KAIST Fall 2016

CS492: Crowdsourcing

http://kixlab.org/courses/crowdsourcing/

Class 1:
Introduction & Course Overview

2016.9.1

Juho Kim

Who am I? Prof. Juho Kim

- Assistant Professor in SoC
- *1st semester & 1st course at KAIST!*

- Research Interests: HCI, crowdsourcing, learning at scale, civic engagement
- Ph.D. from MIT CSAIL

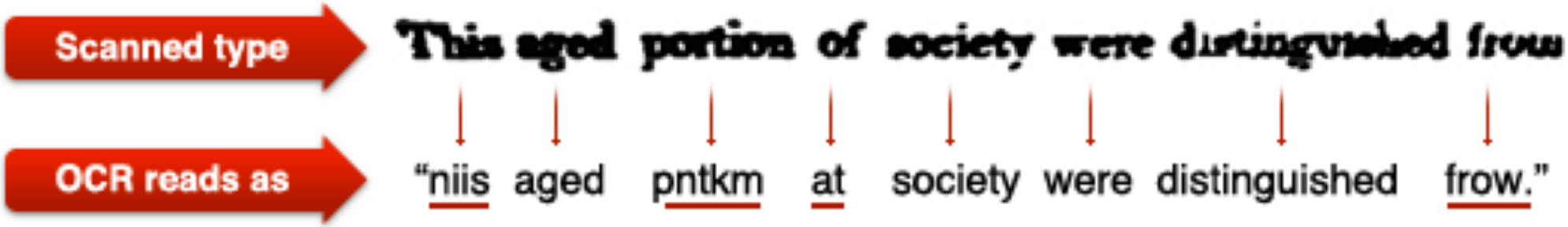
- KIXLAB (KAIST Interaction Lab): kixlab.org
- juhokim.com

TA: Seunghyun Han

- 1st year Master's student, SoC

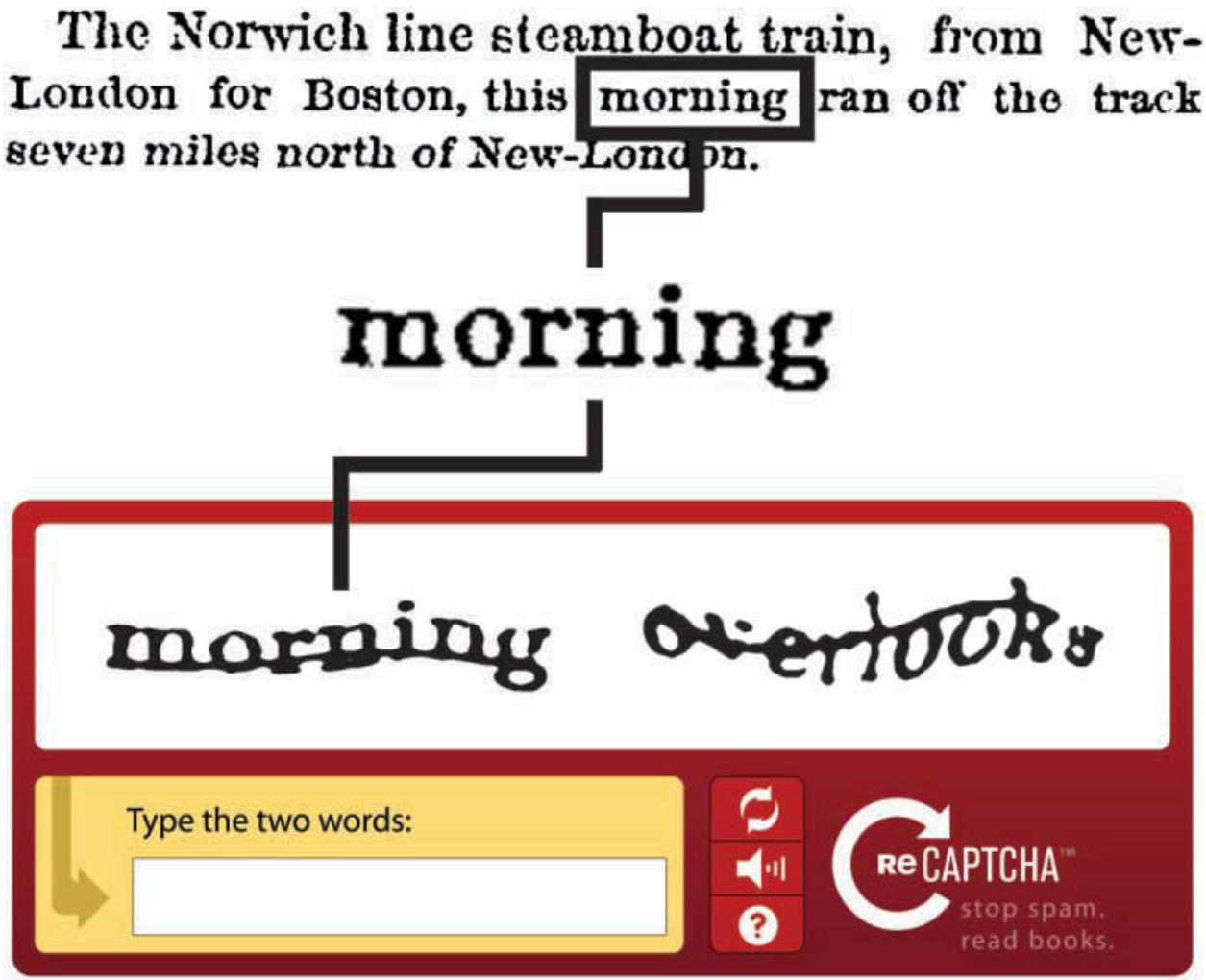
Type the two words:
reCAPTCHA™
stop spam.
read books.

This aged portion of society were distinguished from



~30% error rate

Human brain as OCR?



reCAPTCHA: Human-Based Character Recognition via Web Security Measures.
Luis von Ahn, Ben Maurer, Colin McMillen, David Abraham and Manuel Blum. Science, September 12, 2008. pp 1465-1468.

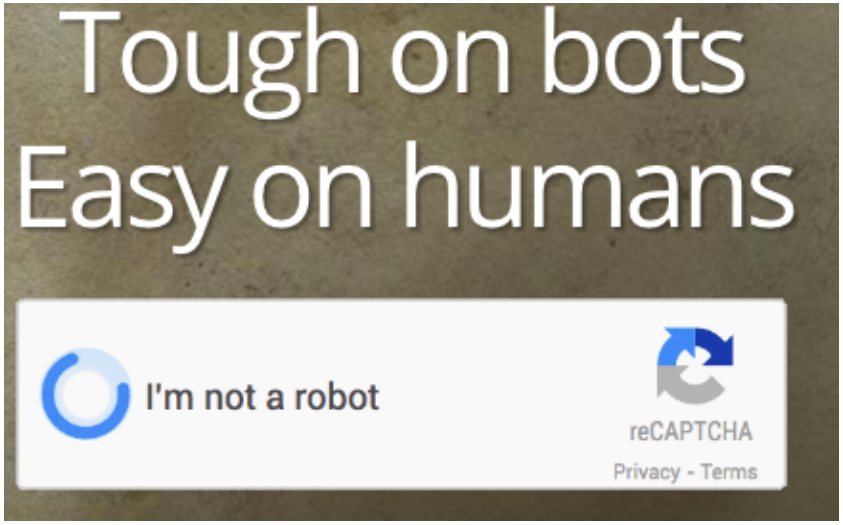
350K websites (FB, Google...)

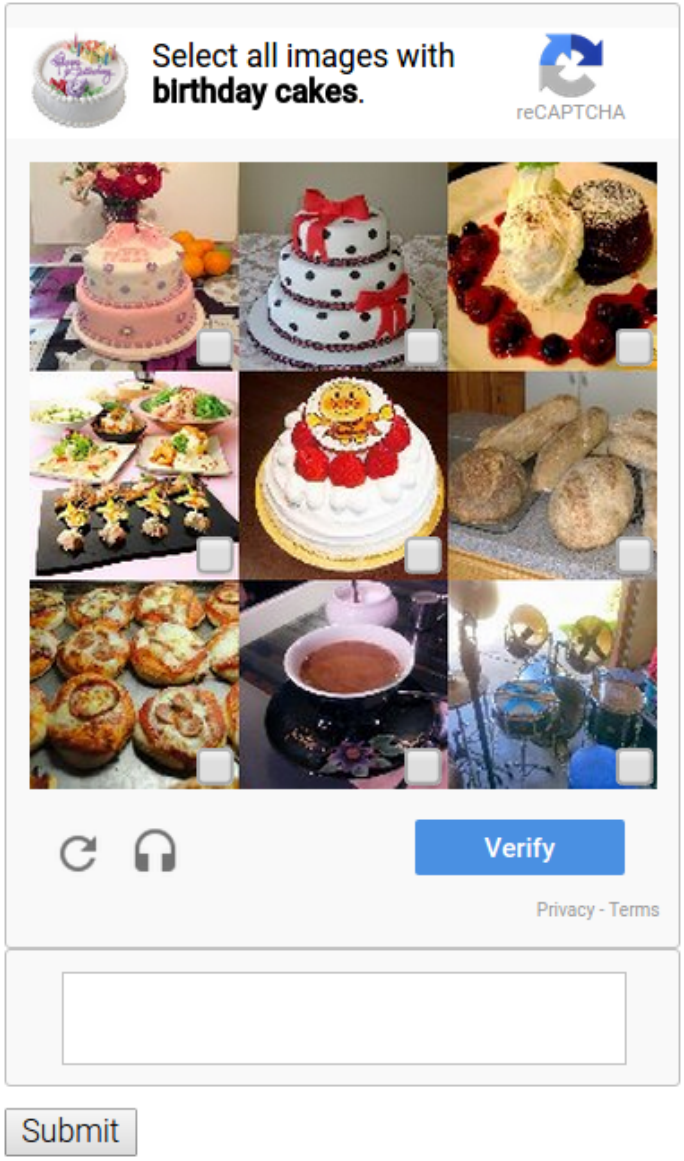
100M words / day

2.5M books / year

750M people (>10% of humanity)

reCAPTCHA version 2





reCAPTCHA

Human	Computer
human verification	digitize text

Google

golden gate bridge

Search

SafeSearch moderate ▼

About 2,550,000 results (0.08 seconds)

