Chapter 5
Hybrid Machine-Crowd Interaction for Handling Complexity: Steps Toward a Scaffolding Design Framework

António Correia, Shoaib Jameel, Hugo Paredes, Benjamim Fonseca and Daniel Schneider

Abstract Much research attention on crowd work is paid to the development of solutions for enhancing microtask crowdsourcing settings. Although decomposing difficult problems into microtasks is appropriate for many situations, several problems are non-decomposable and require high levels of coordination among crowd workers. In this chapter, we aim to gain a better understanding of the macrotask crowdsourcing problem and the integration of crowd-AI mechanisms for solving complex tasks distributed across expert crowds and machines. We also explore some design implications of macrotask crowdsourcing systems taking into account their scaling abilities to support complex work in science.

5.1 Introduction

In recent years, we have seen a flourishing of crowd-powered systems intended to support computer-hard tasks that cannot be solved by simple machine algorithms (Li et al. 2016). A large body of work exists around the integration of human inputs into microtask crowdsourcing environments (Lasecki 2014). Consistently, many studies attempt to tackle tasks that can be easily decomposed into simpler subtasks and accomplished independently (Cheng et al. 2015). With the growth of expert crowdsourcing settings comprising non-decomposable macrotasks, there is an increasing need to support complex work (Schmitz and Lykourentzou 2018). Such open-ended worker inputs often implicate a high level of dependency and expertise (Zakaria and

A. Correia (✉) · H. Paredes · B. Fonseca
University of Trás-os-Montes e Alto Douro, UTAD, Quinta de Prados, Apartado 1013, Vila Real, Portugal
e-mail: antonio.g.correia@inesctec.pt

INESC TEC, Porto, Portugal

A. Correia · S. Jameel
University of Kent, School of Computing, Medway Campus, Canterbury, UK

D. Schneider
Tércio Pacitti Institute of Computer Applications and Research (NCE), Federal University of Rio de Janeiro, Rio de Janeiro, Brazil

V.-J. Khan et al. (eds.), *Macrotask Crowdsourcing*,
Human–Computer Interaction Series, https://doi.org/10.1007/978-3-030-12334-5_5

Abdullah 2018). In particular, macrotasking projects go beyond data processing to produce new information through social interaction among crowd members (Walsh et al. 2014). The crowdsourcing tasks in this line of work need contextual information and can imply overheads regarding the increasing levels of coordination required to generate information socially. Haas and colleagues (2015) go even further by arguing that "a key challenge in macrotask-powered work is evaluating the quality of a worker's output" due to the absence of an aggregation method in which the inputs of the crowd can be easily combined and evaluated. In this sense, we should state at the outset that supporting macrotasks is particularly challenging and there is still a need to identify new pathways along which complex crowd work can be effectively accomplished.

This work furthers an existing strand of research in leveraging collaborative efforts between humans and machine agents to handle the complexity of the work that can be performed by IT-mediated crowds in science. Crowdsourcing has been successfully used as a tool for supporting scientific research (Law et al. 2017), and research problems of massive scale can be distributed among a sizeable pool of experts and volunteers who contribute actively by handling massive quantities of assorted data (Hochachka et al. 2012). Researchers attempting to perform complex scientific tasks (e.g., systematic literature reviews) usually decompose them into smaller, more manageable chunks of work that can be used to generate training data for AI algorithms (Krivosheev et al. 2018). Such small-scale scientific work settings must be further expanded to incorporate the untapped potentials of combining crowd interactions with automated reasoning at a large-scale given the value of the crowd-AI integration to produce large amounts of data and attain novel discoveries on multivariate topics. Adding on to this line of inquiry, this chapter explores some theoretical underpinnings of crowd-AI hybrids in the context of complex work while depicting a research agenda with a vast set of gaps reported in the literature.

