Springer Encyclopedia manuscript No.
(will be inserted by the editor)

Crowdsourcing and Human Computation, Introduction

Matthew Lease · **Omar Alonso**

1 Introduction

The first computers were actually people [10]. Later, machines were built, known at the time as as Automatic Computers (ACs), to perform many routine computations. While such machines have continued to advance and now perform many of the routine processing tasks once delegated to people, human human capabilities still continue to exceed state-of-the-art Artificial Intelligence (AI) on a variety of important data analysis tasks, such as those involving image [33] and language understanding [32]. Consequently, today's Internet-based access to 24/7 online human crowds has sparked the advent of crowdsourcing [12] and a renaissance of human computation [28, 20] relying upon participation of online crowds.

The resulting wealth of new opportunities has brought a disruptive shift to research and practice for how computer scientists are designing and implementing intelligent systems today. Not only can labeled data for training and evaluation be collected faster, cheaper, and easier than ever before, but we now see human computation being integrated into the systems themselves, operating in concert with automation during run-time execution. While AI will certainly continue to improve, strategic use of human computation in tandem with AI and routine data processing enables us to offer greater system capabilities today. As a result, we have suddenly found ourselves occupying a much richer design space for navigating traditional tradeoffs between processing time, cost, effort, or accuracy.

Whereas the traditional mixed-initiative model of human-computer interaction explores the balance between automation vs. user effort, crowdsourcing generalizes this beyond interactions with an individual user to broader social participation. In addition to human-system interactions being driven by an engaged user, we increasingly see interactions where the crowd is engaged by the system to perform work on its behalf. For example, in designing the search engines of tomorrow, the rise of online crowds has created new opportunities to

Matthew Lease
School of Information
University of Texas at Austin
Austin, TX USA
E-mail: ml@ischool.utexas.edu

Omar Alonso
Microsoft Corp.
Mountain View, CA USA
E-mail: Omar.Alonso@microsoft.com

leverage social wisdom and foster dynamic collaborations on-the-fly in order to better address information needs than would be possible using purely automated systems and isolated individuals [11,39].

2 Origins and Forms

Use of people to perform computations has a long history predating the ACs we have built to perform many routine computations today [10]. Jeff Howe coined the term *crowdsourcing* and defined it as the process of taking a job that is traditionally performed by a known agent (often an employee) and outsources it to an undefined, generally large group of people via an open call [12]. The application of open source principles is another way to think about crowdsourcing. Wikipedia is often mentioned as the most representative example of crowdsourcing. We define crowdsourcing as a tool that a human computation system can use to distribute tasks. Next, *human computation* [28,20] describes computation performed by a human, and human computation systems organize human efforts (typically online crowds) to carry out computation.

Many distinct forms of crowdsourcing co-exist under the same umbrella term. For example, *citizen science* represents a movement within scientific fields that seeks to engage non-domain experts in scientific work, e.g., GalaxyZoo (`galaxyzoo.org`) and eBird (`ebird.org`). Many other applications are now being built based on collective intelligence or wisdom of crowds [34]. A wide variety of volunteer-based crowdsourcing initiatives have been explored, such as transcribing historical weather reports (`OldWeather.org`) and ancient manuscripts (`AncientLives.org`). Other models include peer-production and "games with a purpose" (e.g. `gwap.com` [27]. Wikipedia represents one of the most successful examples of complex, quality peer-production [36].

3 Commercial Platforms

While volunteer-driven crowdsourcing initiatives can be very effective, participation can also be limited by a variety of factors. To complement such volunteer-driven initiatives, commercial crowdsourcing services provide another, alternative form of crowdsourcing. When financial incentives are well-aligned to work objectives, industrial crowdsourcing been shown to be very effective in motivating both scale and quality of work performed [27]. *Crowd work* [16] encompasses a range of online work in which tasks are posted online for completion by an online workforce.

In recent years, a large and vibrant pay-based crowdsourcing industry and workforce has emerged (crowdsortium.org) which advertises a wide range of tasks via the Internet to an on-demand, 24/7 global population of workers. In this Internet clearinghouse for online work, global market forces of supply and demand often yield lower labor costs (akin to traditional outsourcing), while Internet infrastructure enables rapid, low-cost distribution and submission of micro-tasks by domestic and international workers alike. By reducing artificial economic friction of physical geography, labor costs vary more naturally as a function of task complexity and any requirements on time and quality. For example, crowdsourcing industry heavy-weights LiveOps (`www.liveops.com`) and eLance (`www.elance.com`) use crowd workers to provide online call center and technical support services for Fortune-500 companies.

One of the most prominent commercial platform for crowdform today, especially for "micro-work", is Amazon Mechanical Turk (AMT, `mturk.com`). Launched in 2005 AMT

provides a massive globally-distributed, online marketplace for work in which *Requesters* post tasks to be completed by *Workers* (aka, *Providers*). By 2007, AMT had grown to encompass 100,000 Workers in more than 100 countries [20]. Today, AMT boasts over 500,000 Workers from 190 countries (`requester.mturk.com/tour`). While the scope of tasks one can post on AMT is largely unrestricted, AMT has predominantly attracted entry-level *micro-tasks* in which human workers engaged in data-processing continue to produce higher quality results than even state-of-the-art, automated methods. In addition to AMT, today's vibrant crowdsourcing industry now boasts a myriad of new vendors offering a wide range of additional features and workflow models for accomplishing complex, quality work. Other popular platforms include: CrowdFlower (formerly Dolores Labs, `www.crowdflower.com`) [32, 21], and MobileWorks (`www.mobileworks.com`) [19].