Advanced search

- Everything
- Images
- Videos
- News
- Shopping
- More

Sort by relevance
Sort by subject

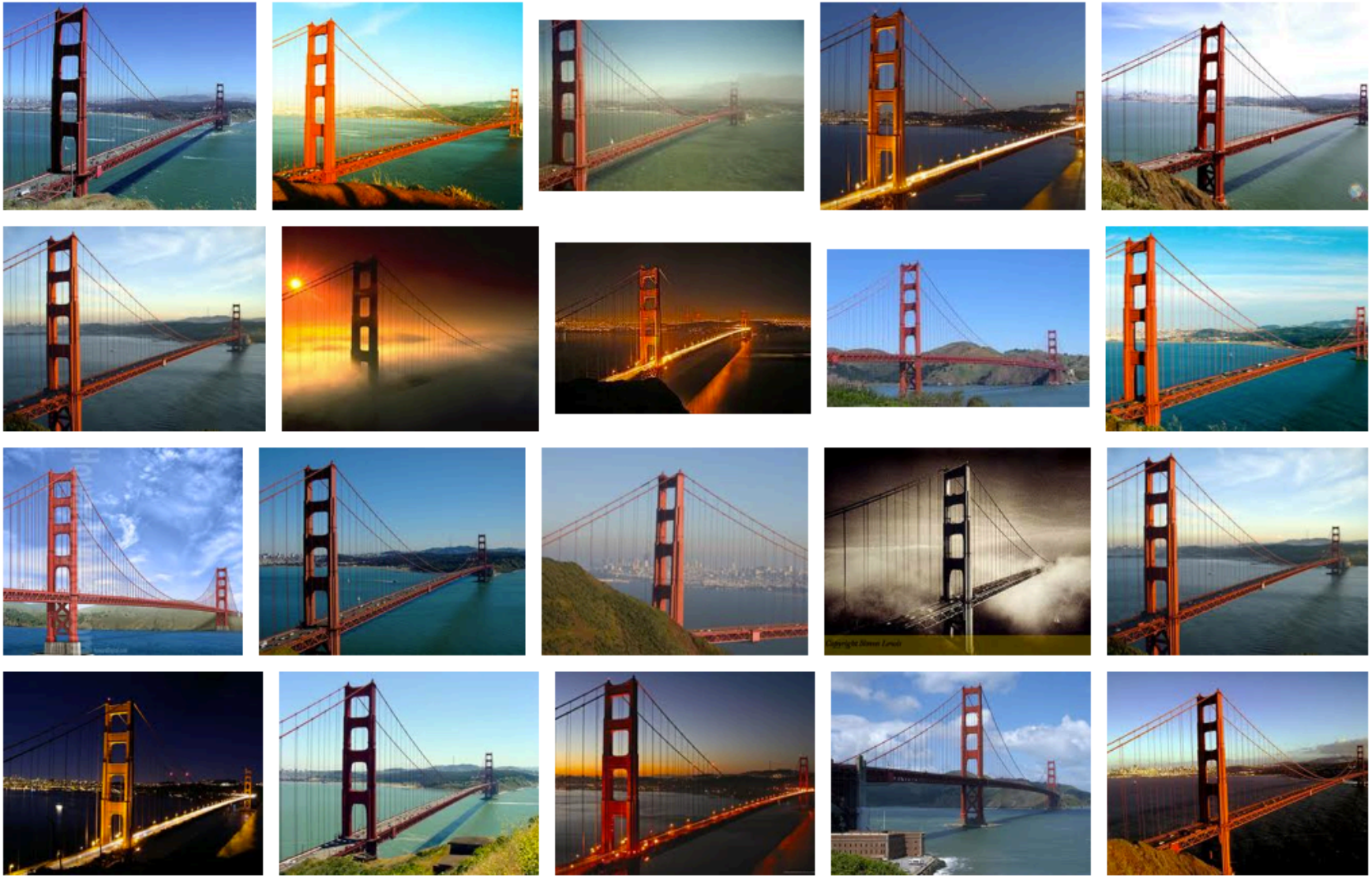
Any size
Large
Medium
Icon
Larger than...
Exactly...

Any color
Full color
Black and white

Any type
Face
Photo
Clip art
Line drawing

Standard view
Show sizes

Related searches: golden gate bridge **at night** golden gate bridge **wallpaper** golden gate bridge **cartoon** golden gate bridge **black and white** golden gate bridge **sunset**



gwap | ESP Game | Tag a Tune | Verbosity | Squigl | Matchin | FlipIt | PopVideo

Sign In ☑ remember me forgot password?

We have partially resumed the operations, this message will disappear when all services are online.

Play the Games, Change the Web.

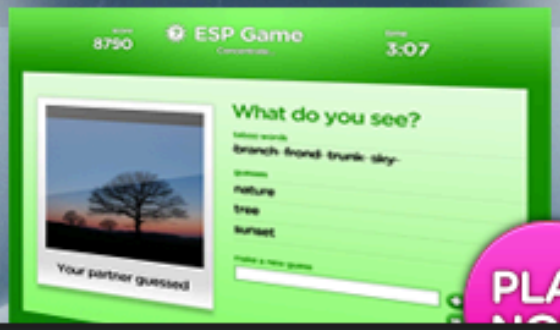
When you play a game at Gwap, you aren't just having fun.

Learn More Register

ESP Game

Concentrate...

What do you see in an image? Do you have ESP?