The rest of this chapter proceeds as follows. In Sect. 5.2 we present some background on macrotask crowdsourcing and hybrid machine-crowd interaction in the context of scientific work. In this section, we also illustrate examples of current systems and frameworks intended to support expert crowdsourcing. In Sect. 5.3, we describe some design claims and general aspects of crowdsourcing and AI applications found in the literature. We close in Sect. 5.4 with some remarks and future directions on the combination of crowd-computing hybrids.

5.2 Macrotask Crowdsourcing in Science: From HCI to Hybrid Crowd-Machine Applications

As the number of publications continues to increase, discovery, and acquisition of useful scholarly data from academic literature impose several challenges (Dong et al. 2017). In addition, a large amount of resources are usually spent on research practices (Rigby 2009). Crowd science can be a trustworthy solution for tackling scientific problems that are beyond the capabilities of computer algorithms by engaging academic researchers and nonprofessional scientists (Franzoni and Sauermann 2014). Although several research studies have demonstrated the potential uses of crowdsourcing in science, many researchers are still reluctant regarding the adoption of crowdsourcing (Law et al. 2017). Researchers have been studying crowdsourcing as a way to reduce the cost and speed of a research project while enhancing the quality of the work (Ranard et al. 2014; Tsueng et al. 2016). On the other hand, reviews of the prior research on crowdsourcing show that there are some challenges on scaling up complexity maintaining high-quality responses (Barowy et al. 2012). Human-Computer Interaction (HCI) reaching a high level of engagement over time is another concern in crowd science (Nov et al. 2014). Past research in HCI has explored the use of platforms like Amazon Mechanical Turk (AMT)[1] for crowdsourcing research. For example, Good et al. (2014) recruited nonscientists to identify disease concepts in biomedical paper abstracts and showed that crowd-powered systems can be a reliable instrument for creating annotated corpora. Basing their approach on the general assumption that crowd annotations can be of equal (or even better) value when compared to experts, several authors have used AMT to systematically evaluate scientific literature (e.g., Brown and Allison 2014; Mortensen et al. 2017; Krivosheev et al. 2017). Nevertheless, very little is known about the adoption of alternative platforms such as Prolific Academic[2] and Crowdcrafting[3] for crowdsourcing research (Peer et al. 2017). While this is an obvious limitation, there are several reasons why this fact may be acceptable. In comparison to other crowdsourcing platforms used for research, these platforms usually lack a large and active user base and a suitable API to programmatically access the platform's functionalities.

As previously noted, crowdsourcing tasks can be categorized into microtasks and macrotasks (Luz et al. 2015). Microtask-level settings are characterized by repetitive tasks that are simple for individuals to perform (e.g., image labeling). Such tasks comprise context-free units of work, do not require special skills, and the reward for each task is usually small (Xie and Lui 2018). In macrotasking, requesters create high-level tasks without microtask decomposition while paying workers fair hourly wages (Marcus and Parameswaran 2015). In the literature, there are several examples of expert crowdsourcing systems and general online macrotask-powered work platforms (see Table 5.1). As the table shows, these tools differ from microtasking platforms due to their focus on solving innovative and complex tasks that require high levels of expertise to complete. In contrast to AMT, expert crowdsourcing platforms allow requesters and workers to participate in persistent one-on-one discussions (Salehi et al. 2017). The macrotasks supported by these platforms are usually freeform and large in the sense that they need a vast amount of time to complete.

Macrotasks have particular dependencies, changing requirements, and require expert skills and varied types of expertise. In addition, they are socially mediated in the sense that they require collaboration and may take more time to complete

[1] https://www.mturk.com/.

[2] https://prolific.ac/.

[3] https://crowdcrafting.org/.

