

# Human Computation Architectures

Swati Mishra

Applications of Machine Learning (4AL3)

Fall 2024



---

ENGINEERING

# Review

- Week 1 : Linear Regression, Gradient Descent
- Week 2: Polynomial Regression, Logistic Regression,
- Week 2-3: Machine Learning Evaluation, Data Pre-Processing
- Week 3: Support Vector Machines

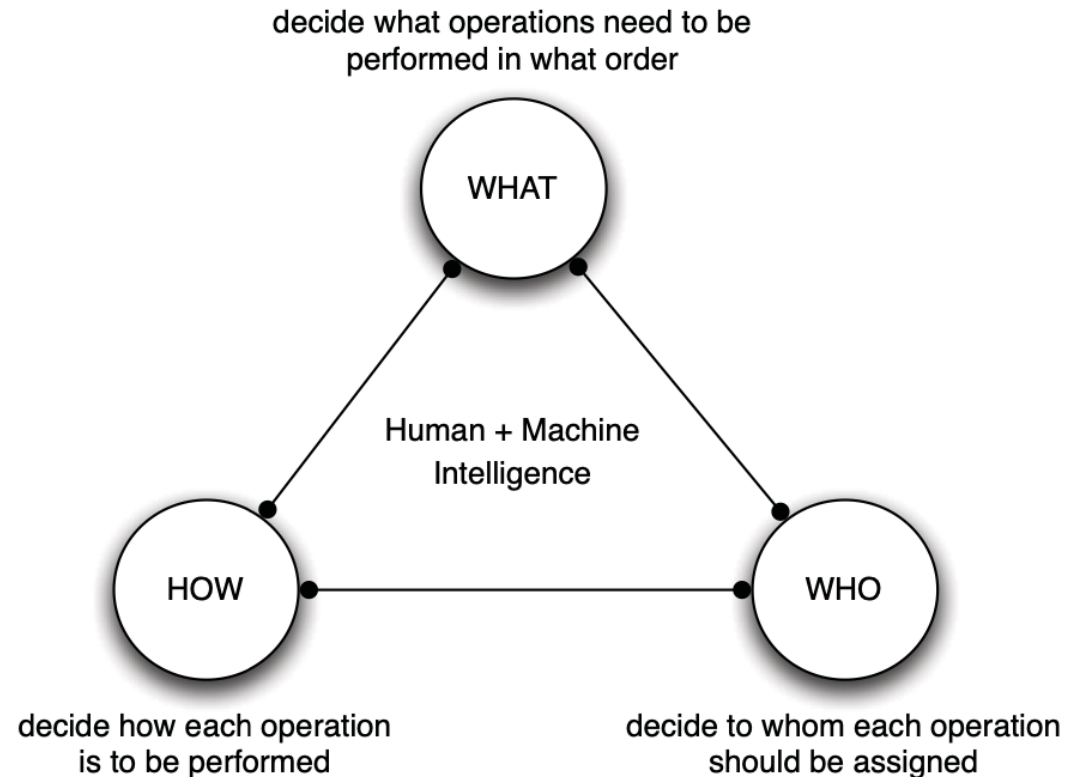
# What is Human Computation

- Computation is the process of mapping some input representation using an explicit, finite set of instructions (i.e. an *algorithm*).
- **Human computation systems** can be defined as intelligent systems that organize humans to carry out the process of computation for performing:
  - the basic operations (or units of computation),
  - taking charge of the control process itself (e.g., decide what operations to execute next or when to halt the program),
  - synthesizing the program itself (e.g., by creating new operations and specifying how they are ordered).