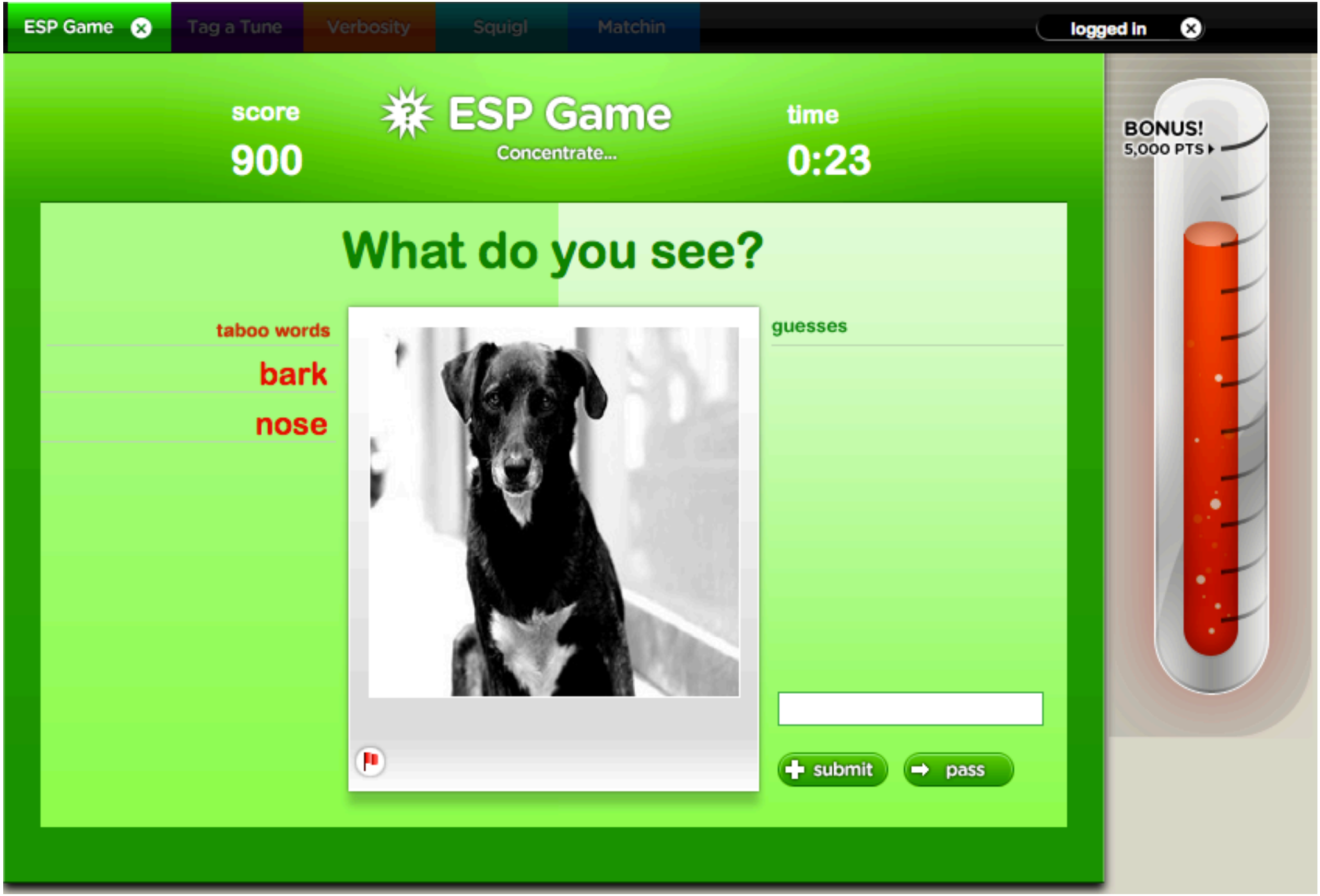
PLAY NOW

Gender Guesser | Prizes | Best Images

Today's All Time Top 10

6	welloff	190 K
7	PlasticBiddy	178 K
8	fo jo	149 K
9	Lottie	145 K
10	fling	128 K

Blog | About | Contact



Luis von Ahn & Laura Dabbish. Labeling images with a computer game. *CHI '04*

ESP Game

Human	Computer
play mind-guessing game	collect image tags

Easy for humans, hard for computers

Large scale & Small contributions

Solve novel problems

AI: computers imitate humans

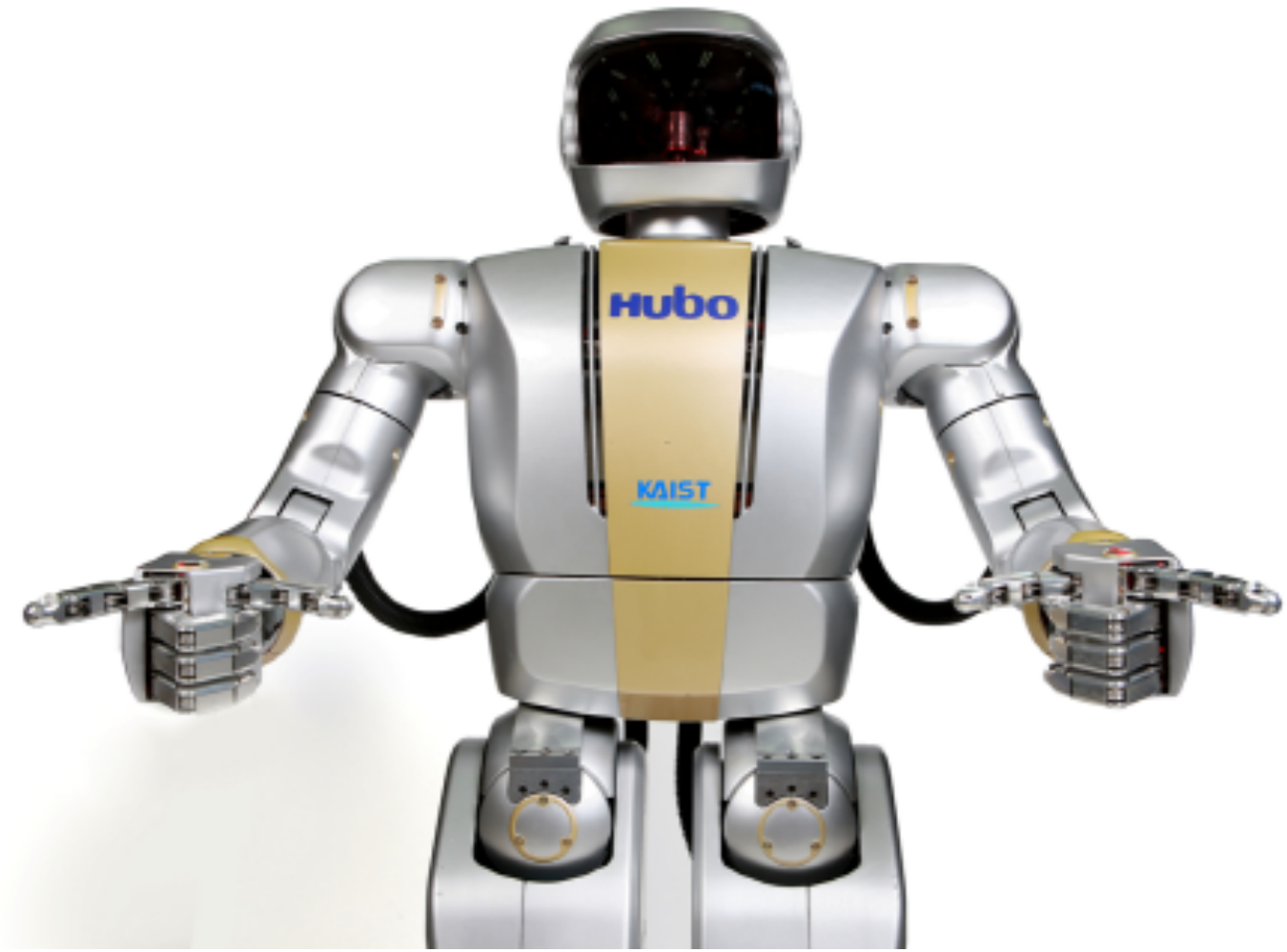


photo credit: KAIST Hubo Lab

Human Computation: humans help computers

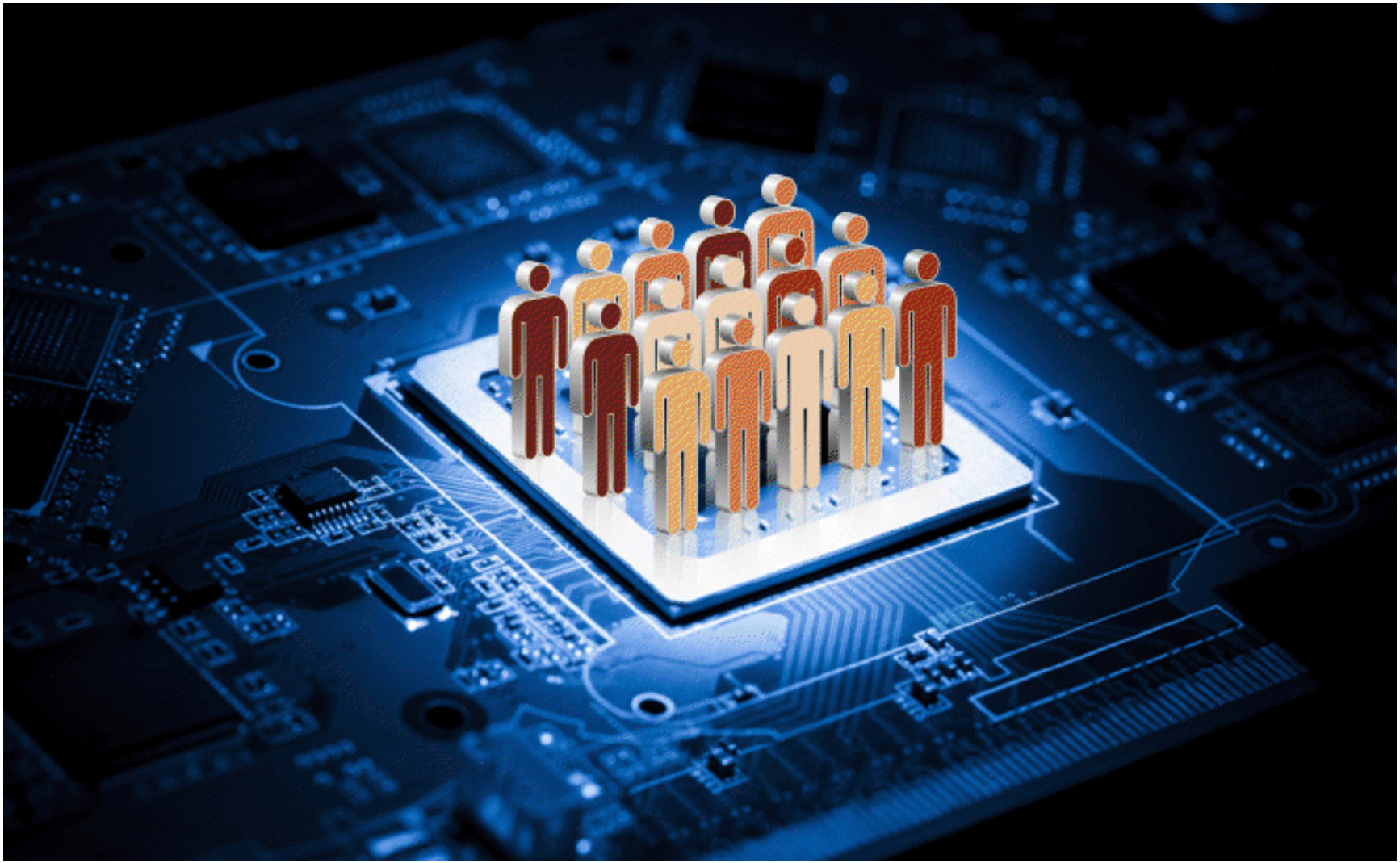


Image courtesy of Christine Daniloff at MIT

Can we generalize?

Crowdsourcing Marketplace

amazonmechanical turk
Artificial Artificial Intelligence

Your Account | HITs | Qualifications

Already have an account?
Sign in as a Worker | Requester

Introduction | **Dashboard** | **Status** | **Account Settings**

Mechanical Turk is a marketplace for work.
We give businesses and developers access to an on-demand, scalable workforce.
Workers select from thousands of tasks and work whenever it's convenient.

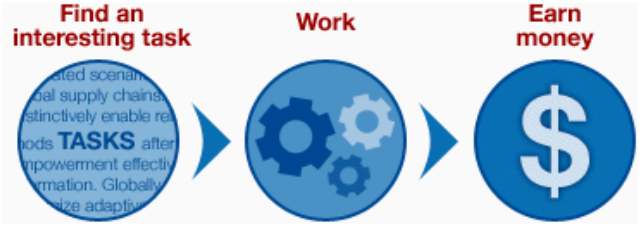
910,392 HITs available. View them now.

Make Money
by working on HITs

HITs - *Human Intelligence Tasks* - are individual tasks that you work on. Find HITs now.

As a Mechanical Turk Worker you:

- Can work from home
- Choose your own work hours
- Get paid for doing good work



Find HITs Now

or learn more about being a **Worker**

Get Results
from Mechanical Turk Workers

Ask workers to complete HITs - *Human Intelligence Tasks* - and get results using Mechanical Turk. Get Started.

As a Mechanical Turk Requester you:

- Have access to a global, on-demand, 24 x 7 workforce
- Get thousands of HITs completed in minutes
- Pay only when you're satisfied with the results

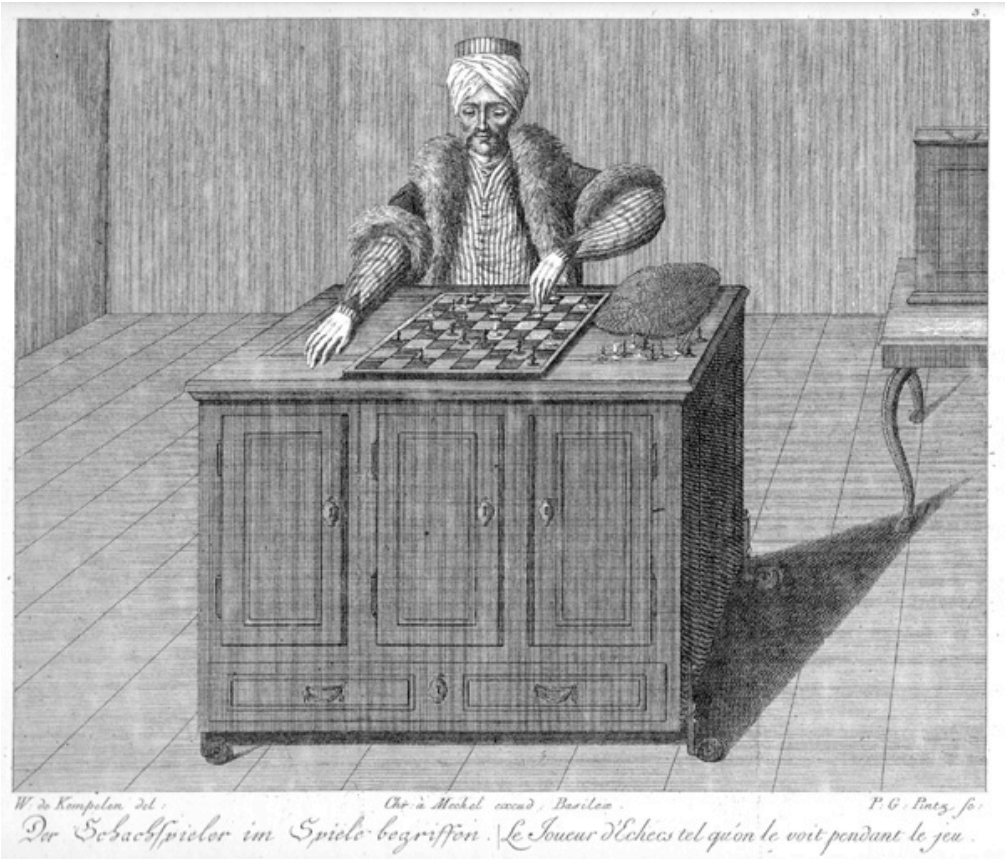
Get Started

FAQ | Contact Us | Careers at Mechanical Turk | Developers | Press | Policies | State Licensing | Blog | Service Health Dashboard

©2005-2016 Amazon.com, Inc. or its Affiliates

An amazon.com company

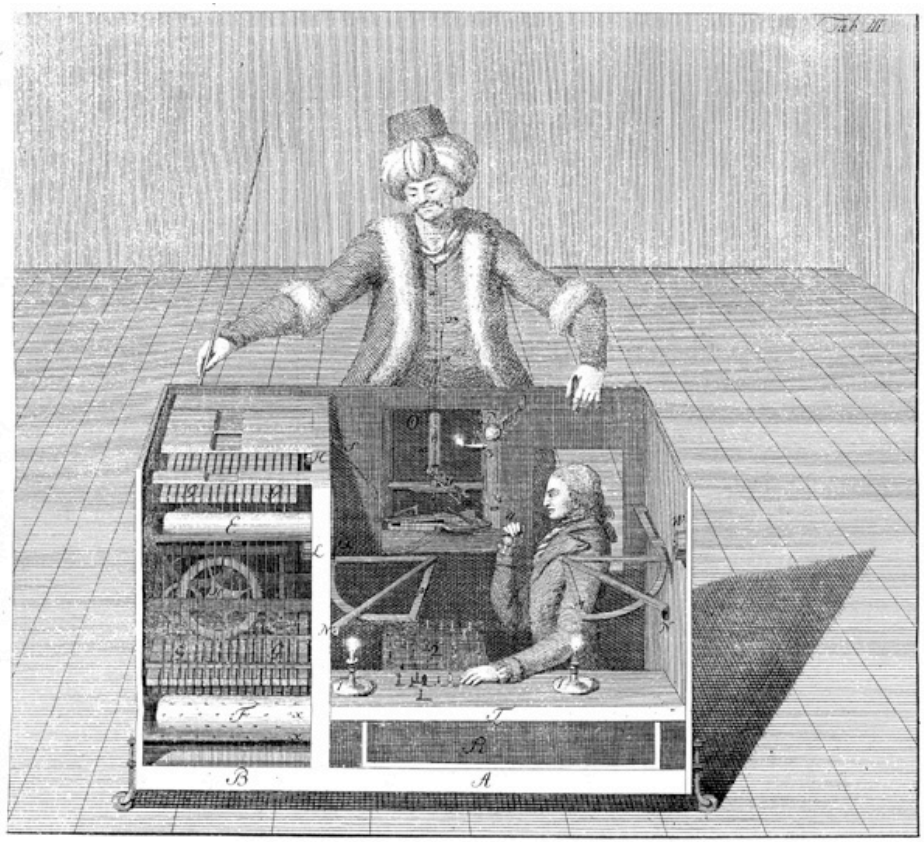
Artificial Intelligence



Copper engraving from the book: Karl Gottlieb von Windisch, Briefe über den Schachspieler des Hrn. von Kempelen, nebst drei Kupferstichen die diese berühmte Maschine vorstellen. 1783.

