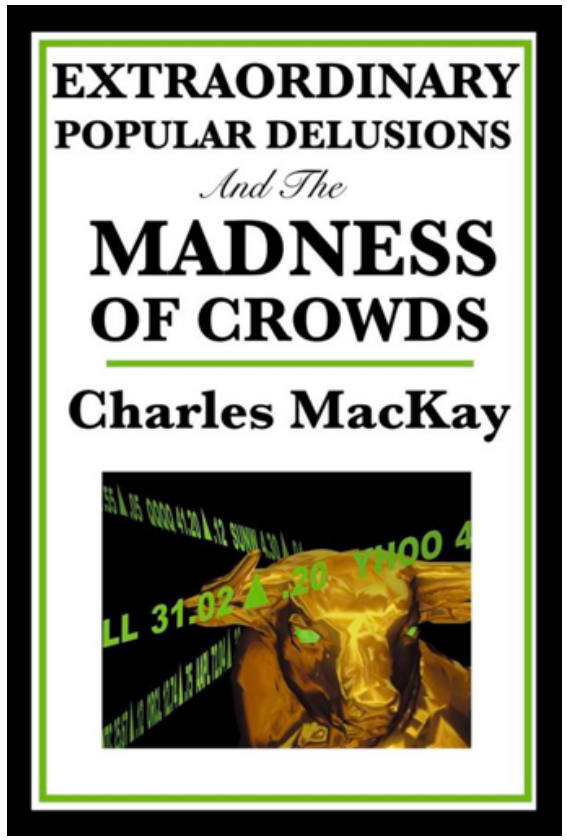
The Wisdom of Crowds

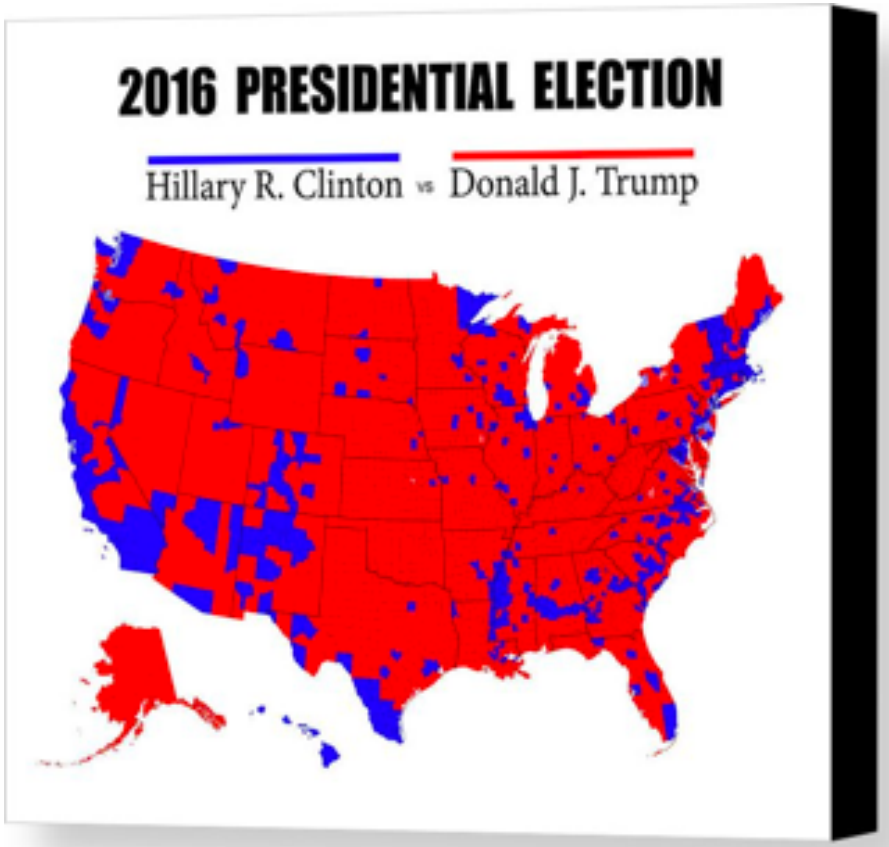


- At a 1906 country fair, ~800 people participate in a contest to guess the weight of an ox.
- Reward is given to the person with the closest guess.
- The average guess is 1,197lbs.
  The true answer is 1,198lbs.

When is the Crowd Wise?







Warm-up Discussion

- Under what conditions is the crowd wise?
- How should we (mathematically) **model** the crowdsourcing process?
  - How to combine the results?
  - Can we estimate how accurate the combined result is?
  - Can we better design the process to motivate higher-quality data?
  - …
- Notes about discussion
  - Please spend a minute for introducing yourselves to others!
  - **Be prepared to share** what you have discussed. We will ask volunteers and then also randomly pick people to share.

When is the crowd wise?

- Some general “beliefs” for crowd to be wise
  - Informed individuals
  - Diversity of opinions
  - Independence
  - Good aggregation
  - Aligned incentives
  - …
- Can we make these ideas more formal?
  - Can we “model” the process?

All models are wrong
but some are useful

George E.P. Box

Example Model on Aggregation

- People have unbiased estimates of the true answer

user guess = **true answer** + **Gaussian noise**

Observations | Latent values we want to know | Zero-Mean Noises

- What can we do with this model?
  - Assume the model somewhat captures the reality,
    - we can infer the **true answers** from the **user guesses** we observe (average is a good estimate)
    - we can further analyze the ”number of guesses” we need to achieve certain accuracy
  - Does this model capture the reality?
    - What if everyone has different abilities in guessing? How do we perform inference?
    - What if there are some unknown biases?

Sep 22	Humans as Data Sources: Label Aggregation Review of Techniques - Label aggregation - Concentration bounds - Maximum likelihood estimation	**No required reading** (No reviews needed) **Reference materials** Probability cheatsheet You should feel very comfortable with at least the first 1.5 pages of this cheatsheet to take this course.
Sep 24	Label Aggregation: EM-based Algorithms	**Required** Whose Vote Should Count More: Optimal Integration of Labels from Labelers of Unknown Expertise. Whitehill et al. NIPS 2009. **Optional** Learning from Crowds. Raykar et al. JMLR 2010. Maximum Likelihood Estimation of Observer Error-Rates Using the EM Algorithm. Dawid and Skene. Applied Statistics. 1979.
Sep 29	Label Aggregation: Matrix-based Methods	**Required** Who Moderates the Moderators? Crowdsourcing Abuse Detection in User-Generated Content. Ghosh, Kale, and McAfee. EC 2011. - If you want to refresh your memory on matrix algebra, Matrix Cookbook is a good resource. Section 5 contains the matrix decomposition part. - This reading is mathematically heavy. Try to at least understand the model and the key results. It's ok if you do not fully comprehend all proof/technical details. **Optional** Budget-Optimal Crowdsourcing using Low-rank Matrix Approximations. Karger, Oh, and Shah. Allerton 2011. Spectral Methods Meet EM: A Provably Optimal Algorithm for Crowdsourcing. Zhang et al. JMLR 2016.
Oct 1	Biases in Human-Generated Data	**Required** Towards Fairer Datasets: Filtering and Balancing the Distribution of the People Subtree in the ImageNet Hierarchy. Yang et al. FAT* 2020. **Optional** Understanding and Mitigating Worker Biases in the Crowdsourced Collection of Subjective Judgments. Hube et al. CHI 2019. How Do We Talk about Other People? Group (Un)Fairness in Natural Language Image Description. Otterbacher et al. HCOMP 2019.

Example Model on Incentives

- People choose effort level in generating their true answer

  user-answer = true-answer + noise(effort)

  user utility = reward * Pr(getting reward | effort) – cost(effort)

  user **chooses effort** that **maximizes their utility**

- What can we do with the models
  - Designing the optimal incentive schemes
  - How much we should pay, should we offer a bonus?
  - What to design non-financial incentives (badges, leaderboards, etc).