While these platforms above largely regard crowd workers as interchangeable, some research has reported that establishing long-term relationships with Workers made it unnecessary to apply many of the statistical procedures commonly used for quality assurance on these other platforms [18]. For example, oDesk (`odesk.com`) workers are known to those who hire them for often longer-term relationships. A topic for further research is studying the extent to which supposed anonymity itself may fundamentally contribute toward poor quality work, leading to ephemeral working relationships in place of more enduring ones. While AMT's workforce is not anonymous [22], they have largely been treated as such historically. A past research pilot project [17] explored inviting AMT Workers to voluntarily de-anonymize to establish trust relationships, advertise skills and expertise, and gain access to higher paying work and economic mobility, etc. At the time, this was achieved by letting AMT Workers link their WorkerID to an external identity source (e.g., OpenID, Facebook, Google+, etc.). Now that it is clear Amazon provides every Worker a built-in opportunity to voluntarily de-anonymize by creating an Amazon Profile page [22], these Profile pages could become the LinkedIn equivalent for crowd workers engaged in micro-work. Workers without Profiles could continue to perform entry-level data processing, and those with skills might more easily gain access to work requiring greater expertise. A diverse ecology of work and skills may come to flourish on AMT, in time.

4 Motivating Applications

An ever-growing variety of data processing work is being crowdsourced today. This includes: common tasks (transcription, translation, content generation, and assessment), accelerating innovation and discovery, creating new content, making predictions, fund-raising, information-seeking, and monitoring. While computers have let us automate many routine data processing operations once performed by people, some forms of computation remain difficult or impossible to automate. These include: "AI-hard" content analysis (e.g., understanding text, images, or multi-media), harvesting and digitizing information about the physical world, new tasks which can be more easily and rapidly assigned to people than automated, polling people for their subjective opinions or feedback, etc. Common forms of micro-tasks today include: data verification and correction, information harvesting, collecting objective assessments or subjective opinions, transcription, translation, and content generation (writing text and creating other media).

Eric Raymond once remarked that "Every good work of software starts by scratching a developer's personal itch", characterizing a part of the software development process. Similarly, the use of crowdsourcing in different computer science areas started by a personal itch as well: gathering labels. Designing experiments, recruiting subjects, implementing the necessary infrastructure in place, and collecting meaningful data is an intensive and expensive

process that certain institutions were able to do. Three fields that require enormous amounts of data that lead the adaption of crowdsourcing are natural language processing (NLP) [32], machine translation (MT) [5] and the already mentioned information retrieval (IR) [2, 1].

The research work by Snow et al. [32] shows the quality of workers in the context of four different NLP tasks, namely, affect recognition, word similarity, textual entailment, and event temporal ordering. Machine translation investigates the use of software to translate text or speech from one natural language to another. The work by Calison-Burch [5] has shown that it is possible to produce judgments similar to non-experts and evaluating translation quality through reading comprehension can be done via crowdsourcing. Crowdsourcing techniques have also been used to improve support for humanitarian crises. For example, Munro reports on the findings of Mission 4636, a real-time humanitarian crowdsourcing initiative that processed 80,000 text messages (SMS) sent from within Haiti following the 2010 earthquake [25].

Information retrieval evaluation is an essential part of the development and maintenance of search engines and related systems. Many Web search engines reportedly use large editorial staffs to judge the relevance of web pages for queries in an evaluation set [2, 1]. This is expensive and has obvious scalability issues. Academic researchers, without access to such editors, often rely instead on small groups of student volunteers. Because of the students' limited time and availability, test sets are often smaller than desired, making it harder to detect statistically significant differences in performance by the experimental systems being tested. While behavioral data, such as obtained automatically from search engine logs of user activity, is much cheaper than the editorial method, it requires access to a large stream of data, something not always available to a researcher testing an experimental system. These challenges provide an ideal setting for demonstrating both the potential and practical challenges for the crowdsourcing paradigm.

In addition to collecting labeled data, crowdsourcing can also be used to streamline interactive user studies or increase breadth of demographics vs. what would be available in a traditional small study. For example, Zuccon et al. describe an experimental methodology that can be an alternative to laboratory-based user studies in information retrieval experiments [40]. They show that their crowdsourcing based approach can capture user interactions and searching behaviors at a lower cost, with more data, and within a shorter period than traditional laboratory-based user studies.

5 Understanding The Crowd

Different people are motivated by many different incentives, which can operate independently or in combination. Design of appropriate incentives for the given task at hand is referred to as *incentive engineering*. To motivate people to perform work (and do it well), incentive mechanisms explored inlucde: pay, fun, prestige, socializing, altruism, barter, and learning. For example, with citizen science, scientists or non-profits distribute work to volunteers motivated by altruism, learning, and/or opportunities to socialize (e.g., galaxyzoo.org, ebird.org). Games with a Purpose (`www.gwap.com`), on the other hand, motivate work via opportunities for fun, socialization, and prestige. The re-Captcha project (`recaptcha.net`) re-purposes an existing, ongoing activity (solving captchas) to extract useful work as a by-product. von Ahn's most recent DuoLingo project (duolingo.com) incentivizes translation work by the opportunity to learn a foreign language.