Table 5.1 Examples of crowd-powered systems and frameworks for supporting macrotasks

	System	*Description*	*Reference*
Global online macrotask powered work platforms and commercial products	Fiverr	*A platform for outsourcing challenging and innovative tasks*	Xie & Lui (2018)
	Upwork	*Generic online outsourcing marketplace for creative tasks*	Marcus & Parameswaran (2015)
	OpenIDEO	*Social innovation platform for collaboratively tackling global issues*	Schmitz & Lykourentzou (2016)
	Freelancer	*Global online work platform for freelancers*	Borromeo & Toyama (2016)
	Quirky	*Community-led invention platform for product design*	Schmitz & Lykourentzou (2016)
	Science Exchange	*Outsourcing platform for solving scientific problems through a network of qualified crowd workers*	Yan et al. (2016)
	Crowdspring	*Online marketplace for crowdsourced creative designs and ideas*	Schmitz & Lykourentzou (2016)
	InnoCentive	*Open innovation marketplace for tackling complex problems*	Sieg et al. (2010)
Expert crowdsourcing systems	Argonaut	*Framework that uses hierarchical review to improve complex work*	Haas et al. (2015)
	CrowdWeaver	*A system to track crowd workers and task progress*	Kittur et al. (2012)
	Crowd4U	*A prototype system for the deployment of collaborative tasks*	Ikeda et al. (2016)
	CrowdSCIM	*Allows novice crowd workers to learn historical thinking skills while completing historical research tasks*	Wang et al. (2018)
	CrowdForge	*A framework for performing complex work and interdependent tasks*	Kittur et al. (2011)
	Data Tamer	*Expert crowdsourcing system for handling uncertainties in entity resolution and schema integration*	Stonebraker et al. (2013)
	Prism	*Allows a user to upload text documents for collective interpretation*	Walsh et al. (2014)
	MobileWorks	*Operates as an algorithmically managed service, routing work to qualified participants*	Kulkarni et al. (2012)
	Crowd	*A platform for crowdsourcing complex work where one can submit a workflow*	Chettih et al. (2014)
	WearWrite	*Allows a user to perform complex tasks (e.g., writing) from a wearable device*	Nebeling et al. (2015)
	Wish	*A system that uses expert crowd members to carry out complex (creative) tasks*	Kulkarni et al. (2014)
Crowd-AI hybrid systems in science	PANDA	*Hybrid, crowd-computing system for academic knowledge discovery and acquisition*	Dong et al. (2017)
	Solvent	*Mixed-initiative system for finding analogies between research papers*	Chan et al. (2018)
	CrowdRev	*A platform for crowd-enabled screening of systematic literature reviews*	Ramirez et al. (2018)
	SciCrowd	*Crowd-AI system for supporting research work in academic settings*	Correia et al. (2018a)
	Apolo	*A mixed-initiative system that interactively explores large networks of scientific data*	Chau et al. (2011)

(Schmitz and Lykourentzou 2018). In macrotasking settings, requesters reward workers according to the quality of the evaluated solution. In this sense, a requester will only give a large reward to a worker if the quality of the solution is high. Workflows are needed to facilitate the decomposition of tasks into subtasks, management of dependencies between subtasks, and assembly of results (Kittur et al. 2013). Current approaches do not encompass human factors in assessing the quality of the solution, do not address the challenge of free riding of workers, nor denial of payment of requesters (Xie and Lui 2018). Macrotask crowdsourcing for complex work cannot be realized by using simple parallel approaches like aggregating multiple independent judgments through voting since macrotasks are difficult to be decomposed and require sharing of contextual information. As argued by Niu et al. (2018), a crowd may need to build its own team for solving complex tasks.