Oct 8	Overview: Game Theory and Incentive Design - Game theory basics - Scoring rules - Peer prediction	**No required reading** (No reviews needed) **Reference materials** - Page 47-64 of Multiagent Systems by Shoham and Leyton-Brown - Section 1 and Section 3.1 of Strictly Proper Scoring Rules, Prediction, and Estimation by Gneiting and Raftery. - Tim Roughgarden's lecture notes on scoring rules and peer prediction.
Oct 13	Incentive Design: Financial Incentives	**Required** Incentivizing High Quality Crowdwork. Ho et al. WWW 2015. **Optional** Financial Incentives and the "Performance of Crowds". Mason and Watts. HCOMP 2009. The Effects of Performance-Contingent Financial Incentives in Online Labor Markets. Yin et al. AAAI 2013. The Effects of Pay-to-Quit Incentives on Crowdworker Task Quality. Harris. CSCW 2015. Adaptive Contract Design for Crowdsourcing Markets: Bandit Algorithms for Repeated Principal-Agent Problems. Ho, Slivkins, and Vaughan. JAIR 2016.
Oct 15	Incentive Design: Badges and Attention	**Required** Steering User Behavior with Badges. Anderson et al. WWW 2013. **Optional** Incentives, Gamification, and Game Theory: An Economic Approach to Badge Design. Easley and Ghosh. EC 2013. Social Status and Badge Design. Immorlica, Stoddard, and Syrgkanis. WWW 2015 Incentivizing High-Quality User-Generated Content. Ghosh and McAfee. WWW 2011. Learning and Incentives in User-Generated Content: Multi-Armed Bandits with Endogenous Arms. Ghosh and Hummel. ITCS 2013.
Oct 20	Application: Darpa Network Challenge	**Required** Time Critical Social Mobilization. Pickard et al. Science 2011. Here is the long version of the article if you want to learn more about the details. **Optional** Mechanisms for Multi Level Marketing. Emek et al. EC 2011. Task Routing for Prediction Tasks. Zhang et al. AAMAS 2012. Maximizing the Spread of Influence through a Social Network. Kempe, Kleinberg, and Tardos. KDD 2003.
Oct 22	Application: Prediction Markets	**Required** Results from a Dozen Years of Election Futures Markets Research. Berg et al. 2001. You should first read the first page of the following two-page (not too technical) article for a brief explanation of prediction markets. The Promise of Prediction Markets, K.J. Arrow et. al., Science. 2008

The focus of this course

- Design and analysis of human-in-the-loop computation

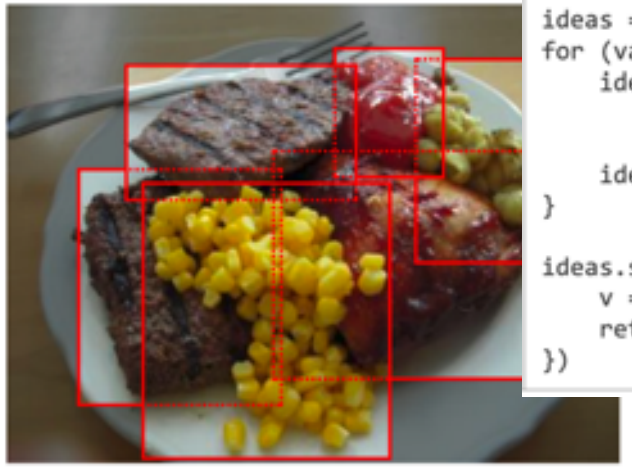
- Develop some “models” of how users take actions in our systems
- Based on the model,
  - design methods to **aggregate** users information
  - design **incentives** to encourage high effort and truthful reports

- Deal with **practical challenges (complex tasks)**
- Discussion on **selected recent topics** **on human-AI interactions**

The focus of this course

- Design and analysis of human-in-the-loop computation
- Deal with **practical challenges (complex tasks)**

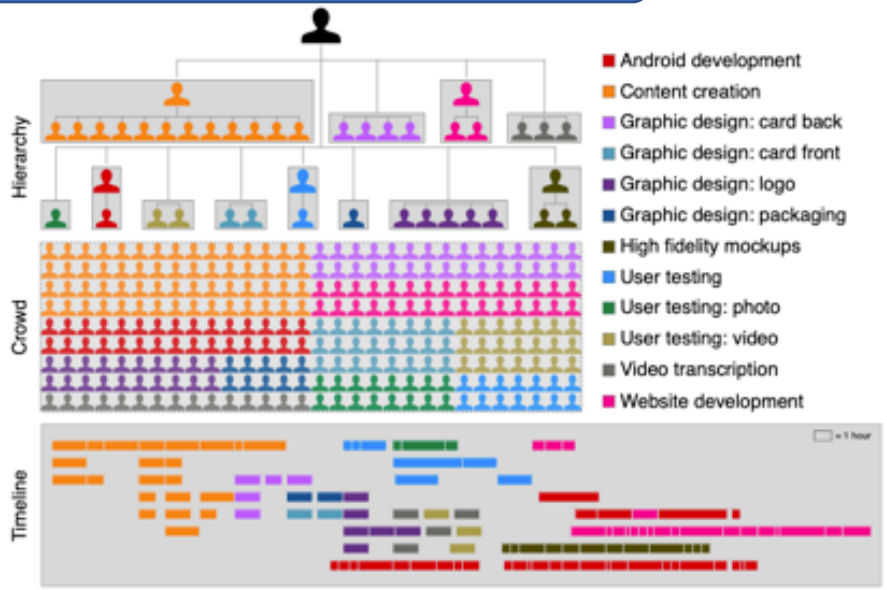
New Programming Paradigm



```
ideas = []
for (var i = 0; i < 5; i++) {
    idea = mturk.prompt(
        "What's fun to see in New York City?
         Ideas so far: " + ideas.join(", "))
    ideas.push(idea)
}

ideas.sort(function (a, b) {
    v = mturk.vote("Which is better?", [a, b])
    return v == a ? -1 : 1
})
```

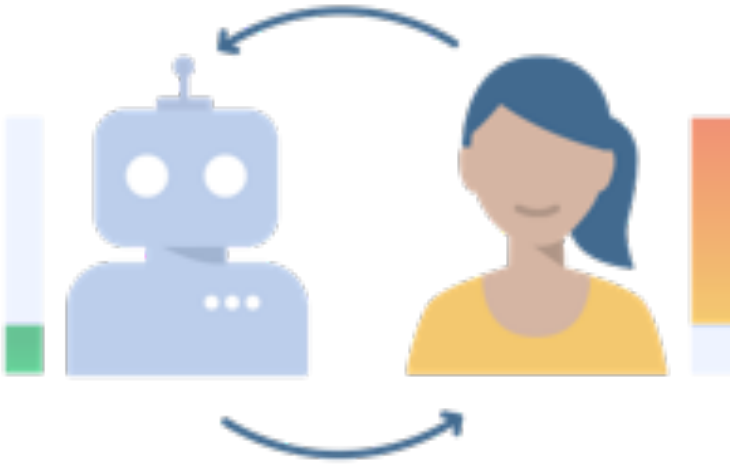
New way to organize teamwork



The focus of this course

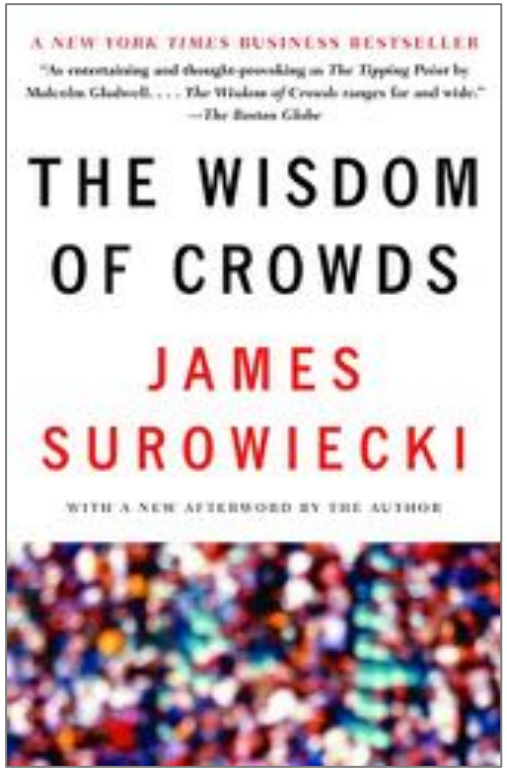
- Design and analysis of human-in-the-loop computation

- Discussion on **selected recent topics** **on human-AI interactions**



- Ethical decision making
- Fairness and privacy in AI/ML
- Interpretable machine learning
- Strategic machine learning
- Human-AI teaming

The Wisdom of Crowds



- At a 1906 country fair, ~800 people participate in a contest to guess the weight of an ox.
- Reward is given to the person with the closest guess.
- The average guess is 1,197lbs.
  The true answer is 1,198lbs.