There are many market or research studies for which demographics of participants are valuable to ascertain. For example, the uTest (`utest.com`) crowdsourcing platform requires

its workforce to provide demographic information during registration. uTest customers specify test requirements such as demographics, OS, browser, etc., and uTest proceeds to identify and invite qualified testers from its online community. In contrast, AMT's *seemingly* anonymous workforce Lease-ssrn13 has made access to demographic information far more difficult, requiring Requesters to solicit demographic information directly from Workers [29, 13] for tasks like remote usability testing [23].

On AMT, the individual or organization who has work to be performed is known as the *requester*, while a person who wants to sign up to perform work is described in the system as a *worker* (or *provider*). The unit of work to be performed is called a Human Intelligence Task (HIT). For example, say that a researcher would like to label 100 images and creates an experiment in a crowdsourcing platform. In the platform, the researcher is the requester, each person who is interested in labeling an image is a worker, and the specific task of labeling an image is a HIT.

Many AMT studies have found distribution of HITs completed across the workforce tends to follow a power-law distribution, with a few workers doing much of the work and many workers doing very little work [32, 35]. It is not clear to what extent that is a natural product of online crowd work, of AMT in particular, of the type of tasks being posted, or due to other factors. This has led some requesters to view crowdsourcing not as utilizing a large crowd to perform work, but filtering a large crowd to find a few people to do the work. For statistical approaches to quality assurance, a resulting challenge is estimating quality of individual worker contributions for those workers who contribute few labels, representing a sparse data problem.

6 Task Design

As crowdsourcing is inherently a human-centric enterprise, attention (or inattention) to human factors will clearly impact the quality of data collected from the crowd ([15]). While poor quality of crowd data has been often blamed on lazy and/or ignorant workers, task instructions and/or interfaces poorly designed for non-experts may be equally culpable. Moreover, if task instructions are unclear, or a task interface poorly designed, the inevitable low quality of crowd data can at best be ameliorated by statistical quality assurance methods. In general, approaches to quality assurance based on human factors vs. statistical methods are largely complementary, allowing each to be independently studied in controlled experimentation. However, it is important to recognize that the quality of data collected from the crowd may vary significantly under different task designs. Consequently, investigation of statistical quality assurance algorithms should take into consideration the highly variable quality of crowd data to be encountered in practice. This impacts both comparative benchmarking of alternative techniques, as well as measuring the robustness of any single technique.

A very important aspect of any crowdsourcing activity involves asking the right questions. Intuitively, this step seems very easy and straightforward to implement but in practice it could produce undesirable results if it is not taken seriously. Humans design a task that needs to be completed by another human so a precondition for proper communication is the use of simple language for describing the request and what is expected.

A worker is part of multilingual and multicultural distributed workforce and they need to understand questions consistently. The requester of the task has to make sure that all workers have a shared, common understanding of the meaning of the question. What constitutes a good answer or good work should be communicated to the workers. Examples of good and bad responses are encouraged. There is no need to use specific terminology unless all

workers are expected to be experts on a particular subject. A common mistake when setting up crowdsourcing experiments is to condense too many questions on a single task. At any given day, requesters compete for workers so a simple and precise task is a much better technique for attracting the right crowd. Partitioning a larger task into atomic micro-tasks that can be easily completed in a very short period of time is the right use of the paradigm.

To summarize the main points:

- Present clear and consistent instructions to the workers on what they have to do. Ideally, it should consist of one or two items that need to get done. This is not a lengthy survey. Brevity and simplicity are essential.
- Show examples whenever is possible.
- Make use of use highlighting, bold, italics, typefaces, and color when needed to improve the content presentation. A lot of tasks require reading text so instructions and content have to be legible.
- Minimize the effort to accomplish a task.

7 Quality Assurance

Ensuring the quality of the work and overall worker performance involves an end-to-end commitment to detail and diligence: clear instructions, a well-designed user interface, content quality, inter-rater agreement metrics, and worker feedback analysis [1]. Moreover, designing and implementing experiments that require thousands or millions of labels presents fundamentally different challenges than conducting small scale experiments, and enabling a framework for continuous crowdsourcing experiments requires even more rigorous design. In some cases, the workers may lack sufficient expertise to perform the task well, depending on its complexity. Managing quality control is an on-going activity and different types of checks have to be applied before, during and after a task. Task payment should be done after the work quality has been assessed.

7.1 When to Assess Quality

Beforehand. The purpose of this quality control check is to help on the screening, selection, recruiting, and training of workers before the task is available on a particular crowdsourcing market place. The most common technique used is to apply a qualification test or similar mechanism. A qualification test is very similar to a task and involves asking workers a few questions and qualifying those who pass. Essentially is the equivalent to taking a test. This method is useful for weeding out workers who are performing really poorly or those intentionally committing fraud: *spammers*. The drawbacks are that, like in any test, maintenance is needed and for certain tasks, workers may not be willing to take a qualification test.