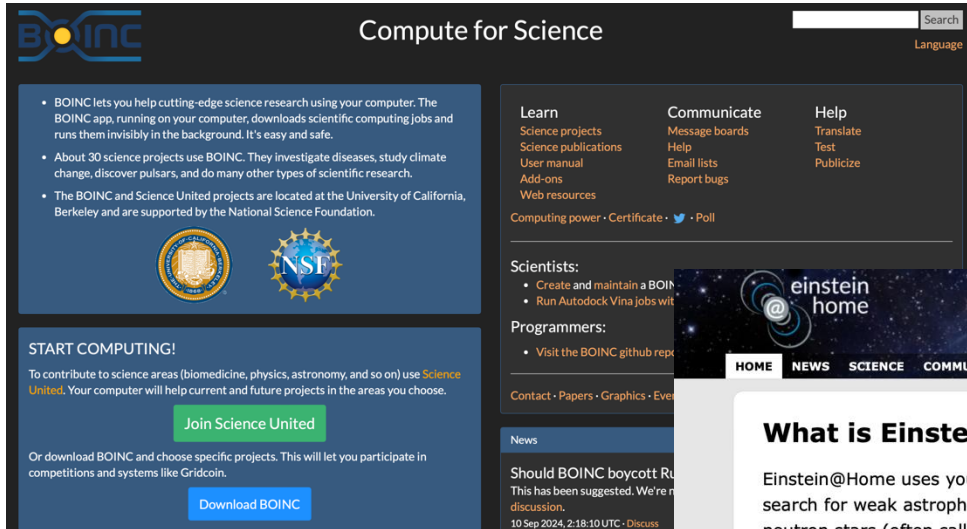
# What is Human Computation

- The concept of human computation is often related to:
  - **Crowdsourcing:** The act of outsourcing tasks, traditionally performed by an employee or contractor, to an undefined, large group of people or community (a crowd) through an open call.
  - **Collective Intelligence:** A shared or group intelligence that emerges from the collaboration and competition of many individuals and appears in consensus decision making in bacteria, animals, humans and computer networks.
  - **Social Computing :** Technology for supporting social behavior in or through computational systems, e.g., blogs, email, instant messaging, social network services, wikis and social bookmarking. Technology for supporting computations that are carried out by groups of people, e.g., collaborative filtering, online auctions, prediction markets, reputation systems, computational social choice, tagging and verification games.

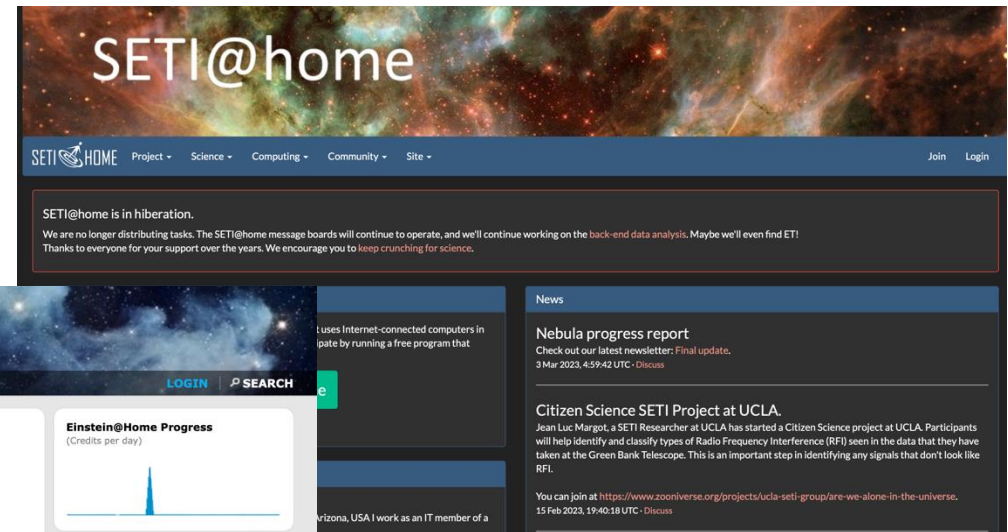
# Aspects of Human Computation Systems



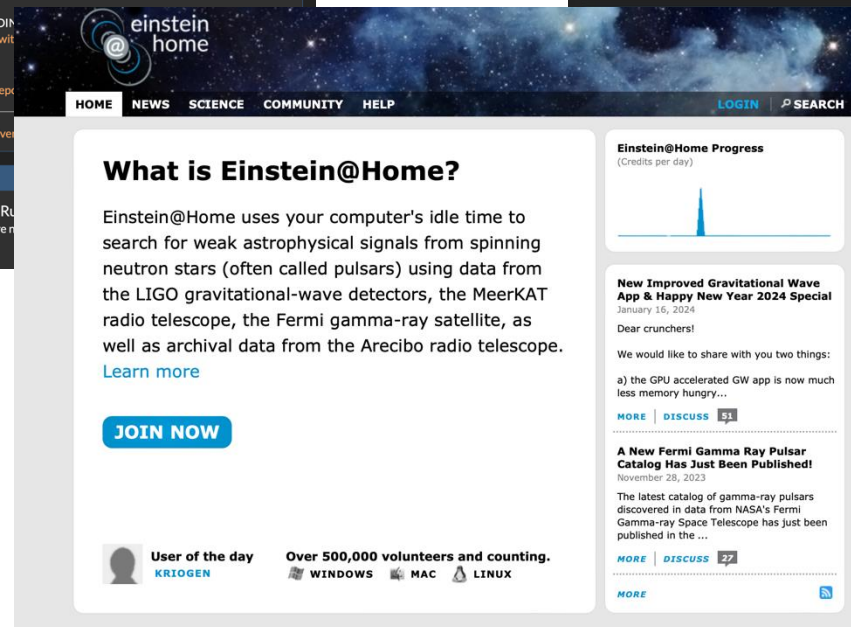
# Examples of Human Computation Systems



Bionic



Seti@Home



Einstein@Home

# What is Human Computation

- Human computation: When humans actively decide when and where to collect the data.
  - Example: Playing game with AI
- Not Human Computation: When they have no **conscious** role in determining the outcome.
  - Example: Logging sensor data from phone.