It’s hard to bring together a large crowd in the old times.

Time Person of the Year 2006

>4 Billion Internet users worldwide as of 2018!

21% of Americans go online almost constantly!

The Internet Age and the Rise of Crowdsourcing

Connect

Create

Share

Collaborate

Let’s look at a few more applications

Games with a Purpose

ESP Game: Crowdsource image labeling via games

boy
white
black
window
wall
…

playground
children
running
boy
…

Potential Label: Boy

von Ahn and Dabbish. Labeling Images with a Computer Game. CHI'04

Games with a Purpose

Peekaboom: Crowdsource image segmentation via games

von Ahn et al. Peekaboom: A game for locating objects in images. CHI'06

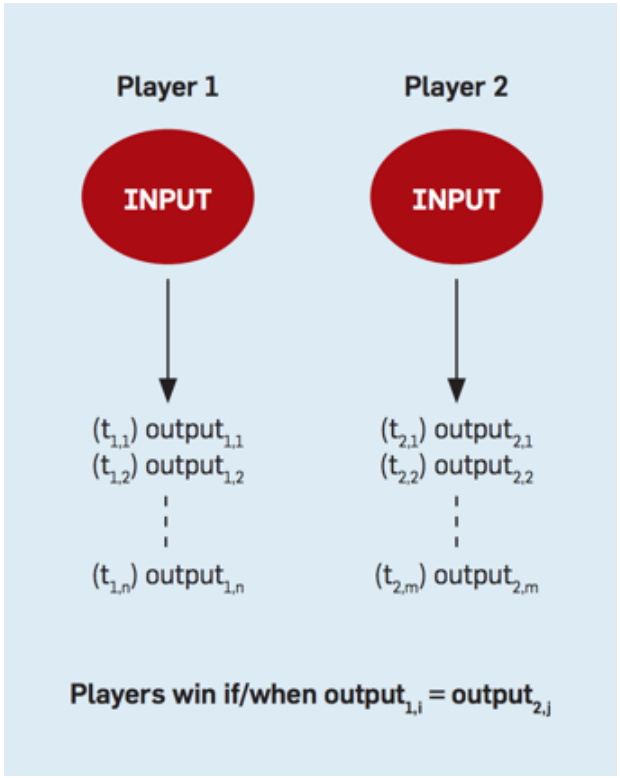
Games with a Purpose

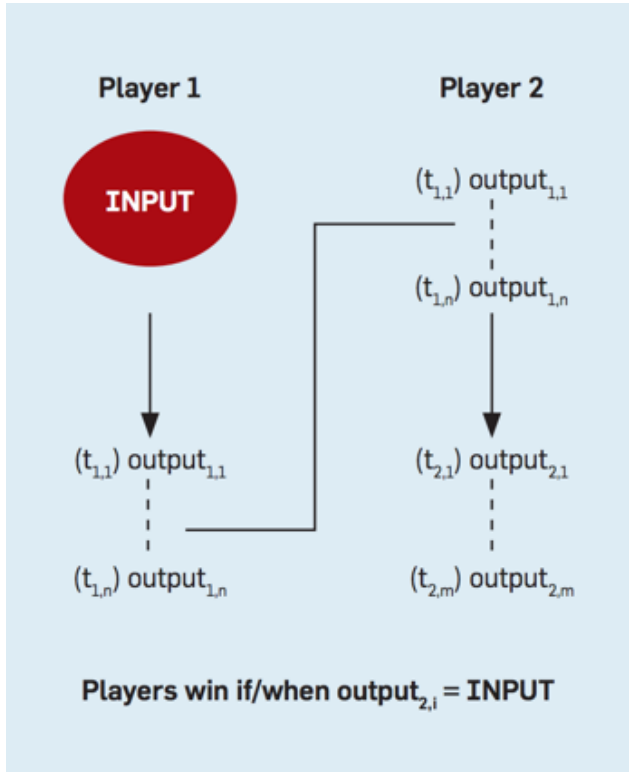
TagATune: Crowdsource audio annotation via games

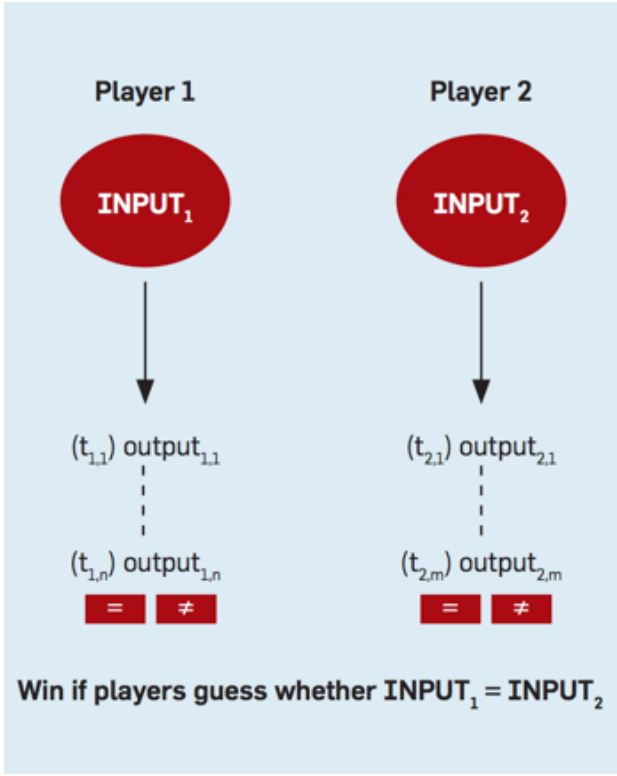
Law et al. TagATune: A Game for Music and Sound Annotation. ISMIR'07

Design and Analysis of Game with a Purpose?

How to ensure the data quality is high?







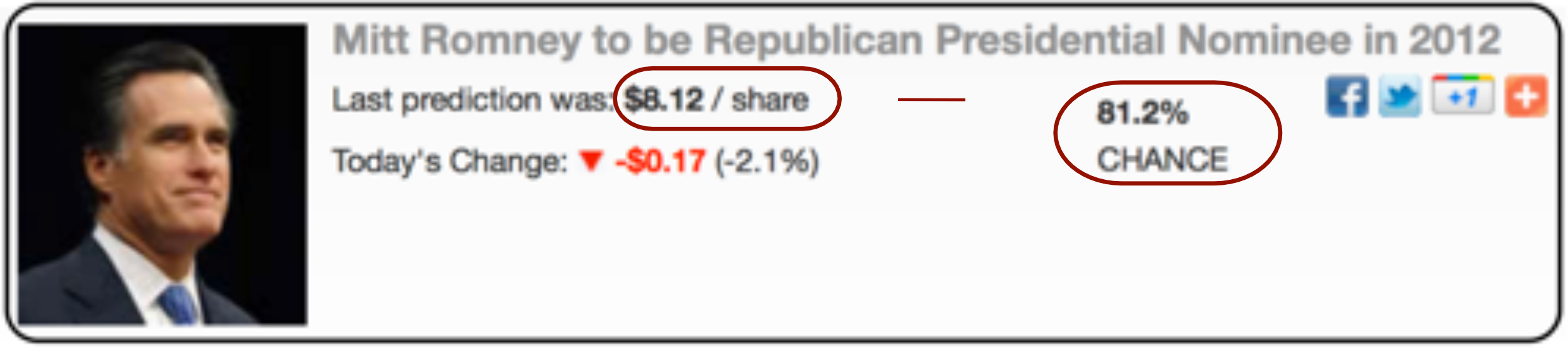
Output Agreement

Inversion Problem

Input Agreement

von Ahn and Dabbish. Designing Games with a Purpose. Comm of ACM, August 2008

Prediction Markets



- Potential payoff is **$10**.
- If I think the probability of Romney winning is ***p***, I should
  - **Buy** this security at any prices **less than $10 p**
  - **Sell** this security at any prices **greater than $10 p**

Price measures the population's collective beliefs

Does the price converge to the collective beliefs?