Research is beginning to emerge in exploring ways to optimize macrotasking scenarios. Retelny and co-workers (2014) proposed *flash teams*, a framework that relies on expert crowdsourcing for solving tasks that require deep domain knowledge. Recently, Valentine and colleagues (2017) extended this expert crowd work framework to *flash organizations*, an approach where crowds are "structured like organizations to achieve complex and open-ended goals". CrowdForge (Kittur et al. 2011) is another example of a framework for executing complex tasks that incorporates some level of automation in the coordination process (Garcia-Molina et al. 2016). In the same vein, Prism (Walsh et al. 2014) was introduced as a system with a shared digital space in which crowd workers can provide creative contributions and interpretations of texts. Argonaut (Haas et al. 2015) is perhaps one of the most widely known examples of a macrotask crowdsourcing system introduced in the literature. The system is intended to support context-aware data processing tasks through a hierarchical review scheme. Platforms such as Crowd (Chettih et al. 2014), Wish (Kulkarni et al. 2014), MobileWorks (Kulkarni et al. 2012), and Data Tamer (Stonebraker et al. 2013) also represent the vast range of solutions that leverage a crowd of domain experts to carry out macrotasks.

The ongoing stream of publications about macrotasking also suggests the use of such applications for learning and research purposes. Crowd4U (Morishima et al. 2012) is a complex data-centric crowdsourcing system that supports collaborative tasks by enabling task decomposition and assignment. Furthermore, CrowdSCIM (Wang et al. 2018) enables a vast set of macrotasks to improve historical research tasks without feedback or intervention from other crowd members. To achieve the full potential of crowdsourcing in science, HCI researchers have also shown a variety of scenarios in which crowd members can be engaged in advanced research tasks such as writing a paper (Gaikwad et al. 2016; Whiting et al. 2017; Crowston et al. 2018). There are other examples of hybrid crowd-AI systems proposed for supporting complex scientific tasks, as can be seen in Table 5.1. To tackle the problem of academic knowledge acquisition, PANDA (Dong et al. 2017) combines hybrid algorithmic-crowdsourcing techniques, while SciCrowd (Correia et al. 2018a) supports research groups on data-driven research tasks (e.g., annotation of large amounts of HCI publications) taking into account a particular research question instead of a simple search for terms. Concomitantly, in research, we have seen systems where humans can annotate aspects of research papers (e.g., findings) in order to find analogies through a computational model (Chan et al. 2018). Others in the community have studied how to combine machine and crowd intelligence in systematic literature reviews (Ramirez et al. 2018). At the same time, Nguyen et al. (2015) combined active learning, domain experts, and crowd workers to support citation screening in systematic literature reviews. However, many aspects regarding crowd-AI interaction have not been investigated by the HCI community intensively so far. While several papers touch on issues of algorithmic crowd-AI hybrids, supporting research macrotasks was not the focus of existing literature since it has predominantly discussed the technology driving mechanisms in microtasking scenarios with little detail on how technology has been adopted as well as the socio-technical aspects required to facilitate a crowd-AI integration for solving complex problems in science.

5.3 Crowd-AI Systems as a Scaffold for Complex Work

When applied to highly complex problem-solving tasks, the depth and breadth of crowd-powered systems are far beyond the traditional definition of macrotask crowdsourcing. In some circumstances, they can benefit from a crowd-AI hybrid approach. However, replicating one second of human brain activity corresponds to more than 80,000 processors and over a petabyte of system memory (Gil and Hirsh 2012). This involves a vast set of challenges for deploying AI algorithms able to systematically explore multidimensional data and autonomously discover patterns at large scale (Gil et al. 2014). On reading the literature, a significant body of research exists on the adoption of crowd intelligence as a scaffold for machine learning (Kamar 2016). For instance, crowd-machine systems like Flock (Cheng and Bernstein 2015) combine the strengths of human crowd workers and computer algorithms to generate hybrid classifiers. As shown in Table 5.2, there are also some design issues that can be taken into account in the deployment of macrotask crowdsourcing systems.