“Artificial” Artificial Intelligence

amazonmechanical turk
beta
Artificial Artificial Intelligence



Copper engraving from the book: Freiherr Joseph Friedrich zu Racknitz,
Ueber den Schachspieler des Herrn von Kempelen, Leipzig und Dresden 1789.

Scene Attribute Labeling					View a HIT in this group
Requester:	James Hays Lab	**HIT Expiration Date:**	Jun 19, 2011 (6 days 19 hours)	**Reward:**	$0.03
		Time Allotted:	15 minutes	**HITs Available:**	5274
NRS V2 FR Electronics Top-Fri Jun 10 20:44:54 PDT 2011					View a HIT in this group
Requester:	Amazon Requester Inc.	**HIT Expiration Date:**	Jul 10, 2011 (4 weeks)	**Reward:**	$0.01
		Time Allotted:	60 minutes	**HITs Available:**	5160
NRS V2 DE Electronics Bottom-Fri Jun 10 22:51:39 PDT 2011					View a HIT in this group
Requester:	Amazon Requester Inc.	**HIT Expiration Date:**	Jul 11, 2011 (4 weeks)	**Reward:**	$0.01
		Time Allotted:	60 minutes	**HITs Available:**	3952
Image Annotation					View a HIT in this group
Requester:	Liyue Zhao	**HIT Expiration Date:**	Jun 14, 2011 (1 day 22 hours)	**Reward:**	$0.01
		Time Allotted:	10 minutes	**HITs Available:**	3642
Copy text from scanned image (Handwritten)					View a HIT in this group
Requester:	HAVI Digital	**HIT Expiration Date:**	Jun 25, 2011 (1 week 6 days)	**Reward:**	$0.18
		Time Allotted:	2 hours	**HITs Available:**	3170
NRS V2 IT Electronics Top-Fri Jun 10 23:30:07 PDT 2011					View a HIT in this group
Requester:	Amazon Requester Inc.	**HIT Expiration Date:**	Jul 10, 2011 (4 weeks)	**Reward:**	$0.01
		Time Allotted:	60 minutes	**HITs Available:**	3062
Copy text from scanned image (Caption)					View a HIT in this group
Requester:	HAVI Digital	**HIT Expiration Date:**	Jun 25, 2011 (1 week 6 days)	**Reward:**	$0.20
		Time Allotted:	2 hours	**HITs Available:**	2768
Create listings for consumer products					View a HIT in this group
Requester:	Buyer's Best Friend	**HIT Expiration Date:**	Jun 19, 2011 (6 days 23 hours)	**Reward:**	$0.05
		Time Allotted:	60 minutes	**HITs Available:**	2759
Moderate Internet Images for Violence, Hate, and Sexual Content (contains mature content) Job 4					View a HIT in this group
Requester:	Dolores Labs	**HIT Expiration Date:**	Jun 19, 2011 (6 days 23 hours)	**Reward:**	$0.04
		Time Allotted:	60 minutes	**HITs Available:**	2038

Typical Tasks

- Photographs of a store-front
- Writing product descriptions
- Business listing verification
- Transcription

Tag 5 Images - Magnum Photos

Requester: Tagasauris **Reward:** $0.04 per HIT **HITs Available:** 1395 **Duration:** 60 minutes

Qualifications Required: HIT approval rate (%) is not less than 95, Location is US

Tagasauris
the smart way to tag!

Provide 3 tags for the image (Add 3 tags) » Show hit instructions

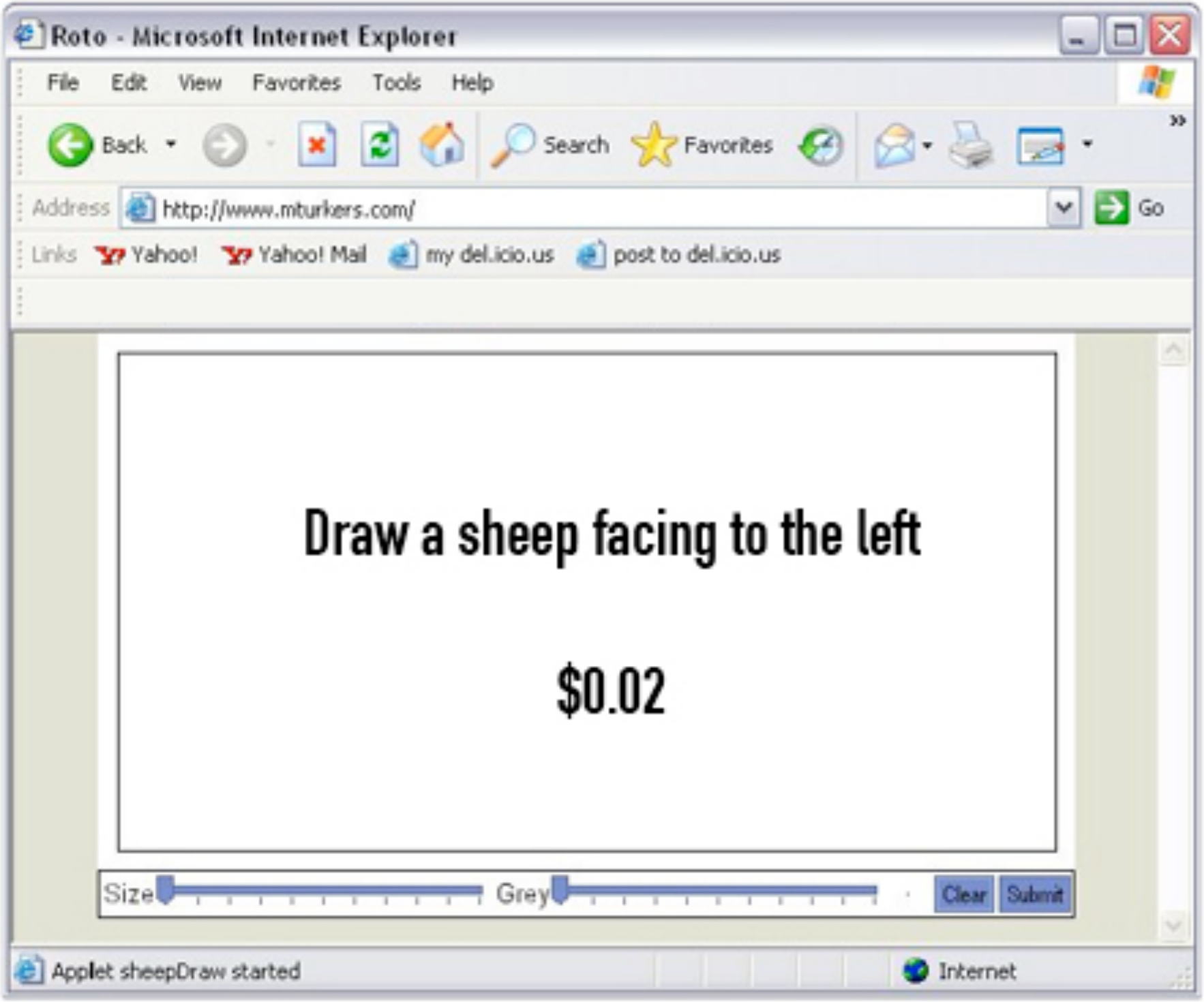
Image 1 of 5 » Send feedback

Preview mode please accept the hit

Please enter 3 tags, confirm with enter:

Want to work on this HIT? Accept HIT

Want to see other HITs? Skip HIT


Roto - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back
Search
Favorites
Address
http://www.mturkers.com/
Go
Links
Yahoo!
Yahoo! Mail
my del.icio.us
post to del.icio.us
Draw a sheep facing to the left
$0.02
Size
Grey
Clear
Submit
Applet sheepDraw started
Internet

The Sheep Market
by Aaron Koblin

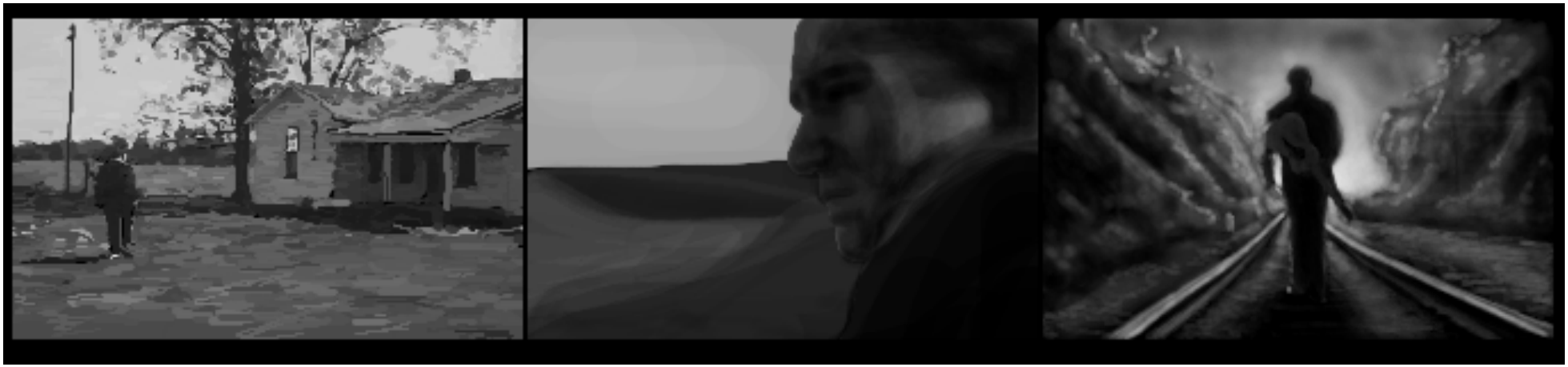
10,000 sheep

$0.02 per sheep

105 seconds / sheep

40 days

The Johnny Cash Project



Help the world become a better place.

photo credit: AFP

Ushahidi.com

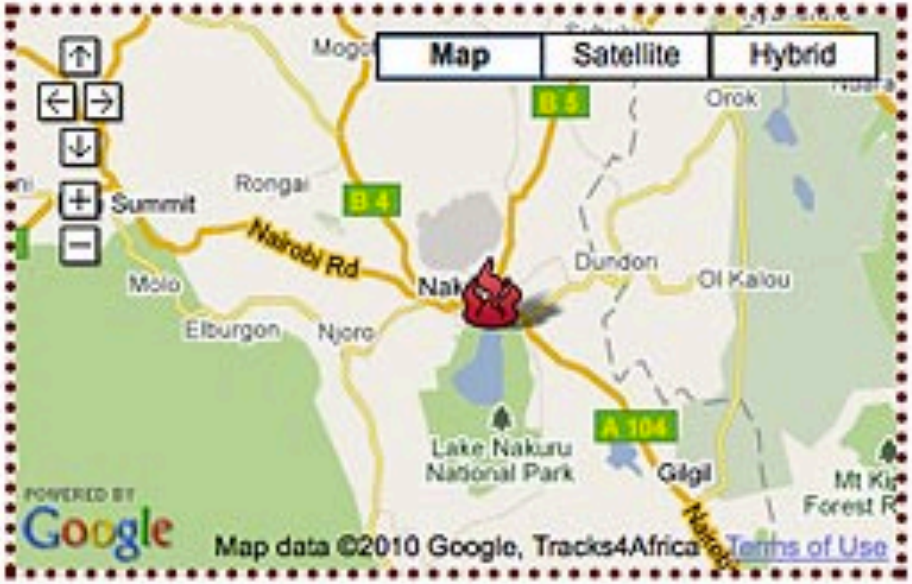
HOME | REPORT AN INCIDENT | CONTACT US | ABOUT | BLOG | HOW TO HELP

Death at trouble Kihingo, Deffo , Mauche in Njoro.