- Imagine the scenario:
  - Alice
    - believes the event won't happen (p=0)
    - very stubborn
    - very rich with unlimited amount of money
  - Bob
    - believes the event will definitely happen (p=1)
    - very stubborn
    - very rich with unlimited amount of money
- The price will just fluctuate between the two extremes

So how good are the predictions?

- In practice…
  - Election markets beat data from polls (BR02)
  - Oscar markets beat expert columnists (PGN01)
  - HP internal markets beat sales forecasts (P00)
  - Racetrack odds beat experts (F79)
  - and many more

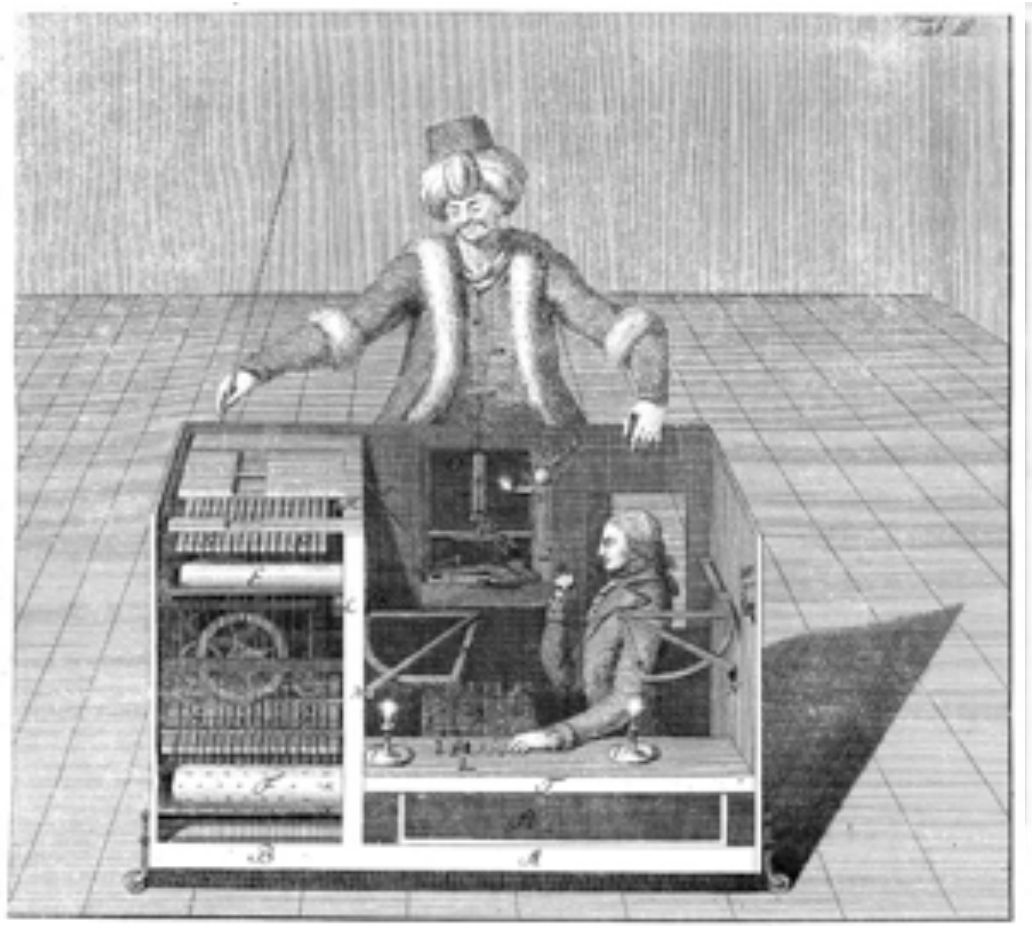
- In theory…
  - Under certain assumptions, prices converge and reflect the traders' collective knowledge
  - Proposing human behavior models to explain the real-world behavior