During. Assuming that some pre-qualification step was taking into consideration, the purpose of this quality control check is to calibrate, reward, penalize, or weight the answers as workers complete the task. We can accomplish this process by assessing the quality of work (e.g. labels, answers, etc.) as workers produce them. The earlier we can detect that a worker is performing badly, the better. A very popular technique is to use random checks, usually called honey pots (i.e., questions that have a pre-computed answer), to validate workers' performance. The advantage of this technique is that enforces the control over the entire task and does not require a lot of setup.

After. The final control check is to compute accuracy metrics after the task is completed. The purpose of this step is to filter, calibrate, weight, and retain the good workers.

7.2 Measuring Agreement

For a given HIT, multiple workers (usually 3 to 5) will work on each assignment. A benefit of crowdsourcing is the ability to access a large and diverse population. Since many users are sampled with a pool drawn from all over the globe, results found have the potential to generalize, more than traditional small user samples limited geographically. We can think of each answer provided by a worker as vote. Measuring and reporting the inter-rater agreement or reliability of the workers is a desirable step when the requester is computing the final results. There is a number of agreement statistics in the literature; see Artstein and Poesio for a short review [3]. Some of the most widely used agreement statistics include Cohen's kappa, Fleiss' kappa and Krippendor's alpha; all are available in statistical software packages.

7.3 Techniques for Quality Assessment

Statistical methods typically measure the quality of a worker's labels in comparison to one or more of the following: 1) expert labels; 2) other workers' labels; or 3) system-predicted labels. Strategy (1) is most easily understood: given an example with a known gold label, check if a crowd worker's label matches it. Supervised statistical methods are trained on such gold labels to learn a statistical model (estimate of model parameters). The great limitation of **Strategy (1)** is that experts are scarce, increasing time and cost of labeling, and often yielding a shortage of gold labels. Consequently, there has been great interest in developing effective unsupervised methods which limit or entirely avoid dependence on expert gold. However, expert labels are particularly important when: 1) the source example distribution is highly skewed and we want to expose workers to a more balanced distribution [21]; 2) the labor pool is sufficiently poor that even consensus labels may be bad; and 3) we want to detect and calibrate workers labels against systematic bias [37]. Some gold labeling is required simply to define the task and verify the output labels match expectations. Time invested labeling gold also helps to ensure that the task creator has seen enough examples to: i) precisely define the desired mapping workers will perform, ii) write clear instructions, and iii) identify helpful cases to present as exemplars.

Strategy (2). A widely used unsupervised approach is repeated labeling and aggregation. Multiple people are asked to label same the example, and the individual labels are the aggregated induce a single, consensus label [7,30]. This strategy can be generally applied to across tasks: e.g., translation, content generation, or even guessing the weight of an ox [34]. The simplest method of aggregation is majority voting, which simply selects the most frequent label. Weighted voting, on the other hand, assigns greater weight to labels produced by more trusted workers. In general, the key idea of repeated labeling is to reduce variance by combining a diverse set of independent responses. A similar principle underlies prediction markets (e.g., `intrade.com`) and the wisdom of crowds [34], as well as use of ensemble methods for integrating multiple statistical models. In general, effectiveness in practice often depends on how well assumptions actually hold. For example, crowd labels may lack independence due to benign causes, such as bias, or adversarial causes, such as collusion or "Sybil attacks" involving misuse of worker identities.

Strategy (3) involves comparing worker labels to system-predicted labels produced by a rule-based system or statistical model. Such a hybrid system integrates some combination of human and system labels to produce output labels. Hybrid systems require a machine-readable representation of source examples (features) that can be processed by the system, as well as an algorithm for mapping example features to an output label. A strength of hybrid systems is the ability to employ automated label generation alongside crowdsourcing to: 1) reduce the number of examples sent to human workers (e.g., when system is already confident in a predicted label); 2) simplify the human processing task by decomposing the problem space or reducing the set of candidate outputs to choose between [39]; 3) aggregate predicted labels with human labels to increase system accuracy beyond what either man or machine could achieve alone; 4) automatically predict labels in order to estimate quality of worker labels; and 5) predict latent worker labels for examples not labeled by a given worker. Whereas hybrid approaches use human labeling at run-time, active learning uses human labeling to train a system.

8 Broader Context

As seen with earlier outsourcing, global market forces are increasingly moving computer work to regions of the world where it can be completed more quickly and affordably. Crowd work savings arise from increase in labor supply, lower cost of living in other geographic regions, and the ability to decompose work into very fine-granularity units which can be efficiently and affordably distributed. While early demographic studies suggested that crowd work was typically performed for supplemental rather than primary income, subsequent studies have indicated crowd work is increasingly become a source of primary income, especially in developing economies [29, 13]. Relatively low wages, depersonalized work, and asymmetric power relationships have led some individuals to raise ethical concerns that we may be building a future of crowd-powered computing on the backs of exploited workers in digital sweatshops[8]. At the same time, crowdsourcing is conversely being seen as "The New Sewing Machine" [26], creating new opportunities for income and social mobility in regions of the world where local economies are stagnant and local governmental structures may discourage traditional outsourcing firms. Just as consumers can choose to buy fair trade goods or invest in social choice funds, some crowdsourcing services now offer guarantees of worker protections and living wages: SamaSource (`samasource.org`, MobileWorks, and CloudFactory (`cloudfactory.com`). Recently filed litigation questions whether individuals engaged in certain forms of crowd work should be legally classified as employees rather than independent contractors under the Fair Labor Standards Act (FLSA), a direction of possible regulation only speculated upon earlier [38]

While optimizing worker behaviors [24] and investigating technological opportunities and challenges is clearly important and valuable work, how might we best wrestle also with questions about what is ethical, legal, and sustainable economic practice in crowd work [31, 9]. For example, is it truly preferable to pay people nothing, i.e., having workers play online games which generate work products as an output (which the workers may not be aware of, and which may generate revenue which the workers do not benefit from) [20], than to pay them low wages and clearly communicate they are performing work [9, 8].