« Back | Submit An Incident Report »

LOCATION:	Kihingo, Deffo , Mauche in Njoro, Kenya
DATE:	Tuesday, March 11, 2008
TIME:	10:00 am (Approximate)
CATEGORY:	DEATHS, PROPERTY,
NAMES OF PEOPLE:	Kalinjin, Kikuyus,
VERIFIED?	✕

Add Information About This Incident »



Incident Report Description:

One person was slashed at mauche belonging to the Kikuyu tribe. The Kikuyu tribe in revenge burnt down acar from the kalenjin community at kihingo trading centre. In the past four days the kalenjins had killed 5mens from the kikuyu tribe at Likia trading centre and closed down road at mauche. After the car burning earley yesterday morning a few kikuyu houses was set on fire and yet to confirm the death toll yet. Heavy army officers mounted on the area.

Ushahidi.com

HOME | REPORT AN INCIDENT | CONTACT US | ABOUT | BLOG | HOW TO HELP

View A Timeline Of Events »

Map | Satellite | Hybrid

- [Feb 9, 2008] Naivasha youth is organizing a peac...
- [Feb 3, 2008] Using sport as antidote to violence...
- [Feb 1, 2008] Reaching out to the Mungiki...
- [Jan 31, 2008] Roadblocks removed after peace effo...
- [Jan 27, 2008] Police quelling an uprising by yout...
- [Jan 27, 2008] Killings in Naivasha...
- [Jan 27, 2008] Naivasha - house burned down with 2...
- [Jan 27, 2008] Violence in Naivasha...

More »

Nairobi

Map data ©2010 Google, Tracks4Africa - Terms of Use

Filter By Category

- ALL CATEGORIES
- RIOTS
- DEATHS
- PROPERTY LOSS
- GOVERNMENT FORCES
- CIVILIANS
- LOOTING
- RAPE
- PEACE EFFORTS
- INTERNALLY DISPLACED PEOPLE

GO

Submit An Incident!

Submit Via SMS

Send your SMS to 6007 on your phone (Safaricom/Celtel)

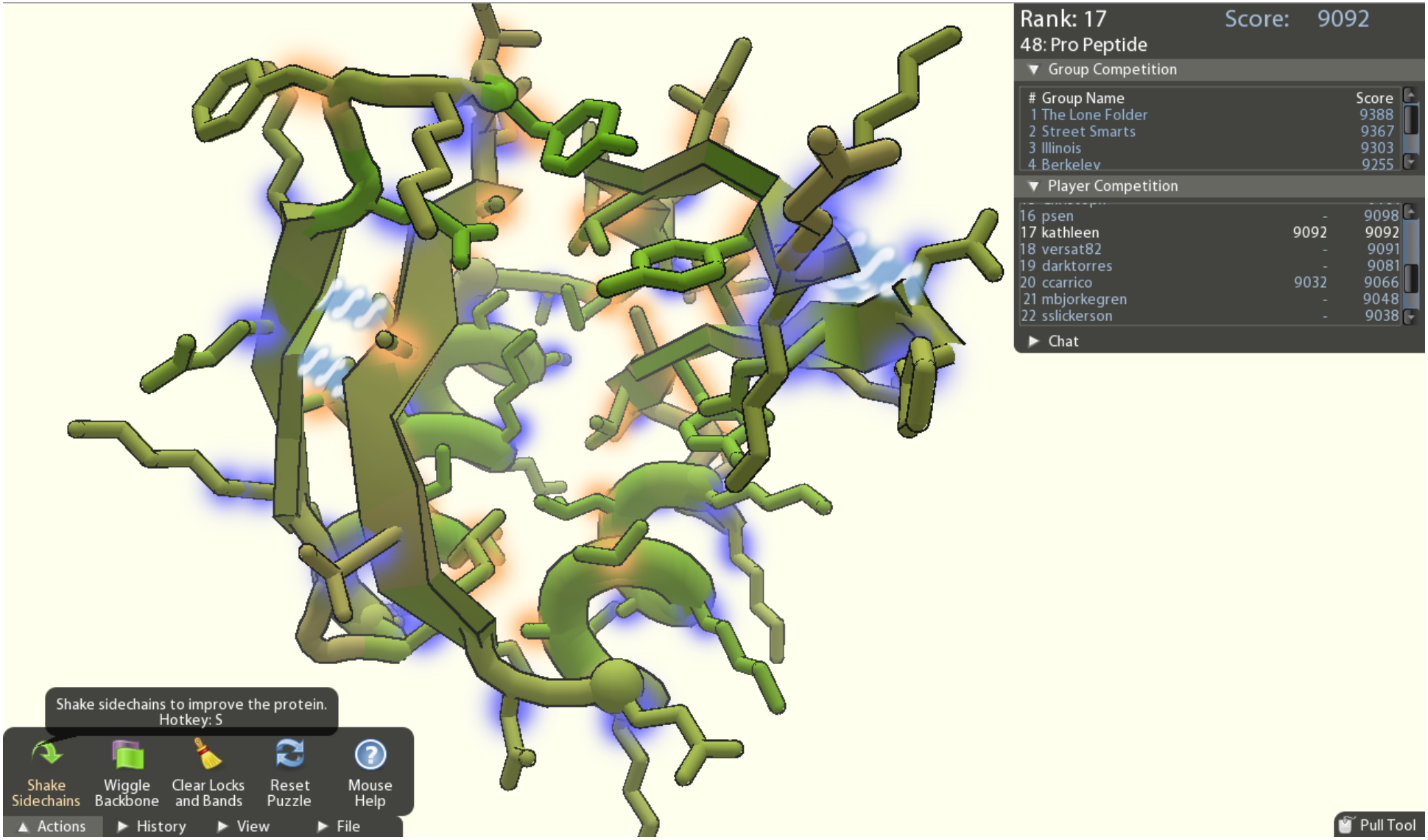
Subscribe To Ushahidi News

TAGS: ALL, Riots, Deaths, Property Damage, Government Forces, Civilians, Looting, Rape, Peace

Incident	Location	Date	Description	Images
Anxiety increases in IDP Camps as supplies	Mathare	Apr 7, 2008	Food and other essential supplies are fast dwindling in the camps for Internally Displaced	

Help Science.

Foldit: Unfold the mystery of science by folding protein



Is crowd any good?

Out of 10 challenges
In 3: Foldit players = best current simulations
In 5: Foldit players > best current simulations
In 2: Foldit players < best current simulations

None of the 5 highest-scoring players took chemistry classes beyond high school.

Crowd is part of academic research

Crystal structure of a monomeric retroviral protease solved by protein folding game players.

Firas Khatib, Frank DiMaio, **Foldit Contenders Group**, **Foldit Void Crushers Group**, Seth Cooper, Maciej Kazmierczyk, Miroslaw Gilski, Szymon Krzywda, Helena Zábranská, Iva Pichová, James Thompson, Zoran Popović, Mariusz Jaskolski and David Baker.

In Nature Structural and Molecular Biology 18, 1175–1177 (2011).

Let’s try a crowdsourcing task.

How many cones in 꼬깔콘? shoutkey.com/good

How many cones in 꼬깔콘?

91

How many cones in 꼬깔콘?
shoutkey.com/good

91

93?

94?

Simple! is not enough...