Crowdsourcing Markets

- The Turk – Automatic Chess Player built in the 18th century

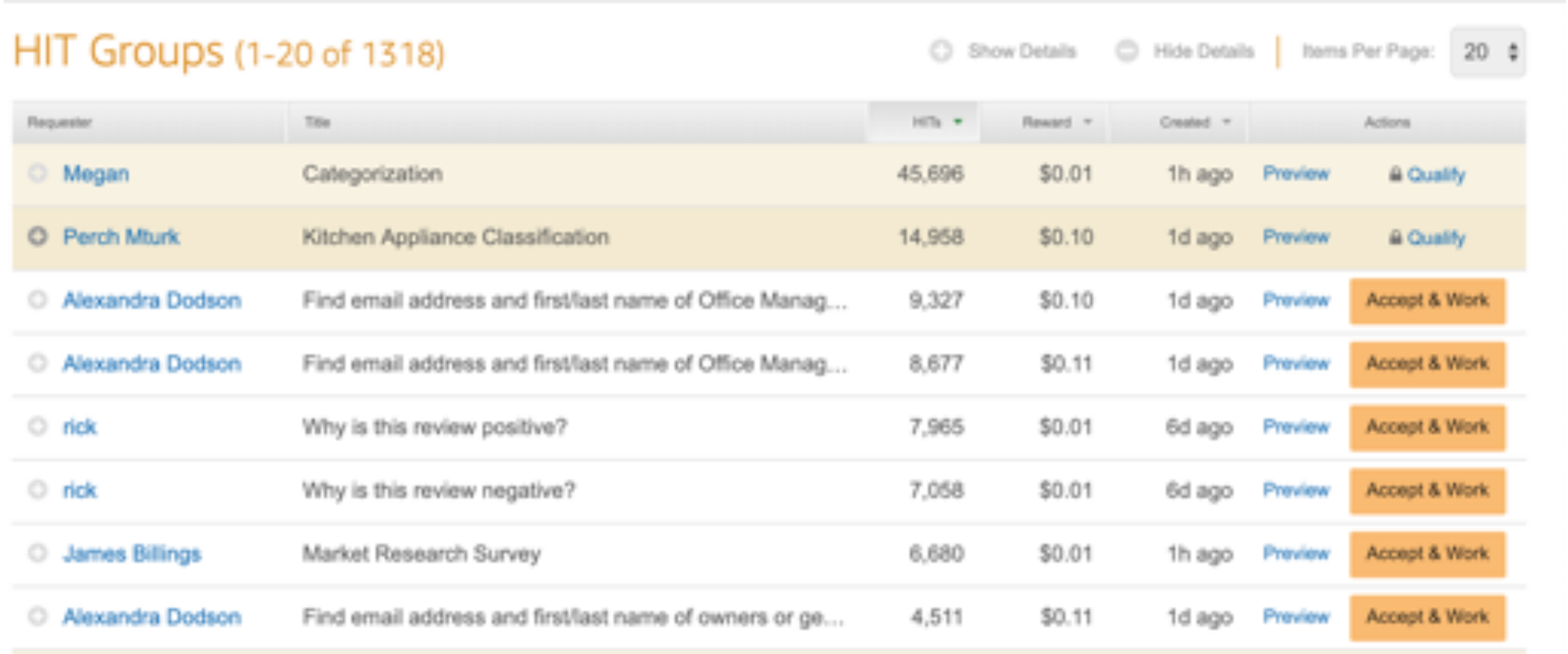
Crowdsourcing Markets

- The Turk – Automatic Chess Player built in the 18th century



Crowdsourcing Markets

- Amazon Mechanical Turk - Artificial Artificial Intelligence



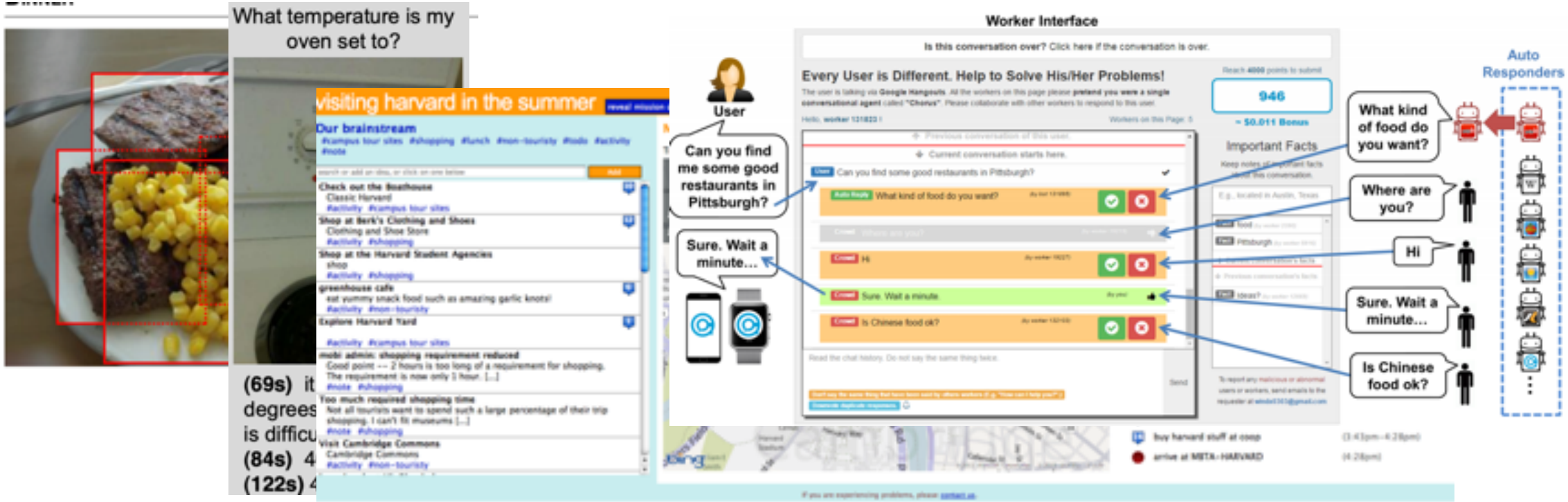
HIT Groups (1-20 of 1318)

Show Details | Hide Details | Items Per Page: 20

Requester	Title	HITs	Reward	Created	Actions
Megan	Categorization	45,696	$0.01	1h ago	Preview Qualify
Perch Mturk	Kitchen Appliance Classification	14,958	$0.10	1d ago	Preview Qualify
Alexandra Dodson	Find email address and first/last name of Office Manag...	9,327	$0.10	1d ago	Preview Accept & Work
Alexandra Dodson	Find email address and first/last name of Office Manag...	8,677	$0.11	1d ago	Preview Accept & Work
rick	Why is this review positive?	7,965	$0.01	6d ago	Preview Accept & Work
rick	Why is this review negative?	7,058	$0.01	6d ago	Preview Accept & Work
James Billings	Market Research Survey	6,680	$0.01	1h ago	Preview Accept & Work
Alexandra Dodson	Find email address and first/last name of owners or ge...	4,511	$0.11	1d ago	Preview Accept & Work

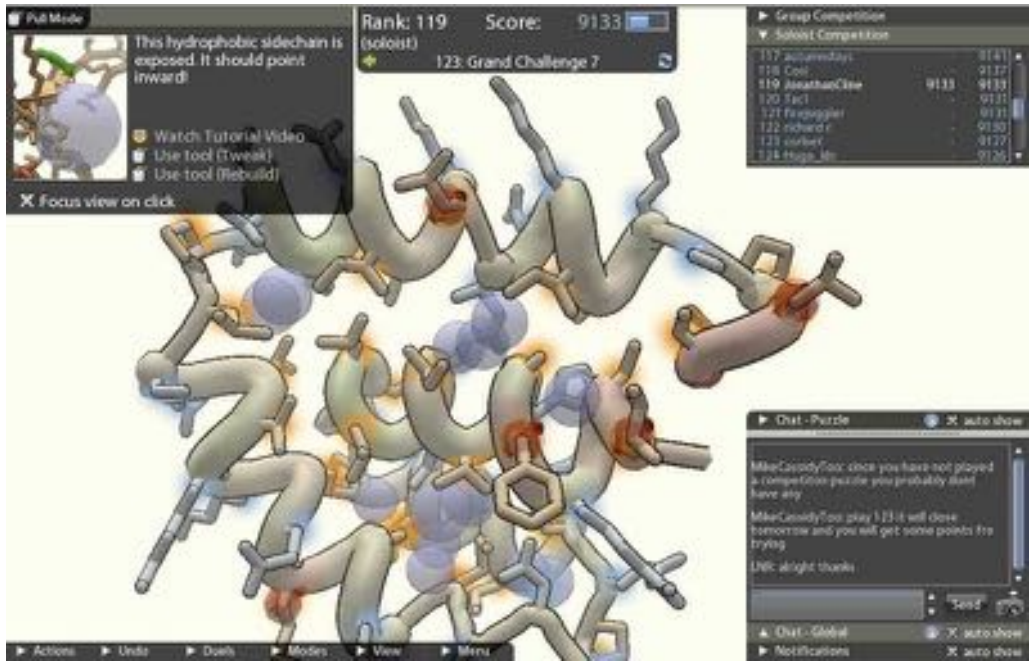
Crowdsourcing Markets

- A general-purpose platform for all kinds of applications
  - Mainly used to engage a crowd, and you can build you applications on top of it
  - In addition, it's easier to design "payment" than to design "fun"



There are more….