8.1 The Danger of Abstraction

Computer Scientists love abstractions that allow us integrate diverse software modules while encapsulating pesky details of the modules' internal characteristics. AMT's novel design lets us write programmatic function calls that look like we are calling upon any other Artificial Intelligence (AI) subroutine, except that the results often outperform those of our traditional AI modules. Crowdsourcing poses a particular danger with regard to the tendency to abstract: we may forget there are real people behind the abstraction, we may impact their lives in ways that do not penetrate the abstraction, and we may not realize or fully appreciate those impacts we are having. A variety of terminology has been coined in order to provide useful computational descriptions for task execution by crowds (e.g., *human computation* [28, 20], "Human Processing Units (HPUs)"[6], "Remote Person Calls (RPCs)" [4], and "the Human API" [14]. While such terminology can be very helpful to us in conceptualizing how task execution by crowds can be effectively integrated with processing by fully-automated algorithms, this same terminology may also serve to perpetuate the invisibility of a global workforce that is by its very distributed nature difficult to put a face on. As has been succinctly noted elsewhere, "abstraction hides detail" [31]: some details may be worth keeping conspicuously present.

To help us understand crowd workers in a way that statistics on crowd demographics do not seem to make as visceral to us, Andy Baio created a collage of worker faces (`waxy.org/2008/11/the_faces_of_mechanical_turk`). Leila Chirayath Janah, who founded the non-profit SamaSource platform for crowd work, regularly gives talks showing people living in African refugee camps performing online crowd work as one of their only opportunities to earn income and exert some measure of control in otherwise chaotic living environments. As a form of design activism, Lilly Irani and collaborators built Turkopticon, combatting the invisibility of crowd workers and raising collective awareness of worker concerns [31, 14]. Despite such efforts, we still know relatively little about the lives and conditions of the many crowd workers powering today's crowdsourcing applications. When we observe only the work products and remain ignorant of its source, we risk assuming a worker's economic or privacy decision is fully informed and freely chosen. We remain unaware of which worker decisions truly are free and informed.

8.2 An Opportunity to Raise Awareness

While exciting applications of crowd work exist with great potential to transform and advance our society, less ideal uses of crowd work also exist. Consider the case of outsourcing of dirty digital jobs, an online equivalent of the traditional practice of offloading of locally undesirable jobs to immigrant workers. Because it is not acceptable for a social network customer to see a pornographic or violent image posted via social media, we might instead hire crowd workers to sift through such images click after click so we are not exposed to it ourselves (`http://www.buzzfeed.com/reyhan/tech-confessional-the-googler-who-looks-at-the-wo`). While such a content moderation task should ideally be automated — after all, computers cannot become emotionally disturbed from constant exposure to such imagery — our best state-of-the-art AI software is unfortunately not effective enough to flag all of the "muck" that gets posted online. So we utilize *human computation* instead via crowd work. Just as janitorial work cleans our local restrooms, crowd work cleans our social networks. To the extent such forms of crowd work are valued and needed, how might we better support and reduce risk to those performing the work?

In another vein, human history shows us that when one group has power over another, with money or other incentives at stake, the lesser group may be compelled to perform work they would otherwise not choose. The recent case of prisoners compelled to perform gold farming in online games is a modern example. As the industry of crowd work grows, these pressures will continue to grow in force. How might we monitor and mitigate the corresponding growth of such practices with at-risk populations?

9 Future Directions

Crowdsourcing continues to rapidly evolve, with increasing impact on both industrial and research practice. For many, crowdsourcing may be interesting only so far as it enables them to better advance their research programs and gain new traction on old problems. For others, crowdsourcing has become a phenomenon to study in its own right, with inter-disciplinary issues spanning engineering, psychology, sociology, economics, policy, and ethics (at least).

While the previous section's discussion of socio-technical issues may seem extraneous to some readers more interested in crowdsourcing's utility, it is precisely this often exclusive focus on utility that should give us pause (a focus that remains widespread across technical research on crowdsourcing today). Our impact extends beyond our technical contributions, and we are affecting the real human lives powering our human computation systems today. There are tremendously exciting applications of crowdsourcing emerging with great potential to transform and advance our society for the betterment of all. We scientists can play a significant, positive role in helping shape this future.

