



# Creating Speech and Language Data With Amazon's Mechanical Turk

Chris Callison-Burch and Mark Dredze

Human Language Technology Center of Excellence & Center for Language and Speech

Processing Johns Hopkins University, NAACL-2010 Workshop

# Introduction

- What kind of research is possible?
- What new tasks should we try to solve?
- How can we ensure high quality?

## Question

How did you convey the task in terms that were simple enough for nonexperts to understand?

Were non-experts as good as experts?

What did you do to ensure quality?

How quickly did the data get annotated?

What is the cost per label?

# Mechanical Turk

The screenshot shows the Amazon Mechanical Turk website interface. At the top, the logo reads "amazonmechanical turk Artificial Intelligence". Navigation tabs include "Your Account", "HITS", and "Qualifications". On the right, there are links for "Already have an account?", "Sign in as a Worker", and "Requester". A secondary navigation bar contains "Introduction", "Dashboard", "Status", and "Account Settings".

The main content area features a yellow banner with the text: "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. 105,755 HITs available. [View them now.](#)"

Below the banner, there are two columns of information:

- Make Money by working on HITs:** This section explains that HITs (Human Intelligence Tasks) are individual tasks. It lists benefits for workers: "Can work from home", "Choose your own work hours", and "Get paid for doing good work". A flow diagram shows "Find an interesting task" (with a circular icon containing text about tasks) leading to "Work" (with a gear icon) and finally "Earn money" (with a dollar sign icon). A "Find HITs Now" button is positioned below the diagram.
- Get Results from Mechanical Turk Workers:** This section instructs requesters to ask workers to complete HITs. It lists benefits for requesters: "Have access to a global, on-demand, 24 x 7 workforce", "Get thousands of HITs completed in minutes", and "Pay only when you're satisfied with the results". A flow diagram shows "Fund your account" (with a credit card icon), "Load your tasks" (with a task list icon), and "Get results" (with a star icon). A "Get Started" button is positioned below the diagram.

At the bottom of the worker section, there is a link: "or [learn more about being a Worker](#)".

# Mechanical Turk

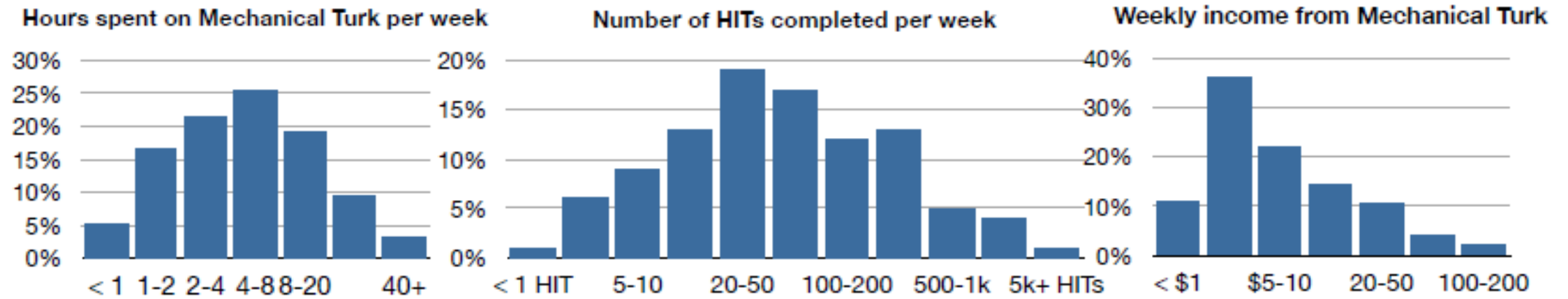


Figure 1: Time spent, HITs completed, and amount earned from a survey of 1,000 Turkers by Ipeirotis (2010).

# Turker demographics

Why do you complete tasks in MTurk?	US	India
To spend free time fruitfully and get cash (e.g., instead of watching TV)	70%	60%
For “primary” income purposes (e.g., gas, bills, groceries, credit cards)	15%	27%
For “secondary” income purposes, pocket change (for hobbies, gadgets)	60%	37%
To kill time	33%	5%
The tasks are fun	40%	20%
Currently unemployed or part time work	30%	27%

Table 1: Motivations for participating on Mechanical Turk from a survey of 1,000 Turkers by Ipeirotis (2010).

# Quality Control

- Amazon provides three mechanisms to help ensure quality:
- Requesters have the option of rejecting the work of individual Turkers, in which case they are not paid. Turkers can also be blocked from doing future work for a requester.
- Requesters can specify that each HIT should be redundantly completed by several different Turkers. This allows higher quality labels to be selected, for instance, by taking the majority label.
- Requesters can require that all workers meet a particular set of qualifications, such as sufficient accuracy on a small test set or a minimum percentage of previously accepted submissions.

# Designing HITs for quality control

- Cheating detection

# Iterative improvements on MTurk

- Quality control on Mechanical Turk is through iterative HITs that build on the output of previous HITs.



# Recommended Practices

- It is critical to convey instructions appropriately for non-experts.
- To calibrate whether the HIT is doable, you should first try the task yourself, and then have a friend from outside the field try it.
- You should insert positive and negative controls so that you can quickly screen out bad Turkers.
- If you are having Turkers evaluate the output of systems, then randomize the order that the systems are shown in.
- you should publish the data that you generate on Mechanical Turk alongside your paper so that other people can verify it.

# Future Directions

- Find a way to improve the cost of Quality Control.
- Use different strategies in different task.
- Use the exist resource to finish the task.