With the rapid growth of crowdsourcing, many scholars have exhaustively discussed aspects such as crowdsourced task features, quality control, crowd and crowdsourcer attributes, motivational factors, crowdsourcing system features, role of contributors, and aggregation mechanisms, to name a few (e.g., Vukovic 2009; Geiger et al. 2011; Dong et al. 2017). Crowd workers can collaborate explicitly to solve a target problem by sharing structured information or building artifacts (e.g., software). On the other hand, implicit collaboration involves "invisible" contributions such as solving *captchas* and play games with a scientific purpose (Doan et al. 2011). A task can vary in terms of complexity (e.g., routine), variety, modularity, solvability (e.g., simple to humans), structure, and reliability (Hosseini et al. 2014). A task may be also difficult or expensive to automate. Task dependency represents a critical aspect of macrotasks since crowd workers need to coordinate and build upon the contributions of the other members (Schmitz and Lykourentzou 2016). Some macrotasks (e.g., perform a qualitative study in the field of HCI) are not easily decomposable (Krivosheev et al. 2018) and a critical factor in crowdsourcing complex work relies

Table 5.2 Design framework of macrotask crowdsourcing systems

Dimensions of macrotask crowdsourcing systems	*Characteristics of crowd-computing applications*
Type of target problem	Application class
Crowdsourcing type	Predetermined purpose
> Passive, Directive, Collaborative	Ubiquity
Nature of collaboration	Persuasion
> Implicit, Explicit	Cloud computing
Communication	Collective intelligence
Coordination	Interests sharing
Modularity	Latency
> Decomposable, Non-decomposable	Integrating human inputs into AI systems
Complexity	> Reasoning abilities for hybrid intelligence, Training
> Routine, Complex, Creative	User interface
Degree of manual effort	Architecture
Recruit and retain users	Connectivity
Combine/aggregate inputs	*Crowdware time-space matrix*
Quality control	Same place, Different places
Crowd size and roles	> Virtual, Physical
Expertise	Same time, Different times
Feedback	> Critical, Non-critical
Implications for designing crowd-powered systems	*Crowdsourcing platform features (facilities)*
Timing	Computing platform
Scalability	> Internal, External
Locality	Type of platform
Reliability	Distinguishing features/facilities
Synergy	> Crowd-related interactions
Dependency	> Crowdsourcer-related interactions
Automation	> Task-related facilities
> Simple, Difficult	> Platform-related facilities
Transparency	System and technology issues
Crowd work regulation and ethics	Common design patterns
Anonymity and privacy	Openness
Design patterns embodied in programming metaphors	> Closed, Internally open, Externally open
Idea ecology	Ownership
Web of dependencies	> Public, Private
Intellectual supply chain	Highlighting data use
Collaborative deliberation	Technology access and proficiency of potential participants
Radically fluid virtual organization	Accessible crowd work and assistive technology
Multi-user games	

on the ability of coordinating crowds by means of reliable tasks, protocols, and feedback (Vaish et al. 2017). As argued by Weiss (2016), crowdsourcing approaches also differ in terms of the type of tasks assigned to the crowd, the amount of time spent, and the level of collaboration between members.

The behavior of a crowd in a crowdsourcing system can also vary taking into account its architecture (Doan et al. 2011). For example, a standalone system deals with challenges like recruiting participants and choosing their potential actions. As a large group of individuals with a shared purpose and emotions, a crowd can be physically or virtually situated and the nature of the task is an influential factor concerning the way in which crowd members might be engaged (Schneider et al. 2012). Previous research has also suggested that crowd workers are classified in terms of diversity, largeness, unknownness, underfinedness, and suitability (Hosseini et al. 2014). In crowdsourcing research settings, possible roles include principal researcher, research assistant, and participants or members of the crowd (Vaish et al. 2017) who have different abilities (e.g., pattern recognition) and use computing devices to interact, coordinate and execute tasks (Parshotam 2013). According to Bigham et al. (2015), there are three main types of crowdsourcing. In passive crowdsourcing, crowd participants are unknown to each other but there is the possibility of tracing their collective behavior. Directed crowdsourcing relies on the recruitment and guidance of crowd members through a single individual or algorithm. In collaborative crowdsourcing, the coordination tasks are usually performed by a group of individuals with a shared purpose and a self-determined structure (e.g., Wikipedia[4]).