High variance is quality

People gaming the system

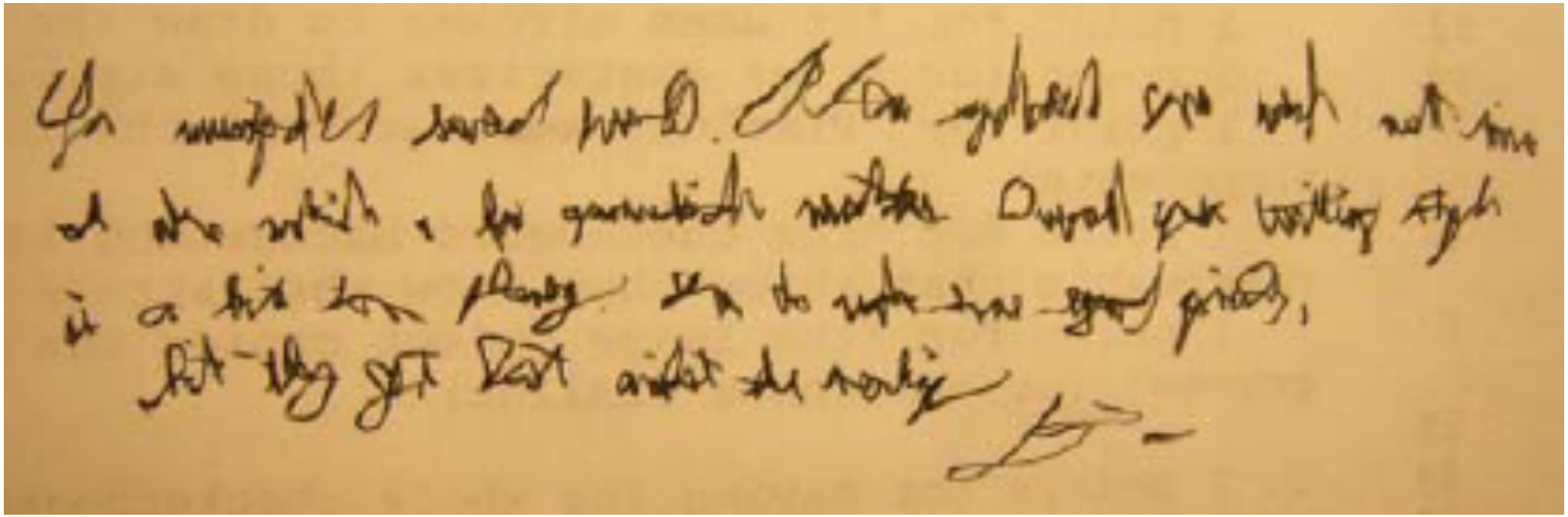
Integration issues

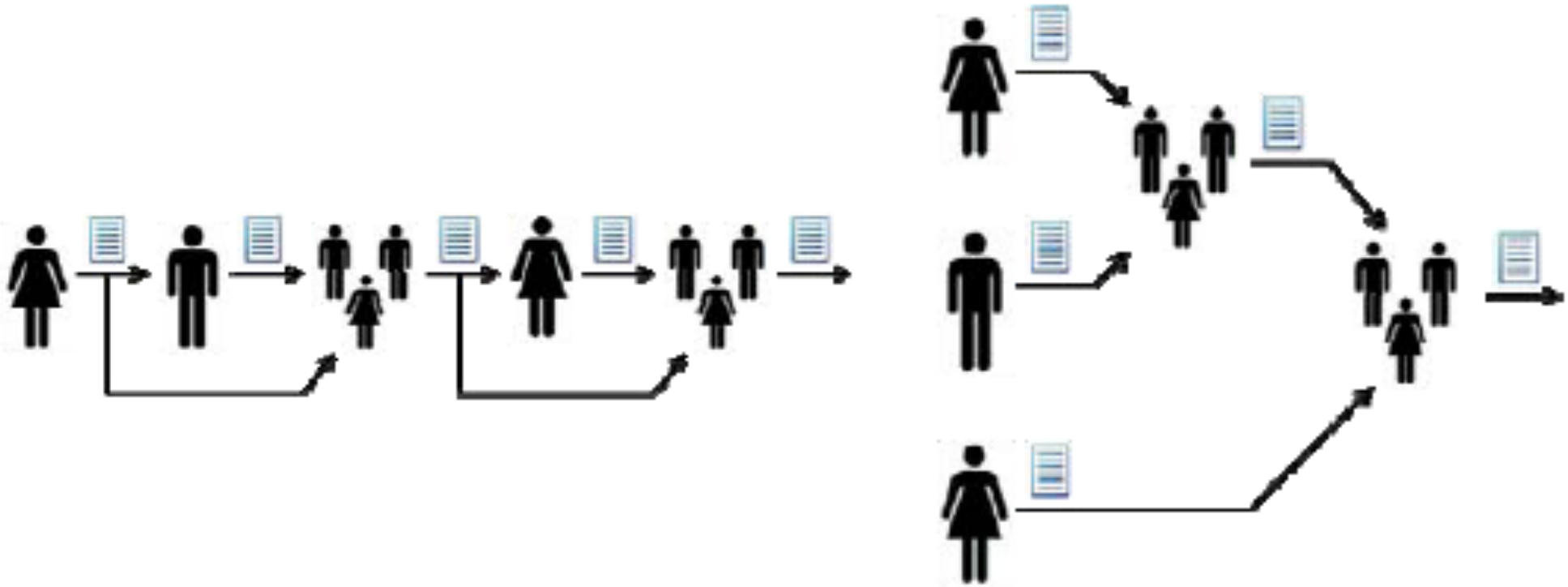
UI design matters

Task design matters

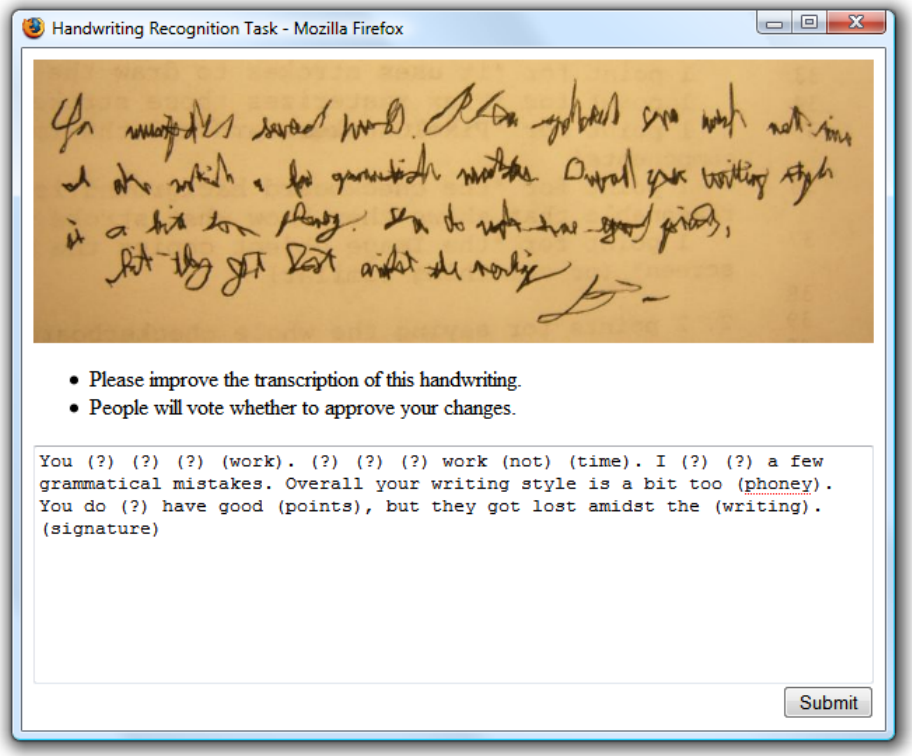
We need to carefully design the process.

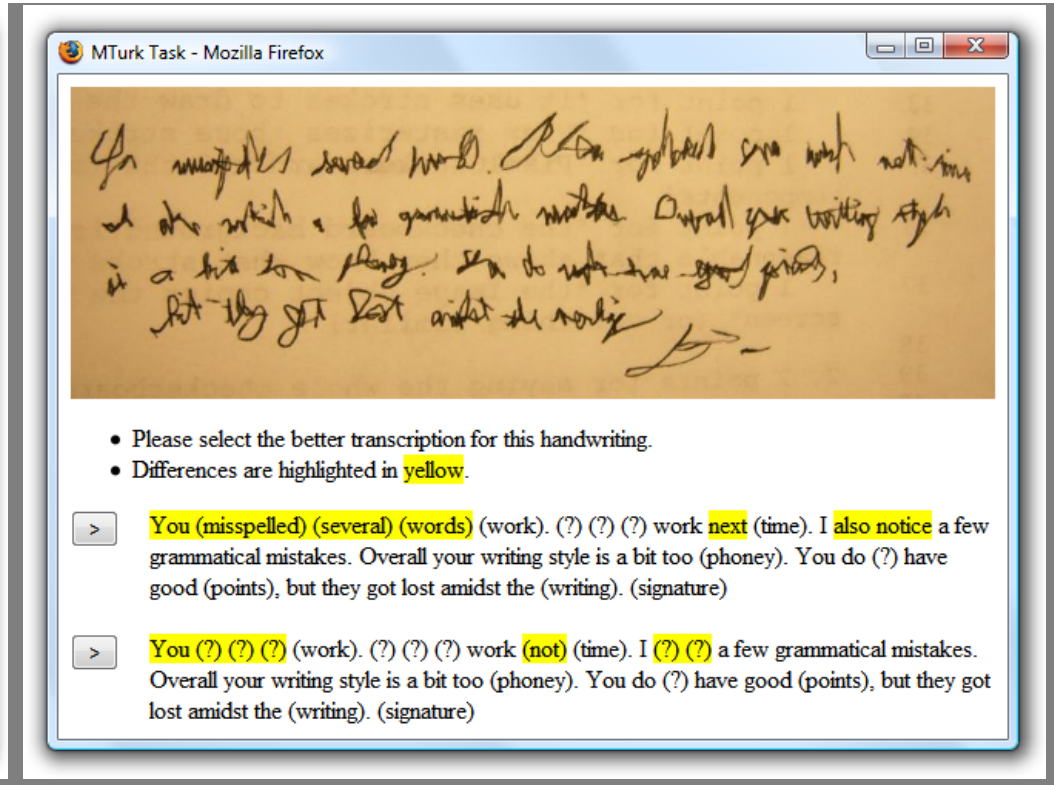
Can you read this?

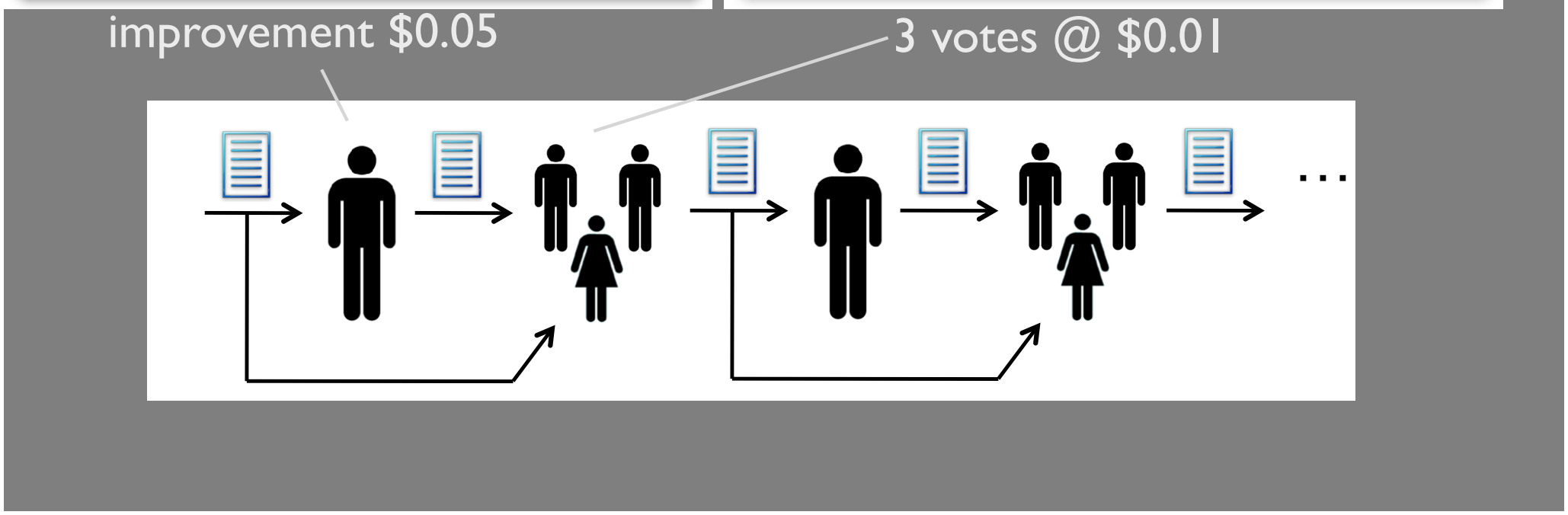


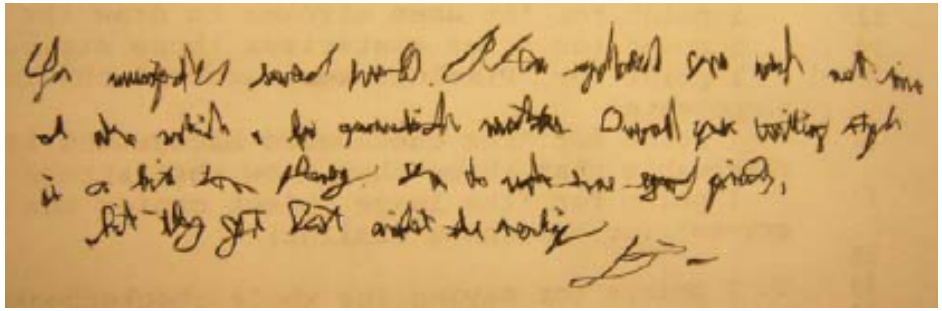


Greg Little et al. Exploring Iterative and Parallel Human Computation Processes. HCOMP 2010.