One of the authors of this article a simple Webpage which tracks related academic research activities, such as conferences, workshops, journals, etc. (`ir.ischool.utexas.edu/crowd`). The CHI community maintains an active shared blog which provides a focal point for promoting community and dissemination of related crowdsourcing research and activities (`crowdresearch.org/blog`). In the KDD and AAAI areas, four years of a Human Computation workshop are now giving rise to a new AAAI Conference on Human Computation (`www.humancomputation.com/2013`). This year will also mark the fourth industrial CrowdConf Conference (`www.crowdconf.com`). A broader Collective Intelligence conference held in 2012 (`www.ci2012.org`) may occur again. Beyond this, a number of Special Issues now exist or are forthcoming: IEEE Internet Computing, Springer Information Retrieval, Springer Machine Learning, VLDB Journal, etc.

Conflict-of-Interest Disclosure

The first author has several ties to AMT and the crowdsourcing industry, having: 1) received Amazon Web Services research awards for use of Amazon's cloud computing; 2) having received Amazon sponsorship funds for the National Institute of Standards and Technology (NIST) Text REtrieval Conference (TREC) Crowdsourcing Track (`https://sites.google.com/site/treccrowd`); and 3) having spent a mini-sabbatical working in residence at CrowdFlower.

Acknowledgements We thank Jessica Hullman for her thoughtful comments and editing regarding broader impacts of crowdsourcing [22]. We also thank AMT personnel for the very useful platform they have built and their clear interest in supporting academic researchers using AMT. Last but not least, we thank the global crowd of individuals who have contributed and continue to contribute to crowdsourcing projects world-wide. Thank you for making crowdsourcing possible.

Matthew Lease was supported in part by an NSF CAREER award, a DARPA Young Faculty Award N66001-12-1-4256, and a Temple Fellowship. Any opinions, findings, and conclusions or recommendations expressed in this paper are those of the authors alone and do not express the views of any of the funding agencies.

Cross-References

References

1. Alonso, O.: Implementing crowdsourcing-based relevance experimentation: an industrial perspective. Information Retrieval Journal, Special Issue on Crowdsourcing (2012)
2. Alonso, O., Rose, D.E., Stewart, B.: Crowdsourcing for relevance evaluation. ACM SIGIR Forum **42**(2), 9–15 (2008)
3. Artstein, R., Poesio, M.: Inter-coder agreement for computational linguistics. Computational Linguistics **34**(4), 555–596 (2008)
4. Bederson, B.B., Quinn, A.J.: Web workers unite! addressing challenges of online laborers. In: CHI Workshop on Crowdsourcing and Human Computation. ACM (2011)
5. Callison-Burch, C.: Fast, cheap, and creative: evaluating translation quality using amazon's mechanical turk. In: Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 1-Volume 1, pp. 286–295. Association for Computational Linguistics (2009)
6. Davis, J., Arderiu, J., Lin, H., Nevins, Z., Schuon, S., Gallo, O., Yang, M.: The HPU. In: Computer Vision and Pattern Recognition Workshops (CVPRW), pp. 9–16 (2010)
7. Dawid, A.P., Skene, A.M.: Maximum likelihood estimation of observer error-rates using the em algorithm. Applied Statistics pp. 20–28 (1979)
8. Felstiner, A.: Sweatshop or paper route?: Child labor laws and in-game work. In: Proceedings of the 1st Annual Conference on the Future of Distributed Work (CrowdConf). San Francisco (2010)
9. Fort, K., Adda, G., Cohen, K.B.: Amazon mechanical turk: Gold mine or coal mine? Computational Linguistics **37**(2), 413–420 (2011)
10. Grier, D.A.: When computers were human, vol. 316. Princeton University Press (2005)
11. Horowitz, D., Kamvar, S.D.: The anatomy of a large-scale social search engine. In: Proceedings of the 19th international conference on World wide web, pp. 431–440. ACM (2010)
12. Howe, J.: The rise of crowdsourcing. Wired magazine **14**(6), 1–4 (2006)
13. Ipeirotis, P.: Demographics of Mechanical Turk. Tech. Rep. CeDER-10-01, New York University (2010)
14. Irani, L., Silberman, M.: Turkopticon: Interrupting worker invisibility in amazon mechanical turk. In: Proceeding of the ACM SIGCHI Conference on Human Factors in Computing Systems (2013)
15. Kazai, G., Kamps, J., Milic-Frayling, N.: An analysis of human factors and label accuracy in crowdsourcing relevance judgments. Information Retrieval Journal, Special Issue on Crowdsourcing (2012)
16. Kittur, A., Nickerson, J.V., Bernstein, M., Gerber, E., Shaw, A., Zimmerman, J., Lease, M., Horton, J.: The Future of Crowd Work. In: Proceedings of the ACM Conference on Computer Supported Cooperative Work (CSCW), pp. 1301–1318 (2013)
17. Klinger, J., Lease, M.: Enabling trust in crowd labor relations through identity sharing. In: Proceedings of the 74th Annual Meeting of the American Society for Information Science and Technology (ASIS&T), pp. 1–4 (2011)
18. Kochhar, S., Mazzocchi, S., Paritosh, P.: The anatomy of a large-scale human computation engine. In: Proceedings of the ACM SIGKDD Workshop on Human Computation, pp. 10–17. ACM (2010)
19. Kulkarni, A., Gutheim, P., Narula, P., Rolnitzky, D., Parikh, T., Hartmann, B.: Mobileworks: Designing for quality in a managed crowdsourcing architecture. IEEE Internet Computing **16**(5), 28 (2012)
20. Law, E., L. von Ahn: Human computation. Synthesis Lectures on Artificial Intelligence and Machine Learning **5**(3), 1–121 (2011)
21. Le, J., Edmonds, A., Hester, V., Biewald, L.: Ensuring quality in crowdsourced search relevance evaluation: The effects of training question distribution. In: SIGIR 2010 workshop on crowdsourcing for search evaluation, pp. 21–26 (2010)
22. Lease, M., Hullman, J., Bigham, J.P., Bernstein, M.S., Kim, J., Lasecki, W.S., Bakhshi, S., Mitra, T., Miller, R.C.: Mechanical turk is not anonymous. In: Social Science Research Network (SSRN) Online (2013). URL `http://ssrn.com/abstract=2228728`. SSRN ID: 2228728