Citizen Science Projects

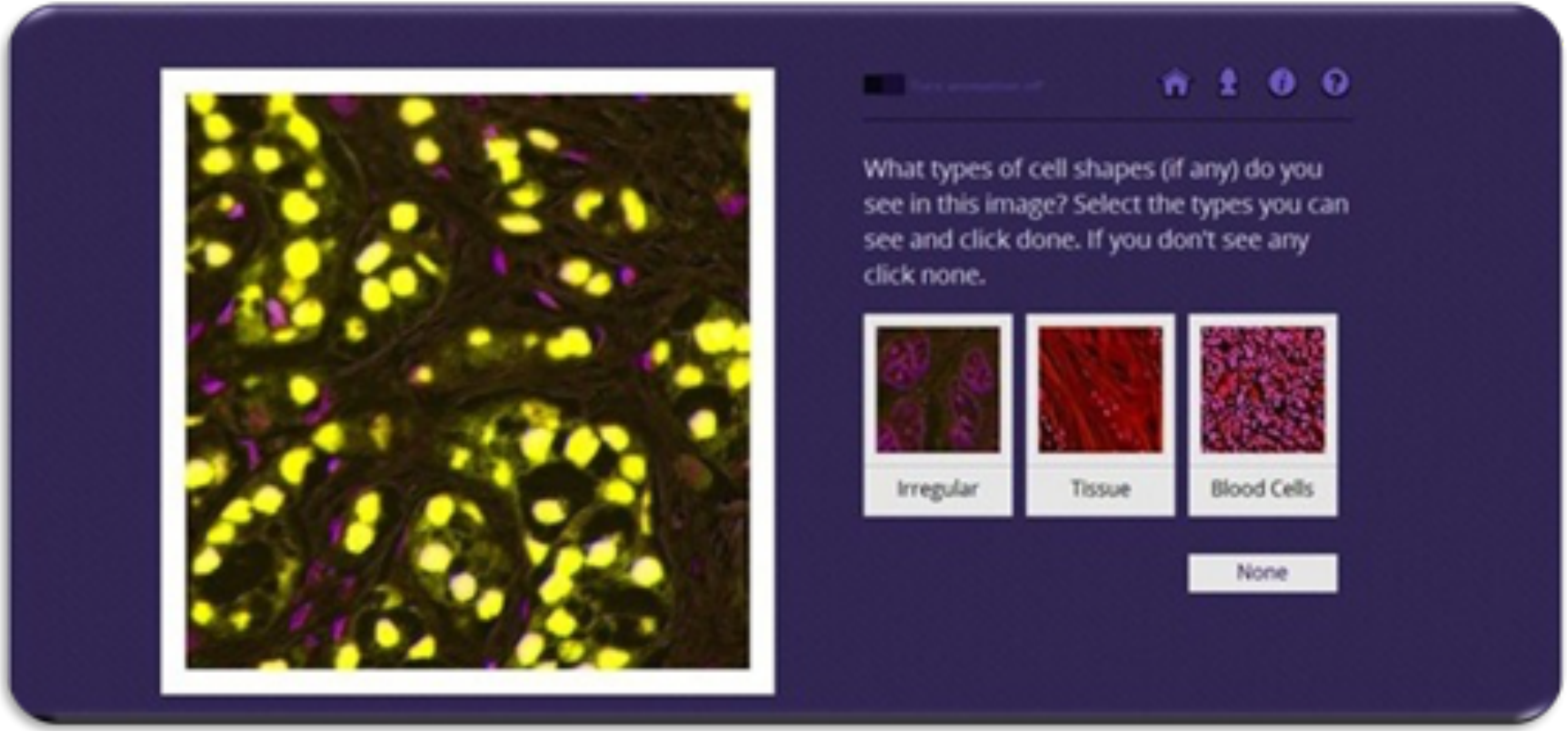


FoldIt: Online video game for protein folding

2011: decipher the crystal structure of an AIDS-related virus
2012: The first crowdsourced redesign of a protein

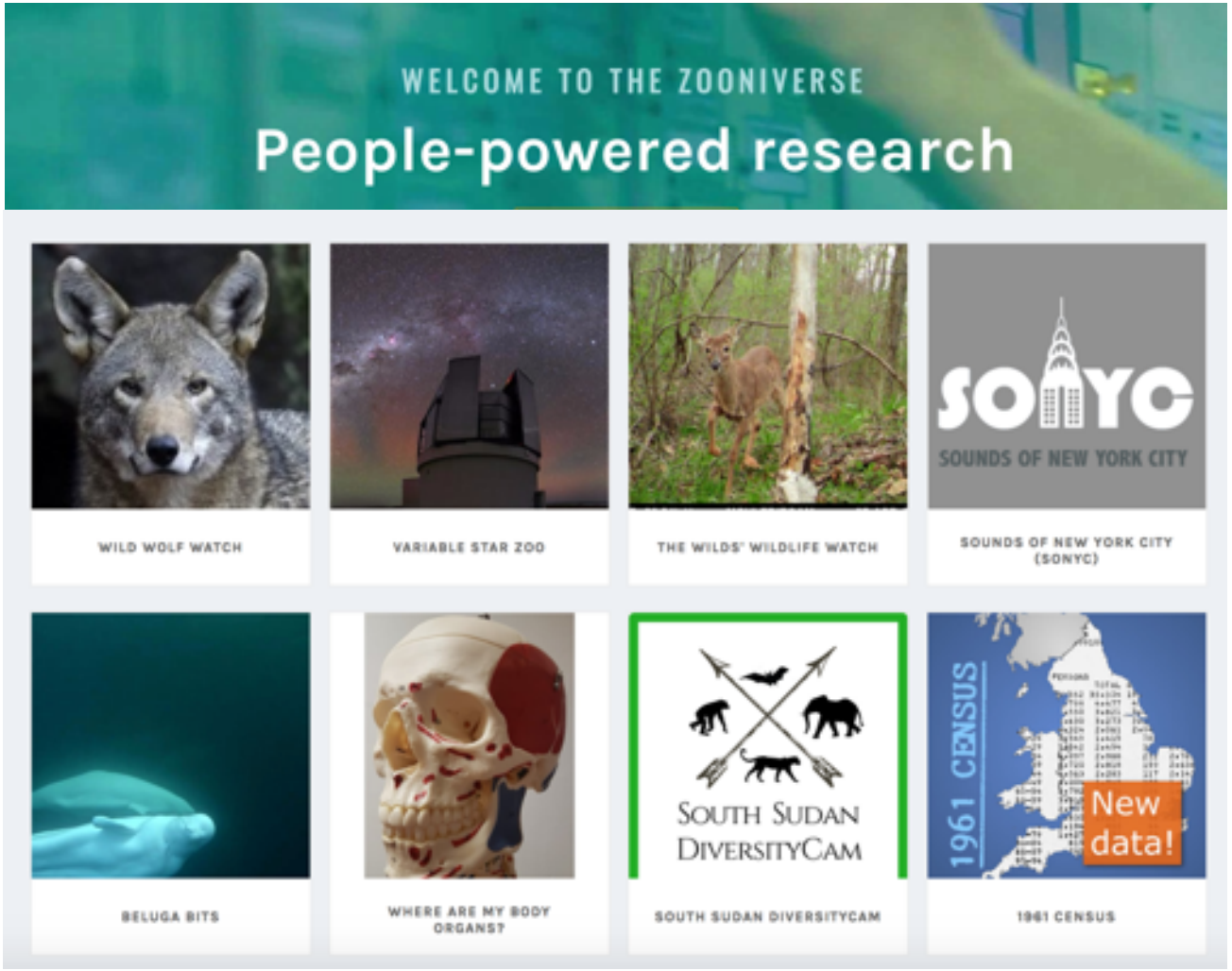
Cooper et al. Predicting Protein Structures with a Multiplayer Online Game. Nature, August 2010