Concerning the characteristics of crowd-computing applications, scalability is a key feature for crowd-AI hybrids in the sense that we need to adapt to different situations and levels of complexity (Talia 2019). Scaling up the crowd reduces the downtime and thus decreases the latency in crowdsourcing (Difallah et al. 2014). The machine must also provide feedback to the user by interactively informing the decision-making process. In a hybrid crowd-AI system such as CrowdFlow (Quinn et al. 2010), complex crowd work outputs are used to provide feedback for machine algorithms and thus enhance their algorithmic power. Dow and colleagues (2012) identified key dimensions of crowd feedback, including timeliness (asynchronous, synchronous), specificity, source (e.g., peer workers), and format. Prior research also suggests that social transparency among crowd workers can be particularly beneficial in crowdsourcing settings (Huang and Fu 2013). Nonetheless, such mechanisms must be implemented with caution to prevent malicious behaviors in crowd-AI interaction (Kittur et al. 2013).

A large body of work (e.g., Hetmank 2013; Daniel et al. 2018) has exploited the use of new techniques for aggregating crowd inputs while controlling the quality of the contributions and the reliability of contributors as critical factors to the success of crowdsourcing since the responses provided by crowd members can be error-prone and biased due to malicious (or less motivated) workers (Lasecki et al. 2014). This calls into question a number of assumptions that lie behind the notion of "quality control". Daniel et al.'s (2018) investigation on quality attributes and assessment techniques found that quality assessment methods range from self-assessment to peer review, voting, gold standards, and feedback aggregation. Crowd participants are usually engaged in complex work through intrinsic motivational factors (e.g., passion, enjoyment and fun, sense of community, personal achievement) and extrinsic motivations such as financial rewards and promotion (Geiger et al. 2011). In addition, crowd work regulation and ethics raise a lot of concerns about privacy and anonymity, worker rights and fair wages, discrimination, and intellectual property (Hansson and Ludwig 2018). In this kind of scenario, sensitive information about crowd workers such as home location and hobbies can be retrieved and used improperly. Furthermore, we should state at the outset that accessible crowd work (Zyskowski et al. 2015) must be leveraged by assistive technology to support people with disabilities and special needs.

An earlier review of the literature on the design components of crowdsourcing platforms (Hetmank 2013) revealed a focus on the functions and operations of a crowd-powered system as an intermediary that distributes Human Intelligence Tasks (HITs) from requesters to the crowd workers. A crowdsourcing system also comprises technical attributes such as software components, functions, and data objects. As these technologies develop, attention to the design processes that support their outputs is essential. Developers of crowd-powered systems must pay attention to

[4]https://www.wikipedia.org/

aspects like awareness, user interface, authentication, quality control, and workflow support. Typically, workflow systems are deployed "ad hoc" and tailored to particular use cases (Lofi and El Maarry 2014). A crowdsourcing platform must support actions such as recruit and evaluate crowd workers, define and assign HITs, submit contributions, set time period, state rewards, and pay crowd workers. As argued by Vukovic (2009), the loss of network connectivity can compromise the interaction in real-time crowdsourcing settings where a failure may be critical to human lives, as in the case of crisis and emergency response.

By virtue of the recent research efforts on crowd-AI hybrids, there are several missing pieces and areas for future work. The literature on this topic is limited and great care must be taken to aspects like task design (Vaish et al. 2015), risk of overspecialization and failing heuristics (Lofi and El Maarry 2014), ambiguity and systematic error biases (Vaughan 2018), and overload of crowd-generated inputs (Barbier et al. 2012). Some requirements for crowd-AI systems include the translation of system states and operations between humans and machines by means of contextual information (Dong et al. 2017) and the adequate support for open-ended, complex scientific activities at different scales (Correia et al. 2018b). These concerns are often overlooked and result from the increasing complexity of algorithms. Within HCI, the adoption of interactive, human-guided machine learning (Gil et al. 2019) constitutes further avenues of research into the intersection of crowdsourcing and AI for supporting macrotasks.