# Frameworks for Human Computation

The screenshot shows the Prolific workspace for a study titled "MATB-II Workload study Part 2a". The interface includes a sidebar with navigation options: My Workspace, Projects, Project, Finance (\$0.01), Team, and Settings. The main content area displays study statistics: Published (4 Jan 2021, 12:38), Adjust reward per hour (\$4.98/hr), Median time (00:07:14), Submissions in progress (218 / 218), and Eligible participants (35,369 of 123,725). Below these statistics are buttons for "Approve all", "Message all", and "Bonus payment all", along with a search bar "Find by ID...". A table lists individual participants with columns for Participant Prolific ID, Started, Time Taken, Completion Code, Status, and Bonus.

PARTICIPANT PROLIFIC ID	STARTED	TIME TAKEN	COMPLETION CODE	STATUS	BONUS
5dd32014b51b3e33ec0d07cc	1 Mar 2021, 11:23	00:11:30	COMPLETED	APPROVED	\$1.00
5e815773285d3883ba2fbcbf	1 Mar 2021, 11:18	00:06:06	COMPLETED	APPROVED	\$1.00
60015f803919e70009ef8dc5	1 Mar 2021, 11:13	N/A		RETURNED	
5e387f2cd070ef099d5e5a4f	1 Mar 2021, 11:11	00:13:34	COMPLETED	APPROVED	\$1.00
5bd7971b0aac450001f951aa	1 Mar 2021, 10:59	N/A		RETURNED	
5eff11ea32f1ca44319deefa	1 Mar 2021, 10:58	00:14:08	COMPLETED	APPROVED	
5c636176bad3560001c27467	1 Mar 2021, 10:58	00:10:04	COMPLETED	APPROVED	\$1.00

Prolific

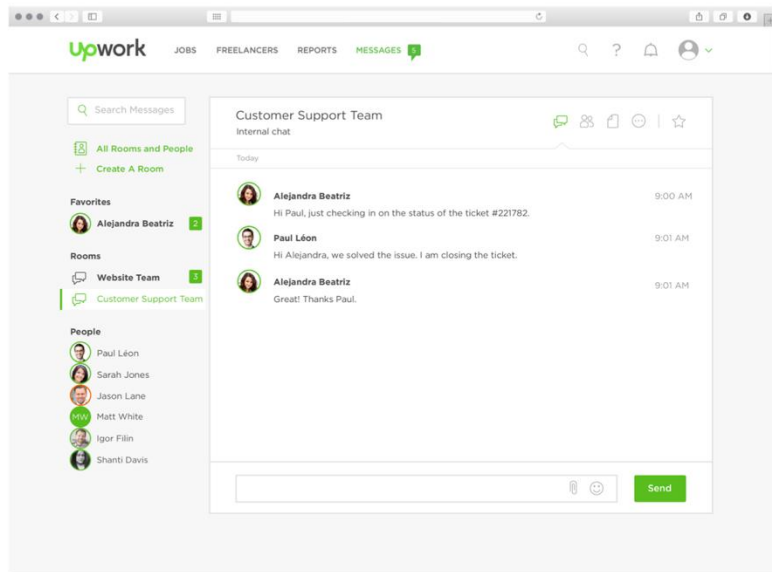
The screenshot shows the Amazon Mechanical Turk HIT Groups page. It includes a search bar "Search All HITs" and a table listing HIT Groups. The table columns are Requester, Title, HITs, Reward, Created, and Actions. The table shows 10 HIT Groups, each with a "Preview" button and an "Accept & Work" button.

Requester	Title	HITs	Reward	Created	Actions
Amazon Requester Inc. - C	[French language proficiency requir...	61,046	\$0.01	17h ago	Preview Accept & Work
Amazon Requester Inc. - C	[日本語能力が必要]商品のインタレ...	59,647	\$0.01	7h ago	Preview Accept & Work
Amazon Requester Inc. - C	Product to Interest Audit (single yes/...	28,379	\$0.01	1h ago	Preview Accept & Work
Amazon Requester Inc. - C	[dominio del idioma español requeri...	27,670	\$0.01	21h ago	Preview Accept & Work
Amazon Requester Inc. - C	[Proficiência no idioma português br...	19,719	\$0.01	20h ago	Preview Accept & Work
Crowdsurf Support	Transcribe up to 35 Seconds of Med...	17,485	\$0.05	3m ago	Preview Qualify
TC Research	Find the Email for These Mental He...	13,896	\$0.12	5d ago	Preview Accept & Work
UnSpun Opinions	Opinion Survey	12,180	\$0.50	1m ago	Preview Accept & Work
KronoPin	Find the Website Address for a Con...	11,846	\$0.03	2/23/2018	Preview Qualify
Assistive Technology Rese	1 minute survey: Smart speakers at ...	10,577	\$0.15	3d ago	Preview Accept & Work
Armin Hamzic	Tell us if a picture shows a specific f...	10,557	\$0.01	1d ago	Preview Qualify
nttkkAN	Image Annotation (WARNING: This ...	8,217	\$0.05	10d ago	Preview Accept & Work