Citizen Science Projects



Cell Slider: Examine tumor tissue samples and spotting cancerous cells

Citizen Science Projects

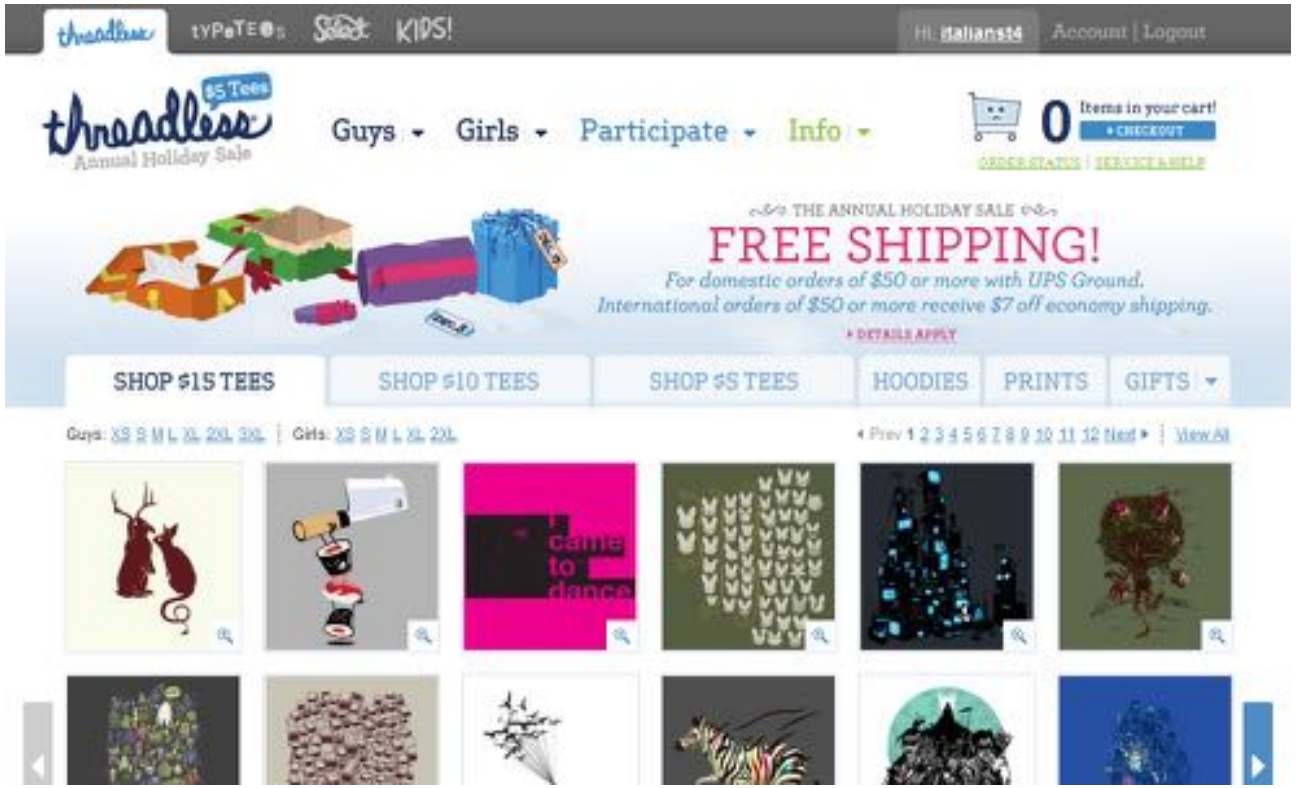


Zooniverse: A citizen science web portal

Crowdsourcing Contests

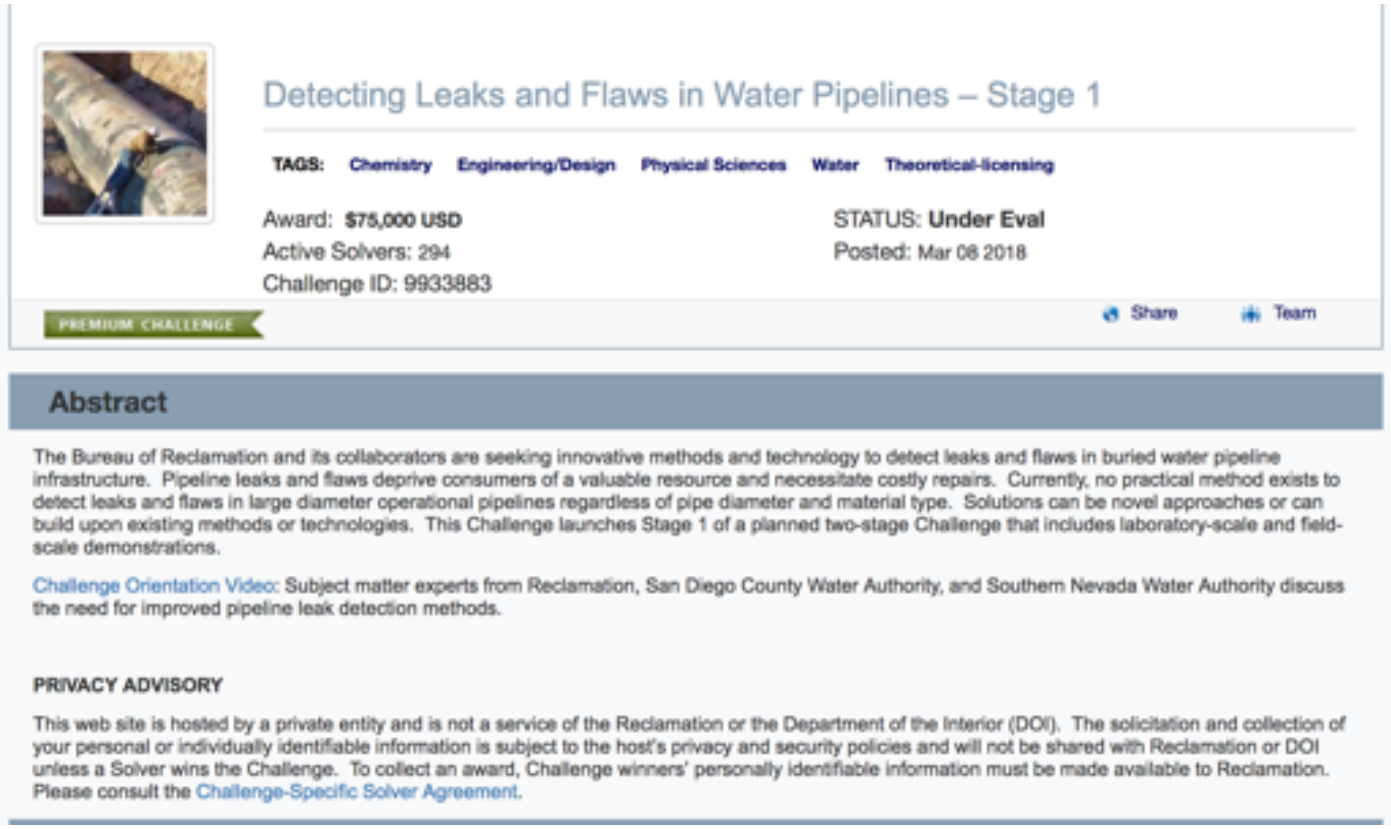
- Netflix Prize, 2006
- 44,014 submissions
- The best solution improve its recommendation engine's performance by 10%
- crowdsourcing within a crowdsourcing competition

Crowdsourcing Contests



- Threadless
- Crowds generate the design
- Crowds vote on the design
- 438,451 designs submitted, 8,805 printed

Crowdsourcing Contests



Detecting Leaks and Flaws in Water Pipelines – Stage 1

TAGS: Chemistry Engineering/Design Physical Sciences Water Theoretical-licensing

Award: $75,000 USD
Active Solvers: 294
Challenge ID: 9933883

STATUS: Under Eval
Posted: Mar 06 2018

PREMIUM CHALLENGE

Share Team

Abstract

The Bureau of Reclamation and its collaborators are seeking innovative methods and technology to detect leaks and flaws in buried water pipeline infrastructure. Pipeline leaks and flaws deprive consumers of a valuable resource and necessitate costly repairs. Currently, no practical method exists to detect leaks and flaws in large diameter operational pipelines regardless of pipe diameter and material type. Solutions can be novel approaches or can build upon existing methods or technologies. This Challenge launches Stage 1 of a planned two-stage Challenge that includes laboratory-scale and field-scale demonstrations.