Handwriting Recognition Task - Mozilla Firefox
• Please improve the transcription of this handwriting.
• People will vote whether to approve your changes.
You (?) (?) (?) (work). (?) (?) (?) work (not) (time). I (?) (?) a few grammatical mistakes. Overall your writing style is a bit too (phoney). You do (?) have good (points), but they got lost amidst the (writing). (signature)
Submit


MTurk Task - Mozilla Firefox
• Please select the better transcription for this handwriting.
• Differences are highlighted in yellow.
>
You (misspelled) (several) (words) (work). (?) (?) (?) work next (time). I also notice a few grammatical mistakes. Overall your writing style is a bit too (phoney). You do (?) have good (points), but they got lost amidst the (writing). (signature)
>
You (?) (?) (?) (work). (?) (?) (?) work (not) (time). I (?) (?) a few grammatical mistakes. Overall your writing style is a bit too (phoney). You do (?) have good (points), but they got lost amidst the (writing). (signature)


improvement $0.05
3 votes @ $0.01
...



version 1:

You (?) (?) (?) (work). (?) (?) (?) work (not) (time). I (?) (?) a few grammatical mistakes. Overall your writing style is a bit too (phoney). You do (?) have good (points), but they got lost amidst the (writing). (signature)

version 2:

~~You (?) (?) (saved) (work). (?) (?) (?) work (not) (time). I (?) (?) a few grammatical mistakes. Overall your writing style is a bit too (phoney). You do (?) have good (points), but they got lost amidst the (writing). (signature)~~

version 4:

You (misspelled) (several) (words). (?) (?) (?) work next (time). I also notice a few grammatical mistakes. …

version 5:

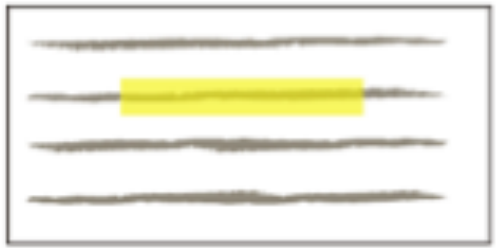
You (misspelled) (several) (words). (Plan?) (spell-check) (your) work next time. I also notice a few grammatical mistakes. Overall your writing style is a bit too phoney. You do make some good (points), but they got lost amidst the (writing). (signature)

version 6:

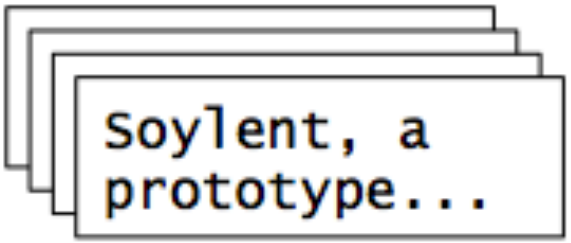
You (misspelled) (several) (words). Please spell-check your work next time. I also notice a few grammatical …

Design Pattern for Quality Control

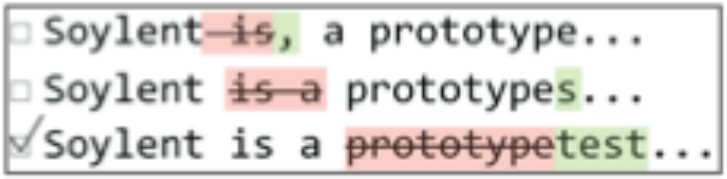
Find ***a problem***



Fix ***each problem***



Verify ***quality of each fix***



Soylent: A Word Processor with a Crowd.
Bernstein, M., Little, G., Miller, R., Hartmann, B.,Ackerman, M., Karger, D., Crowell, D., and Panovich, K.UIST 2010.

Research Topics

Different types of crowd

Real-time crowd at hand

Crowd-powered UI

Crowdsourcing creative, complex work

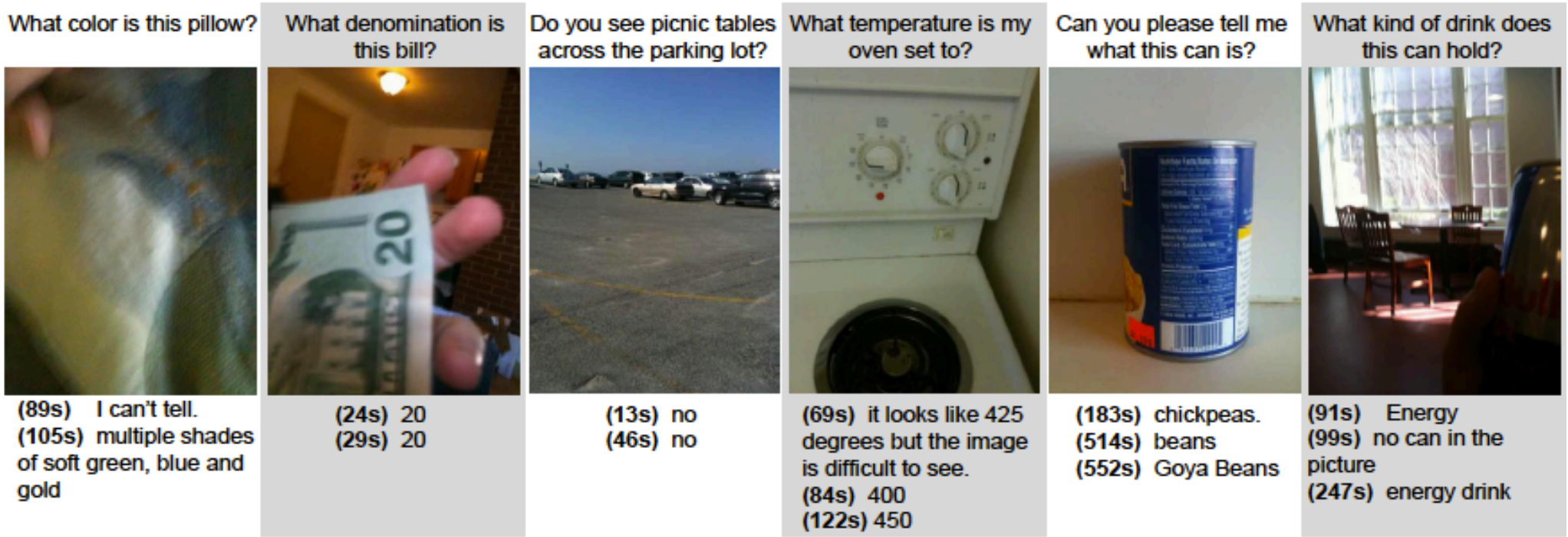
Worker issues, voluntary crowdsourcing

Community-sourcing

Heimarl, Gawalt, Chen, Parikh, and Hartmann. Communitysourcing: Engaging Local Crowds to Perform Expert Work Via Physical Kiosks. CHI2012.

Real-time crowd at hand

Vizwiz: crowd vision for blind people



Jeff Bigham et al. Vizwiz: Nearly Real-time Answers to Visual Questions. UIST 2010

Crowd-powered UI

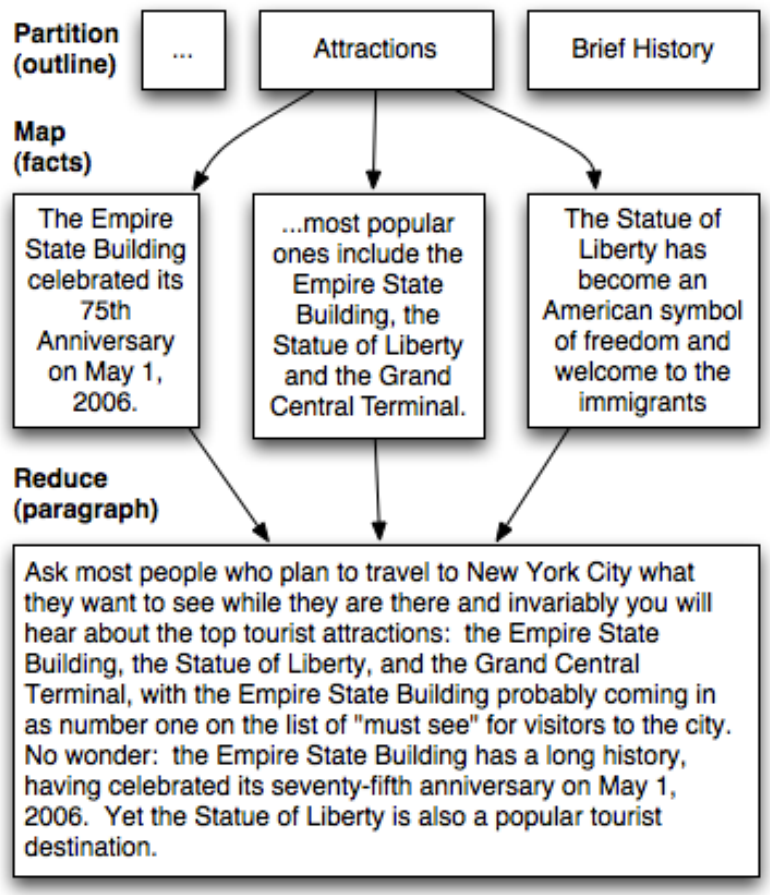
ShadowDraw: Crowd-assisted drawing tool

Slide from Lee's presentation.

Lee, Zitnick, and Cohen. ShadowDraw: Real-Time User guidance for Freehand Drawing. SIGGRAPH 2011.