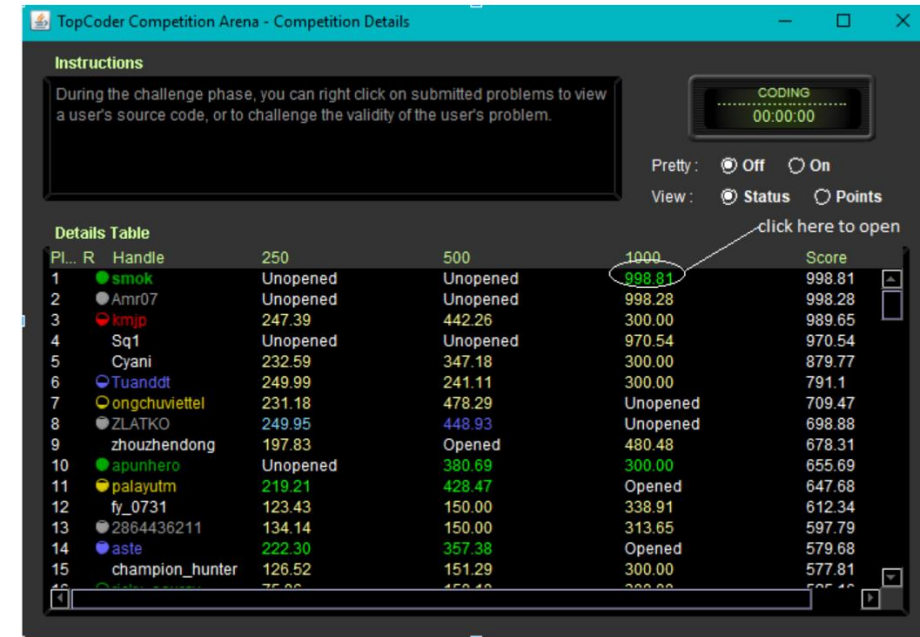
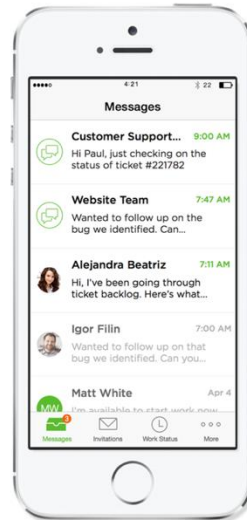
Amazon Mechanical Turk



# Frameworks for Human Computation



Upwork



Top Coder

# Frameworks for Human Computation

**AMAZON SAGEMAKER GROUND TRUTH**

Through SageMaker Ground Truth, you can access the MTurk workforce and implement additional validation and quality checks for a scalable and cost-effective way to train and improve ML models. Use cases include data annotation and human data verification. SageMaker Ground Truth also offers curated workforces for Generative AI use cases including content generation, image captioning, human evaluation, prompt engineering, human feedback, and more.

Learn more and get started on the [SageMaker Ground Truth](#) website.



**Introducing AWS billing with Amazon Mechanical Turk!**

Enable AWS billing to leverage your AWS account to pay for work completed on MTurk.

[Learn more](#)

[Enable AWS Billing](#)

Select a customizable template to start a new project

Survey

[Survey Link](#)

Survey

Vision

Image Classification

Bounding Box

Semantic Segmentation

Instance Segmentation

Polygon

Keypoint

Image Contains

Video Classification

Moderation of an Image

Image Tagging

Image Summarization

Language

Sentiment Analysis

Intent Detection

Collect Utterance

Emotion Detection

Semantic Similarity

**Survey Link Instructions (Click to expand)**

**Survey link:**

**Provide the survey code here:**

You must ACCEPT the HIT before you can submit the results.

amazonmturk Worker

HITS Dashboard Qualifications Search All HITS

All HITS Your HITS Queue

HIT Groups (1-20 of 2106)

Show Details Hide Details Items Per Page: 20

Requester	Title	HITS	Reward	Created	Actions
Amazon Requester Inc. - C	[French language proficiency requir...	61,046	\$0.01	17h ago	<a href="#">Preview</a> <a href="#">Accept &amp; Work</a>
Amazon Requester Inc. - C	[日本語能力が必要]商品のインタレ...	59,647	\$0.01	7h ago	<a href="#">Preview</a> <a href="#">Accept &amp; Work</a>
Amazon Requester Inc. - C	Product to Interest Audit (single yes/...	28,379	\$0.01	1h ago	<a href="#">Preview</a> <a href="#">Accept &amp; Work</a>
Amazon Requester Inc. - C	[dominio del idioma español requeri...	27,670	\$0.01	21h ago	<a href="#">Preview</a> <a href="#">Accept &amp; Work</a>
Amazon Requester Inc. - C	[Proficiência no idioma português br...	19,719	\$0.01	20h ago	<a href="#">Preview</a> <a href="#">Accept &amp; Work</a>
Crowdsurf Support	Transcribe up to 35 Seconds of Med...	17,485	\$0.05	3m ago	<a href="#">Preview</a> <a href="#">Qualify</a>
TC Research	Find the Email for These Mental He...	13,896	\$0.12	5d ago	<a href="#">Preview</a> <a href="#">Accept &amp; Work</a>
UnSpun Opinions	Opinion Survey	12,180	\$0.50	1m ago	<a href="#">Preview</a> <a href="#">Accept &amp; Work</a>
KronoPin	Find the Website Address for a Con...	11,846	\$0.03	2/23/2018	<a href="#">Preview</a> <a href="#">Qualify</a>
Assistive Technology Rese	1 minute survey: Smart speakers at ...	10,577	\$0.15	3d ago	<a href="#">Preview</a> <a href="#">Accept &amp; Work</a>
Armin Hamzic	Tell us if a picture shows a specific f...	10,557	\$0.01	1d ago	<a href="#">Preview</a> <a href="#">Qualify</a>
nttkkAN	Image Annotation (WARNING: This ...	8,217	\$0.05	10d ago	<a href="#">Preview</a> <a href="#">Accept &amp; Work</a>