23. Liu, D., Bias, R., Lease, M., Kuipers, R.: Crowdsourcing for usability testing. In: Proceedings of the 75th Annual Meeting of the American Society for Information Science and Technology (ASIS&T) (2012)
24. Mason, W., Watts, D.J.: Financial incentives and the Performance of Crowds. In: SIGKDD (2009)
25. Munro, R.: Crowdsourcing and the crisis-affected community lessons learned and looking forward from mission 4636. Information Retrieval Journal, Special Issue on Crowdsourcing (2012)
26. Paritosh, P., Ipeirotis, P., Cooper, M., Suri, S.: The computer is the new sewing machine: benefits and perils of crowdsourcing. In: Proceedings of the 20th international conference companion on World wide web, pp. 325–326. ACM (2011)
27. Pickard, G., Pan, W., Rahwan, I., Cebrian, M., Crane, R., Madan, A., Pentland, A.: Time-critical social mobilization. Science **334**(6055), 509–512 (2011)
28. Quinn, A.J., Bederson, B.B.: Human computation: a survey and taxonomy of a growing field. In: 2011 Annual ACM SIGCHI conference on Human factors in computing systems, pp. 1403–1412 (2011)
29. Ross, J., Irani, L., Silberman, M., Zaldivar, A., Tomlinson, B.: Who are the crowdworkers?: shifting demographics in mechanical turk. In: Proceedings of the 28th of the international conference extended abstracts on Human factors in computing systems, pp. 2863–2872. ACM (2010)
30. Sheng, V., Provost, F., Ipeirotis, P.: Get another label? improving data quality and data mining using multiple, noisy labelers. In: Proceeding of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining, pp. 614–622 (2008)
31. Silberman, M., Irani, L., Ross, J.: Ethics and tactics of professional crowdwork. XRDS: Crossroads, The ACM Magazine for Students **17**(2), 39–43 (2010)
32. Snow, R., O'Connor, B., Jurafsky, D., Ng, A.Y.: Cheap and fast—but is it good?: evaluating non-expert annotations for natural language tasks. In: Proceedings of the Conference on Empirical Methods in Natural Language Processing, pp. 254–263. Association for Computational Linguistics (2008)
33. Sorokin, A., Forsyth, D.: Utility data annotation with amazon mechanical turk. In: Computer Vision and Pattern Recognition Workshops, 2008. CVPRW'08. IEEE Computer Society Conference on, pp. 1–8. IEEE (2008)
34. Surowiecki, J.: The wisdom of crowds. Anchor (2005)
35. Tang, W., Lease, M.: Semi-Supervised Consensus Labeling for Crowdsourcing. In: Proceedings of the ACM SIGIR Workshop on Crowdsourcing for Information Retrieval (2011)
36. Viégas, F., Wattenberg, M., Mckeon, M.: The hidden order of wikipedia. Online communities and social computing pp. 445–454 (2007)
37. Wang, J., Ipeirotis, P., Provost, F.: Managing crowdsourcing workers. In: The 2011 Winter Conference on Business Intelligence (2011)
38. Wolfson, S., Lease, M.: Look Before You Leap: Legal Pitfalls of Crowdsourcing. In: Proceedings of the 74th Annual Meeting of the American Society for Information Science and Technology (ASIS&T) (2011)
39. Yan, T., Kumar, V., Ganesan, D.: CrowdSearch: exploiting crowds for accurate real-time image search on mobile phones. In: Proceedings of the 8th International Conference on Mobile Systems, Applications, and Services (MOBISYS), pp. 77–90. ACM (2010)
40. Zuccon, G., Leelanupab, T., Whiting, S., Yilmaz, E., Jose, J.M., Azzopardi, L.: Crowdsourcing interactions: using crowdsourcing for evaluating interactive information retrieval systems. Information Retrieval Journal, Special Issue on Crowdsourcing (2012)

Recommended Reading

1. Jeff Barr and Luis Felipe Cabrera. AI Gets a Brain. *Queue*, 4(4):24–29, 2006.
2. Benjamin B. Bederson and Alex Quinn. Participation in human computation. In *CHI Workshop on Crowdsourcing and Human Computation*. ACM, 2011.
3. R.M. Bell and Y. Koren. Lessons from the Netflix prize challenge. *ACM SIGKDD Explorations Newsletter*, 9(2):75–79, 2007.
4. Yochai Benkler. Coase's penguin, or, linux and" the nature of the firm". *Yale Law Journal*, pages 369–446, 2002.
5. Susan L Bryant, Andrea Forte, and Amy Bruckman. Becoming wikipedian: transformation of participation in a collaborative online encyclopedia. In *Proceedings of the 2005 international ACM SIGGROUP conference on Supporting group work*, pages 1–10. ACM, 2005.
6. Jenny J Chen, Natala J Menezes, Adam D Bradley, and TA North. Opportunities for crowdsourcing research on amazon mechanical turk. In *CHI Workshop on Crowdsourcing and Human Computation*, 2011.