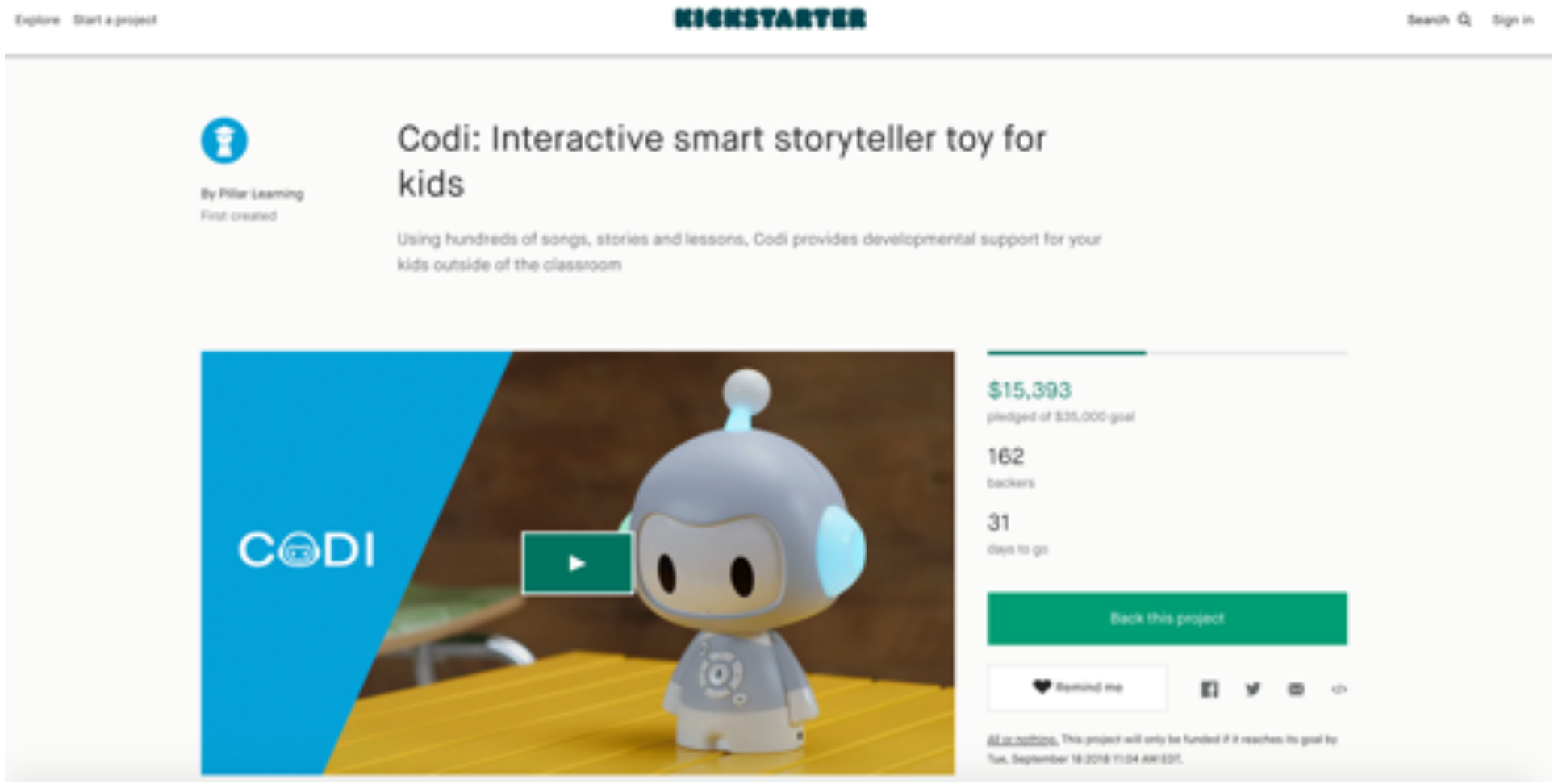
Challenge Orientation Video: Subject matter experts from Reclamation, San Diego County Water Authority, and Southern Nevada Water Authority discuss the need for improved pipeline leak detection methods.

PRIVACY ADVISORY

This web site is hosted by a private entity and is not a service of the Reclamation or the Department of the Interior (DOI). The solicitation and collection of your personal or individually identifiable information is subject to the host's privacy and security policies and will not be shared with Reclamation or DOI unless a Solver wins the Challenge. To collect an award, Challenge winners' personally identifiable information must be made available to Reclamation. Please consult the Challenge-Specific Solver Agreement.

- InnoCentive
- Crowdsourcing problem-solving
- 2000+ problems, 380,000+ solvers
- Problems are supplied by corporations, governments, non-profits, etc.

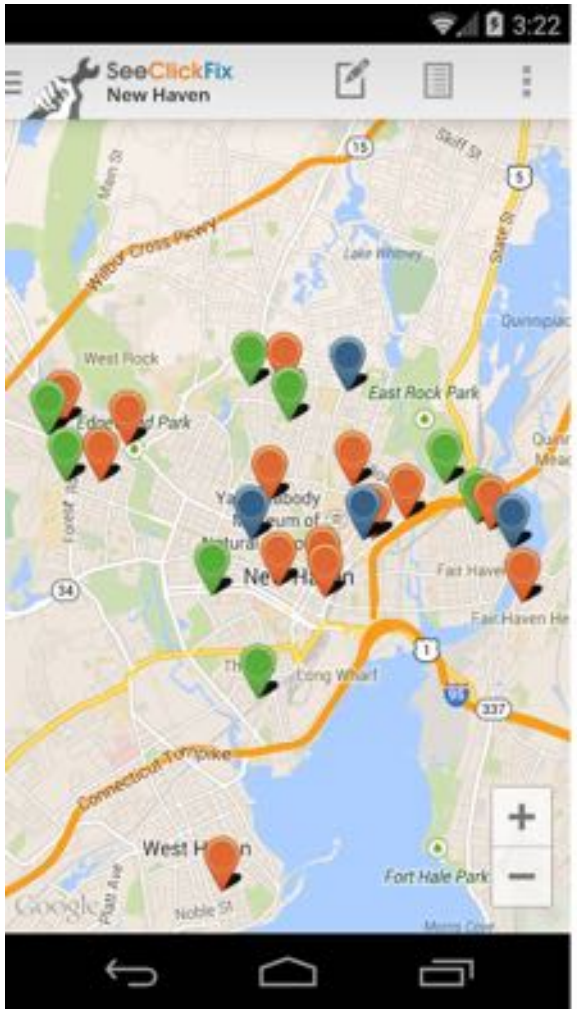
Crowdfunding

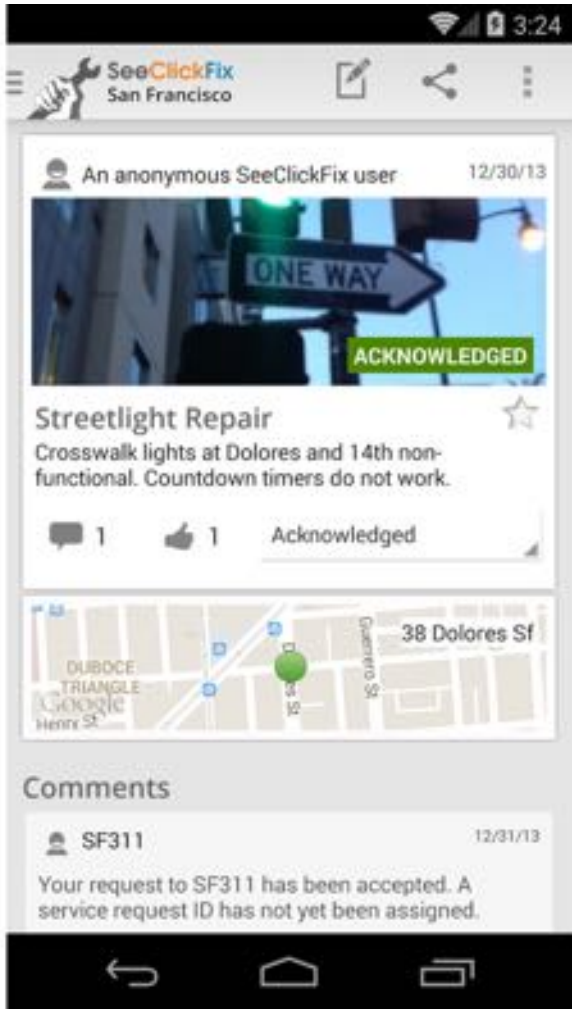


Crowdsensing: Waze

Crowdsourcing: SeeClickFix

SeeClickFix
power to the community





Discussion

- Can you think of some other examples of crowdsourcing applications?

- What are some examples of challenging tasks in your daily life? Can you utilize the idea of crowdsourcing to help with your challenging tasks? Why or why not?