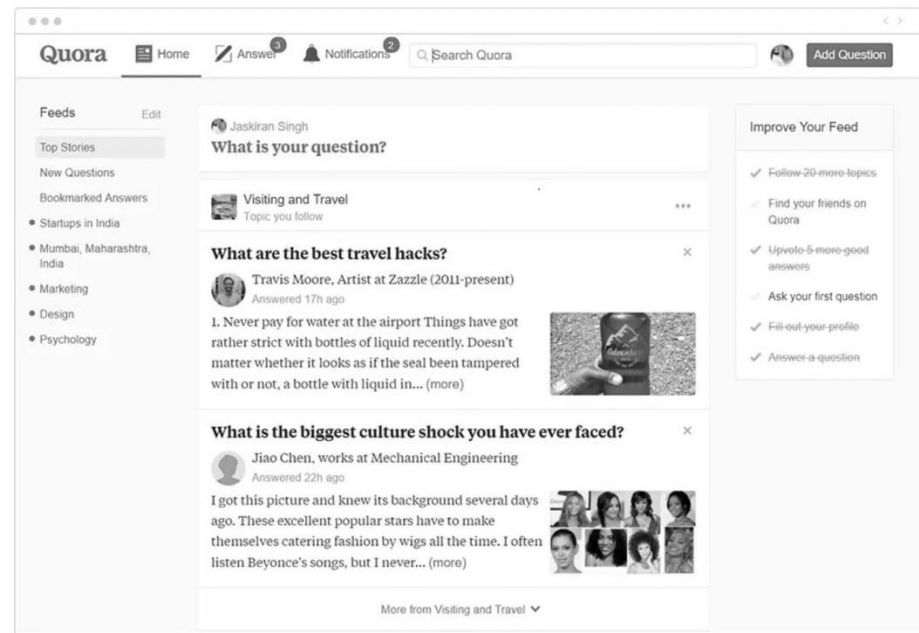
There are two sides of the crowd markets – Workers and Requesters

# Wisdom of the Crowds

- It is the collective opinion of a diverse and independent group of individuals rather than that of a single expert.



WIKIPEDIA  
*The Free Encyclopedia*



# Wisdom of Crowds

When is wisdom of the crowds beneficial.

- Combinatorial problems such as minimum spanning trees and the traveling salesman problem
- Ordering problems such as the order of the U.S. presidents or world cities by population.
- Multi-armed bandit problems, in which participants choose from a set of alternatives with fixed but unknown reward rates.

# Human computation and Machine Learning

## Large Scale Visual Recognition Challenge Case Study



IMAGENET

14,197,122 images, 21841 synsets indexed  
[Home](#) [Download](#) [Challenges](#) [About](#)  
Not logged in, [Login](#) | [Signup](#)

### ImageNet Large Scale Visual Recognition Challenge (ILSVRC)

#### Competition

The ImageNet Large Scale Visual Recognition Challenge (ILSVRC) evaluates algorithms for object detection and image classification at large scale. One high level motivation is to allow researchers to compare progress in detection across a wider variety of objects -- taking advantage of the quite expensive labeling effort. Another motivation is to measure the progress of computer vision for large scale image indexing for retrieval and annotation.

For details about each challenge please refer to the corresponding page.

- [ILSVRC 2017](#)
- [ILSVRC 2016](#)
- [ILSVRC 2015](#)
- [ILSVRC 2014](#)
- [ILSVRC 2013](#)
- [ILSVRC 2012](#)
- [ILSVRC 2011](#)
- [ILSVRC 2010](#)

#### Workshop

Every year of the challenge there is a corresponding workshop at one of the premier computer vision conferences. The purpose of the workshop is to present the methods and results of the challenge. Challenge participants with the most successful and innovative entries are invited to present. Please visit the corresponding challenge page for workshop schedule and information.

#### Download

The most popular challenge is the ILSVRC 2012-2017 image classification and localization task. It is available on [Kaggle](#). For all other data please log in or request access.