7. Ed H Chi and Michael S Bernstein. Leveraging online populations for crowdsourcing: Guest editors' introduction to the special issue. *IEEE Internet Computing*, 16(5):10–12, 2012.
8. Ellen Cushing. Amazon mechanical turk: The digital sweatshop. *UTNE Reader*, January/February 2013. `www.utne.com/science-technology/amazon-mechanical-turk-zm0z13jfzlin.aspx`.
9. danah boyd. What is the role of technology in human trafficking?, December 7, 2011. `http://www.zephoria.org/thoughts/archives/2011/12/07/tech-trafficking.html`.
10. Ofer Dekel and Ohad Shamir. Learning to classify with missing and corrupted features. In *Proceedings of the 25th international conference on Machine learning*, pages 216–223. ACM, 2008.
11. Catherine Grady and Matthew Lease. Crowdsourcing document relevance assessment with mechanical turk. In *Proceedings of the NAACL HLT 2010 Workshop on Creating Speech and Language Data with Amazon's Mechanical Turk*, pages 172–179, Los Angeles, June 2010. Association for Computational Linguistics.
12. B. Hecht, J. Teevan, M.R. Morris, and D. Liebling. Searchbuddies: Bringing search engines into the conversation. *Proceedings of ICWSM 2012*, 2012.
13. Alan Irwin. Constructing the scientific citizen: Science and democracy in the biosciences. *Public Understanding of Science*, 10(1):1–18, January 2001.
14. Matthew Lease and Emine Yilmaz. Crowdsourcing for Information Retrieval: Introduction to the Special Issue. *Information Retrieval*, 16(4), 2013.
15. Brian Neil Levine, Clay Shields, and N Boris Margolin. A survey of solutions to the sybil attack. Technical report, University of Massachusetts Amherst, Amherst, MA, 2006.
16. Richard McCreadie, Craig Macdonald, and Iadh Ounis. Crowdterrier: Automatic crowdsourced relevance assessments with terrier. In *Proceedings of the 35th international ACM SIGIR conference on Research and development in information retrieval*, pages 1005–1005. ACM, 2012.
17. Stewart Mitchell. Inside the online sweatshops. *PC Pro Magazine*, 2010. August 6. `www.pcpro.co.uk/features/360127/inside-the-online-sweatshops`.
18. Prayag Narula, Philipp Gutheim, David Rolnitzky, Anand Kulkarni, and Bjoern Hartmann. Mobileworks: A mobile crowdsourcing platform for workers at the bottom of the pyramid. In *AAAI Human Computation Workshop*, 2011.
19. David Oleson, Alexander Sorokin, Greg Laughlin, Vaughn Hester, John Le, and Lukas Biewald. Programmatic gold: Targeted and scalable quality assurance in crowdsourcing. In *AAAI Workshop on Human Computation*, 2011.
20. Jason Pontin. Artificial intelligence, with help from the humans. *New York Times*, March 25, 2007.
21. Aaron Shaw. Some initial thoughts on the otey vs. crowdflower case, January 9, 2013. `http://fringethoughts.wordpress.com/2013/01/09/some-initial-thoughts-on-the-otey-vs-crowdflower-case/`.
22. Padhraic Smyth, Usama Fayyad, Michael Burl, Pietro Perona, and Pierre Baldi. Inferring ground truth from subjective labelling of venus images. *Advances in neural information processing systems*, pages 1085–1092, 1995.
23. Besiki Stvilia, Michael B Twidale, Linda C Smith, and Les Gasser. Information quality work organization in wikipedia. *Journal of the American society for information science and technology*, 59(6):983–1001, 2008.
24. C.R. Sunstein. *Infotopia:How Many Minds Produce Knowledge*. Oxford University Press, USA, 2006.
25. Danny Vincent. China used prisoners in lucrative internet gaming work. *The Guardian*, May 25, 2011.
26. Luis von Ahn. *Human Computation*. PhD thesis, Carnegie Mellon University, 2005. Tech. Report CMU-CS-05-193.
27. Luis Von Ahn and Laura Dabbish. Designing games with a purpose. *Communications of the ACM*, 51(8):58–67, 2008.
28. Luis Von Ahn, Benjamin Maurer, Colin McMillen, David Abraham, and Manuel Blum. recaptcha: Human-based character recognition via web security measures. *Science*, 321(5895):1465–1468, 2008.
29. Hanna Wallach and Jennifer Wortman Vaughan. Workshop on Computational Social Science and the Wisdom of Crowds. In *NIPS*, 2010.
30. Fabian L Wauthier and Michael I Jordan. Bayesian bias mitigation for crowdsourcing. In *Proc. of NIPS*, 2011.