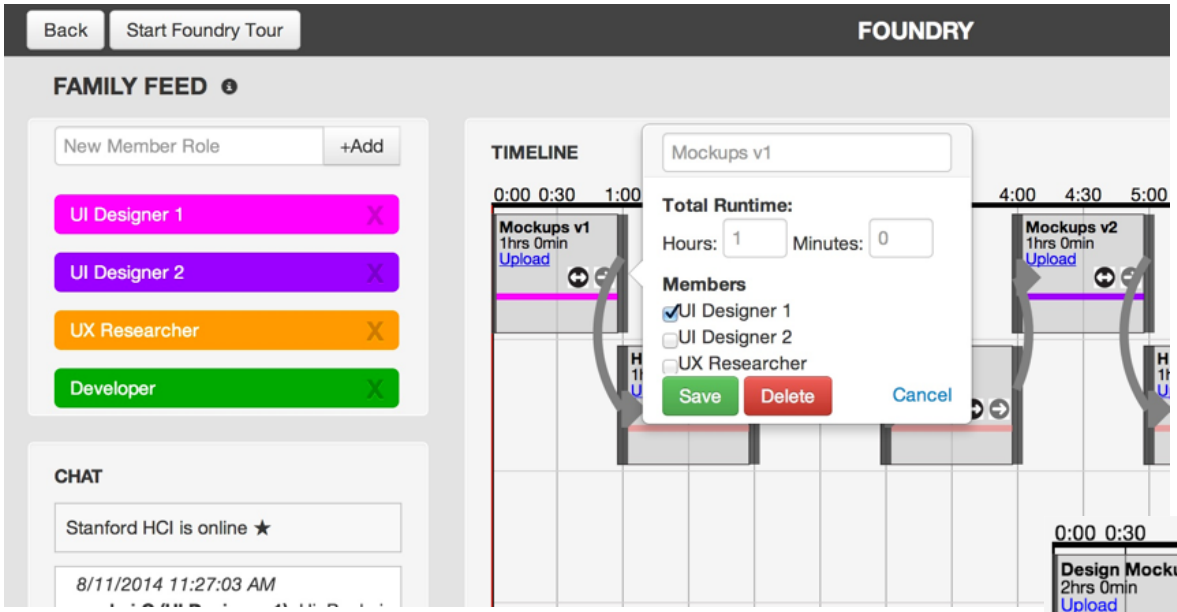
Crowdsourcing creative, complex work

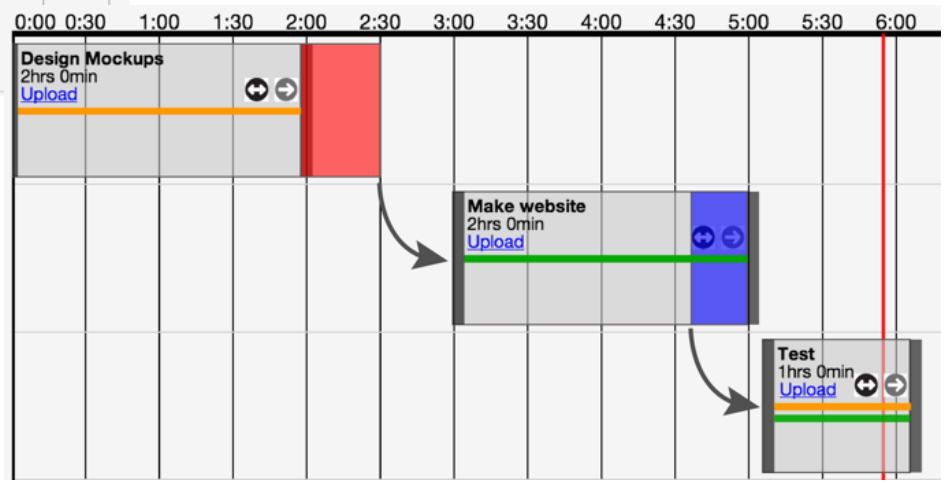
CrowdForge: Article writing with crowd



Kittur, Smus, Khamkar, and Kraut. CrowdForge: Crowdsourcing Complex Work. UIST2011.

Embedding Teamwork: Flash Teams





Retelny et al., Expert Crowdsourcing with Flash Teams. UIST 2014.

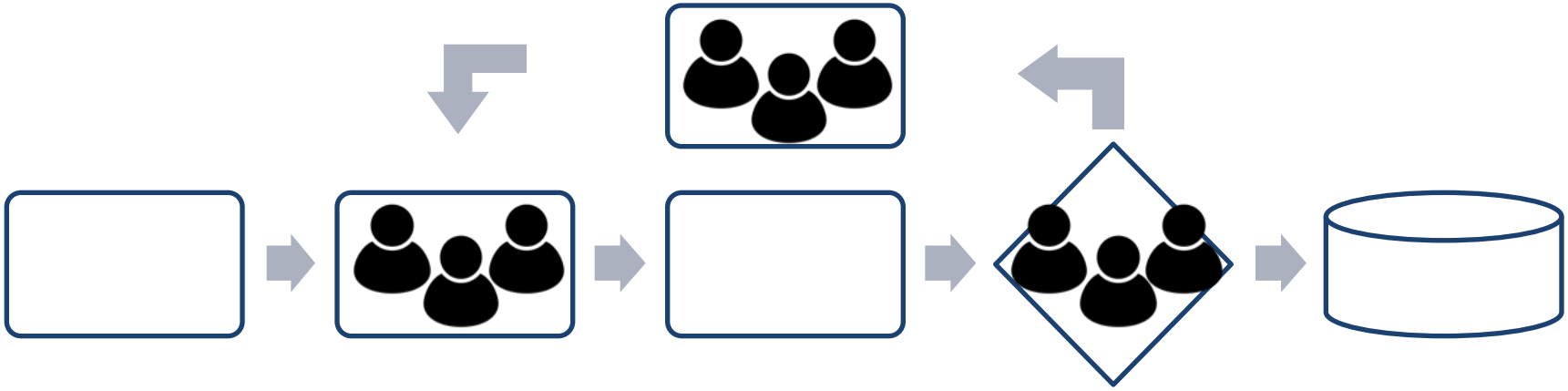
Open Issues

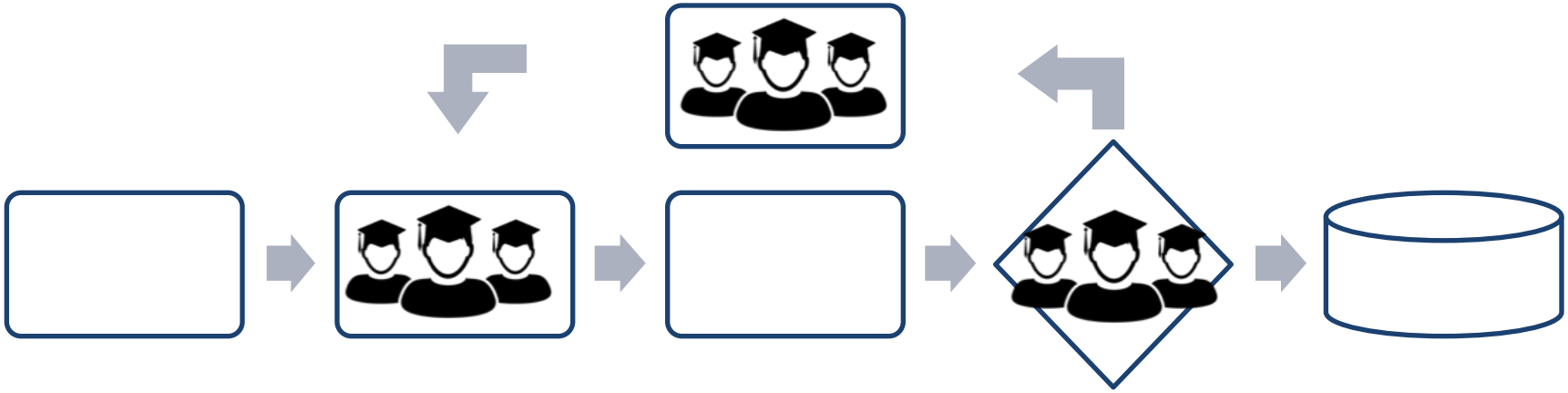
Minimum wage? Labor abuse?

Attribution

Promoting learning & development

Moral valence (Zittrain)



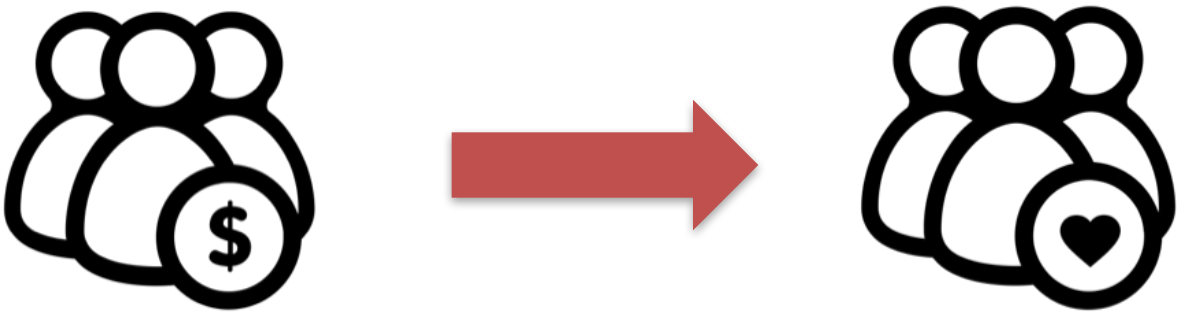


Humans as computational units?

Passive & mechanical crowd

Voluntary, organic crowdsourcing

Individual & collective benefits to the crowd



Course Logistics

No exams
No lectures

In this class, you will

- PRESENT, TEACH
- READ, CRITIQUE
- ANALYZE, IMPLEMENT
- DESIGN, BUILD, TEST
- DISCUSS, SHARE

In this class, you will

- PRESENT, TEACH
  - Topic Presentation (1-2 times / semester)
- READ, CRITIQUE
  - Reading Response (each class)
- ANALYZE, IMPLEMENT
  - Assignments (once every week or two)
- DESIGN, BUILD, TEST
  - Final Project (throughout the semester)
- DISCUSS, SHARE
  - In-class, asynchronous discussion (anytime)

In this class, you will

- PRESENT, TEACH — 20%
  - Topic Presentation (1-2 times / semester)
- READ, CRITIQUE — 20%
  - Reading Response (each class)
- ANALYZE, IMPLEMENT — 20%
  - Assignments (once every week or two)
- DESIGN, BUILD, TEST — 30%
  - Final Project (throughout the semester)
- DISCUSS, SHARE — 10%
  - In-class, asynchronous discussion (anytime)

Topic Presentation

- You'll lead the class
  - Summarize the readings and peer students' critiques
  - Show related examples
  - Design in-class activities
  - Spur the in-class discussion

Reading Response

- For each class, you'll read 1-2 papers and submit your critique.
  - Summarize main ideas and discuss why they matter.
  - What have you learned? What did you like about the paper?
  - Methodological / logical / technical concerns? How would you improve the work?

Assignment

- Hands-on activities
  - Become a crowd worker
  - Survey & Analyze existing crowdsourcing systems
  - Replicate results from an academic paper
  - Implement a crowdsourcing technique

Final Project

- Design, build, and test your own crowdsourcing system.
- Team of 2-3
- Connecting to your own research is encouraged.

Final Project Milestones

- Idea (week 4)
- Abstract (week 6)
- Pitch (week 8)
- Prototype 1 (week 10)
- Prototype 2 (week 13)
- Final presentation (week 15)
- Final paper (week 16)

Participation

- In-class
  - Please speak!
  - Contribute your own (rough) perspective.
- At home
  - Piazza discussion forum
  - Share cool examples
  - Ask and answer questions

Each class will look like...

Topic intro 15 mins	Activities, Examples 15 mins	Student Presentation, Discussion 40 mins

Topics: Intro

- Introduction to crowdsourcing and human computation
- Crowdsourcing systems, platforms, and workers

Topics: Technique

- Programming paradigms
- Quality control
- Crowd agents, realtime crowdsourcing

Topics: Design

- Workflow design & task decomposition
- Incentive mechanisms & gamification
- Experiments & analysis

Topics: Application

- Data labeling, machine learning, computer vision, NLP
- Can the crowd learn? Learnersourcing, education, feedback
- Citizen science & civic engagement
- Teamwork, expert crowdsourcing, accessibility

Office Hours

- Juho Kim
  - 4-5:30 PM Tue/Thu @ N1-605
- Seunghyun Han
  - 10-11:30 AM Wed @ N1-624

TODO items for you

- Visit the course website
  - http://kixlab.org/courses/crowdsourcing/
  - Course updates will be posted here.
  - KLMS is only for assignment submission.

- Enroll in the class on Piazza
  - Course website has more info.

Next class...

- What is crowdsourcing and human computation?