#### Evaluation Server

The [evaluation server](#) can be used to evaluate image classification results on the test set of ILSVRC 2012-2017. Please see [here](#) for our submission policy. Importantly, you should not make more than 2 submissions per week.

#### Updates

- October 10, 2019: The ILSVRC 2012 classification and localization test set has been updated. The [Kaggle challenge](#) and our [download page](#) both now contain the updated data.
- June 2, 2015: [Follow-up update regarding status of the server](#)
- May 19, 2015: [Announcement regarding the submission server](#)

#### Citation

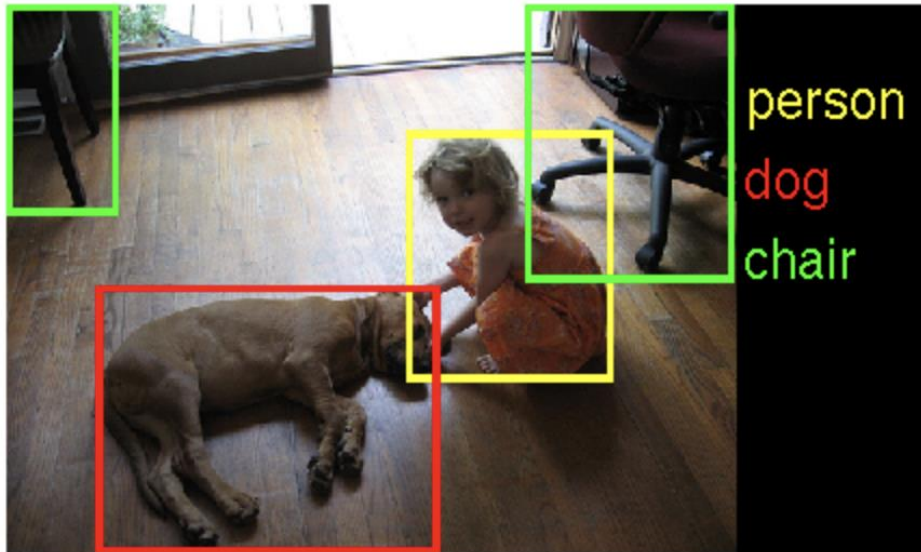
When reporting results of the challenges or using the datasets, please cite:

Olga Russakovsky\*, Jia Deng\*, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, Alexander C. Berg and Li Fei-Fei. (\* = equal contribution) **ImageNet Large Scale Visual Recognition Challenge**. *IJCV*, 2015. [paper](#) | [bibtex](#) | [paper content on arxiv](#) | [attribute annotations](#)



# Human computation and Machine Learning

## Large Scale Visual Recognition Challenge Case Study



Label hierarchy

Man-made objects  
Furniture  
Animals

Labels  
Input

	Table	Chair	Bowl	Dog	Cat	...
	+	+	-	-	-	-
	+	-	+	-	+	-
	+	+	-	-	-	-
	-	-	-	+	-	-

Label correlation

Label sparsity

Annotation task required workers to identify various objects in the image

*ILSVRC2010  
challenge's winning entry  
use stochastic SVM!*

# Human computation and Machine Learning

## Large Scale Visual Recognition Challenge Case Study



~ 2 million images

IMAGENET

14,197,122 images, 21841 synsets indexed  
Home Download Challenges About  
Not logged in, Login | Signup

## ImageNet Large Scale Visual Recognition Challenge (ILSVRC)

### Competition

The ImageNet Large Scale Visual Recognition Challenge (ILSVRC) evaluates algorithms for object detection and image classification at large scale. One high level motivation is to allow researchers to compare progress in detection across a wider variety of objects -- taking advantage of the quite expensive labeling effort. Another motivation is to measure the progress of computer vision for large scale image indexing for retrieval and annotation.

For details about each challenge please refer to the corresponding page.

- [ILSVRC 2017](#)
- [ILSVRC 2016](#)
- [ILSVRC 2015](#)
- [ILSVRC 2014](#)
- [ILSVRC 2013](#)
- [ILSVRC 2012](#)
- [ILSVRC 2011](#)
- [ILSVRC 2010](#)

### Workshop

Every year of the challenge there is a corresponding workshop at one of the premier computer vision conferences. The purpose of the workshop is to present the methods and results of the challenge. Challenge participants with the most successful and innovative entries are invited to present. Please visit the corresponding challenge page for workshop schedule and information.

### Download

The most popular challenge is the ILSVRC 2012-2017 image classification and localization task. It is available on [Kaggle](#). For all other data please log in or request access.

### Evaluation Server

The [evaluation server](#) can be used to evaluate image classification results on the test set of ILSVRC 2012-2017. Please see [here](#) for our submission policy. Importantly, you should not make more than 2 submissions per week.

### Updates

- October 10, 2019: The ILSVRC 2012 classification and localization test set has been updated. The [Kaggle challenge](#) and our [download page](#) both now contain the updated data.
- June 2, 2015: [Follow-up update regarding status of the server](#)
- May 19, 2015: [Announcement regarding the submission server](#)

### Citation

When reporting results of the challenges or using the datasets, please cite:

Olga Russakovsky\*, Jia Deng\*, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, Alexander C. Berg and Li Fei-Fei. (\* = equal contribution) **ImageNet Large Scale Visual Recognition Challenge**. *IJCV*, 2015. [paper](#) | [bibtex](#) | [paper content on arxiv](#) | [attribute annotations](#)

# Human computation and Machine Learning

Human Computation tasks in Machine Learning Applications:

- Generating Binary or Categorical Labels
  - Example: a binary label indicating whether or not the website contains profanity.
  - Popular model to implement human computation architecture in this setting is Dawid-Skene model that assumes instances are homogeneous
  - A worker's probability of labeling any given instance correctly is controlled by one or more worker-specific quality parameters.

$N$  = Number of computational tasks.

$Y_n$  = *true* output each task is unknown.

Our goal is to estimate  $Y_n$  given an output matrix  $O$ , containing the responses from  $M$  workers.

AGGREGATING OUTPUTS

		computational task			
		1	2	...	$N$
worker	1	$O_{11}$	$O_{12}$	$\cdots$	$O_{1N}$
	2	$O_{21}$	$O_{22}$	$\cdots$	$O_{2N}$
	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
	$M$	$O_{M1}$	$O_{M2}$	$\cdots$	$O_{MN}$



# Human computation and Machine Learning

Human Computation tasks in Machine Learning Applications:

- Generating Binary or Categorical Labels
  - Example: a binary label indicating whether or not the website contains profanity.

---

**Algorithm 1** The basic EM framework of Dawid and Skene (1979).

---

**Input:** Sets of worker-generated labels for each instance

Initialize each instance's label based on a simple majority vote

**repeat**

**for all** Workers  $w$  **do**

    Calculate  $w$ 's quality parameter(s), treating each instance's current label as ground truth

**end for**

**for all** Instances  $i$  **do**

    Calculate the most likely label for  $i$ , treating each worker's approximated quality parameter(s) as ground truth

**end for**

**until** Label assignments have converged

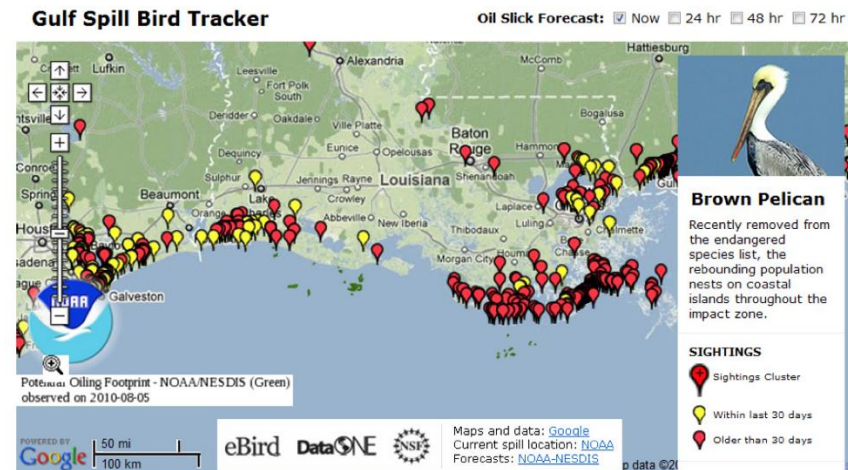
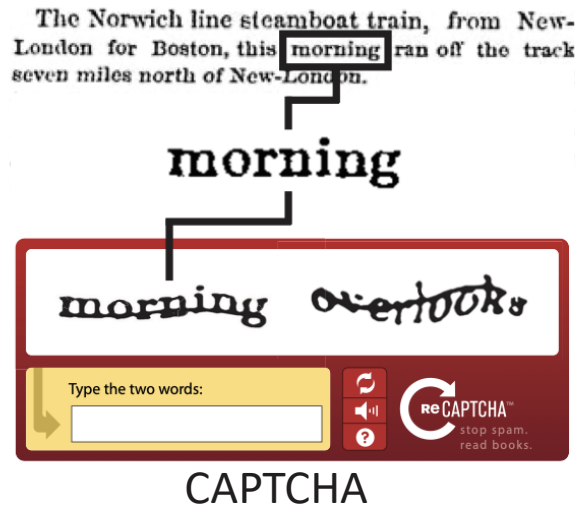
**Output:** The current label assignments for each instance

---

# Human computation and Machine Learning

Human Computation tasks in Machine Learning Applications:

- Generating Binary or Categorical Labels
- Generating Transcriptions, Translations, and Image Annotations
  - Example: ReCAPTCHA, Citizen Science

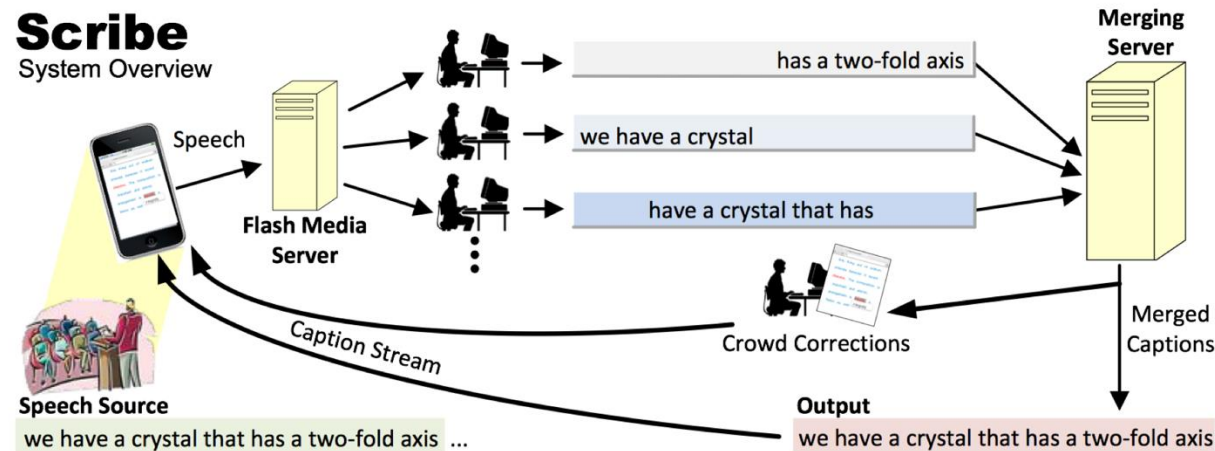


Bird sightings of birds during the Gulf Spill.

# Human computation and Machine Learning

Human Computation tasks in Machine Learning Applications:

- Generating Binary or Categorical Labels
- Generating Transcriptions, Translations, and Image Annotations
  - Example: ReCAPTCHA, Citizen Science



The architecture of Scribe Lasecki et al. (2012).

# Human computation and Machine Learning

Human Computation tasks in Machine Learning Applications:

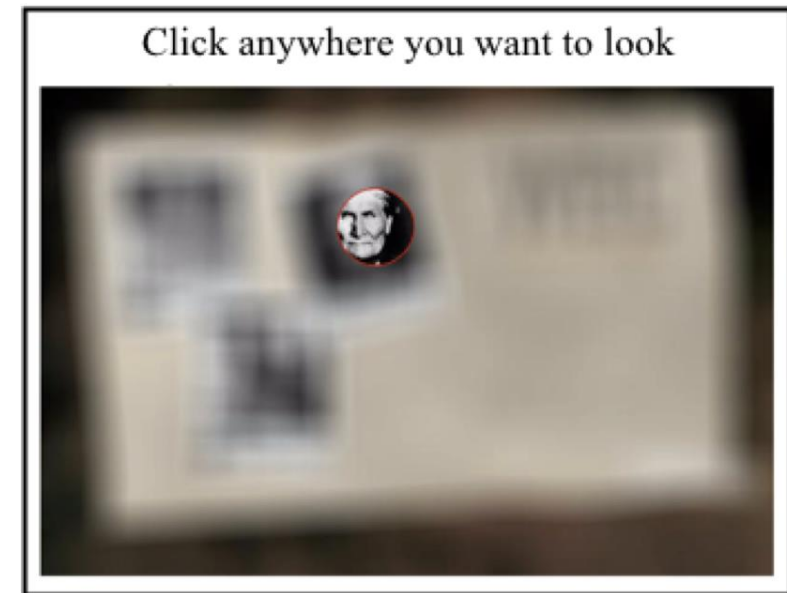
- Generating Binary or Categorical Labels
- Generating Transcriptions, Translations, and Image Annotations
- Evaluating and Debugging Models
  - Crowdsourced evaluation is especially common for unsupervised models, which generally cannot be evaluated in terms of simple metrics like accuracy or precision because there is no objective notion of ground truth.

# Human computation and Machine Learning

Human Computation tasks in Machine Learning Applications:

- Asking the right question.
  - Influence of presented information on task performance.
- Task Granularity:
  - Well defined
  - Not cognitively overwhelming
- Task should be done independently
- Incentive should be adequate
- Outcome must have quality control

**BubbleView (2017)**



# Human computation and Machine Learning

Human Computation tasks in Machine Learning Applications:

- Asking the right question.
  - Influence of presented information on task performance.
- Task Granularity:
  - Well defined
  - Not cognitively overwhelming
- Task should be done independently
- Incentive should be adequate
- Outcome must have quality control

**BubbleView (2017)**



# Readings

## *Reference Material:*

- Human Computation, Edith Law, Luis von Ahn
- Making Better Use of the Crowd: How Crowdsourcing Can Advance Machine Learning Research, Jennifer Wortman Vaughan, Microsoft Research