5.4 Final Remarks

In this chapter, we addressed the need for handling complexity in crowd work through the integration of crowd-AI hybrids. This approach appears to be a viable solution for many areas. Nonetheless, we are aware of very little work that tries to characterize such kind of combination in the context of macrotask crowdsourcing as it moves on from its young age. In framing it as a problem, we want to explore the ways in which the design of intelligent systems can be informed by symbiotic interactions between crowds and machines able to completing complex tasks. The full extent of this crowd-guided AI model will be studied in future stages of this research towards a conceptual framework predicated on the socio-technical aspects that need to be considered when solving complex tasks that require high levels of interdependency and domain expertise.

References

Barbier, G., Zafarani, R., Gao, H., Fung, G., & Liu, H. (2012). Maximizing benefits from crowdsourced data. *Computational and Mathematical Organization Theory, 18*(3), 257–279.

Barowy, D. W., Curtsinger, C., Berger, E. D., & McGregor, A. (2012). Automan: A platform for integrating human-based and digital computation. *ACM SIGPLAN Notices, 47*(10), 639–654.

Bigham, J. P., Bernstein, M. S., & Adar, E. (2015). Human-computer interaction and collective intelligence. *Handbook of Collective Intelligence*, 57.

Borromeo, R. M., & Toyama, M. (2016). An investigation of unpaid crowdsourcing. *Human-Centric Computing and Information Sciences, 6*(1), 11.

Brown, A. W., & Allison, D. B. (2014). Using crowdsourcing to evaluate published scientific literature: Methods and example. *PLoS ONE, 9*(7), e100647.

Chan, J., Chang, J. C., Hope, T., Shahaf, D., & Kittur, A. (2018). Solvent: A mixed initiative system for finding analogies between research papers. In *Proceedings of the ACM Conference on Computer-Supported Cooperative Work and Social Computing*.

Chau, D. H., Kittur, A., Hong, J. I., & Faloutsos, C. (2011). Apolo: Making sense of large network data by combining rich user interaction and machine learning. In *Proceedings of the ACM CHI Conference on Human Factors in Computing Systems* (pp. 167–176).

Cheng, J., & Bernstein, M. S. (2015). Flock: Hybrid crowd-machine learning classifiers. In *Proceedings of the ACM Conference on Computer Supported Cooperative Work and Social Computing* (pp. 600–611).

Cheng, J., Teevan, J., Iqbal, S. T., & Bernstein, M. S. (2015). Break it down: A comparison of macro-and microtasks. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (pp. 4061–4064).

Chettih, A., Gross-Amblard, D., Guyon, D., Legeay, E., & Miklós, Z. (2014). Crowd, a platform for the crowdsourcing of complex tasks. In *BDA 2014: Gestion de Données—Principes, Technologies et Applications* (pp. 51–55).

Correia, A., Schneider, D., Paredes, H., & Fonseca, B. (2018a). SciCrowd: Towards a hybrid, crowd-computing system for supporting research groups in academic settings. In *Proceedings of the 24th International Conference on Collaboration and Technology* (pp. 34–41).

Correia, A., Schneider, D., Fonseca, B., & Paredes, H. (2018b). Crowdsourcing and massively collaborative science: A systematic literature review and mapping study. In *Proceedings of the 24th International Conference on Collaboration and Technology* (pp. 133–154).

Crowston, K., Mitchell, E., & Østerlund, C. (2018). Coordinating advanced crowd work: Extending citizen science. In *Proceedings of the 51st Hawaii International Conference on System Sciences* (pp. 1681–1690).

Daniel, F., Kucherbaev, P., Cappiello, C., Benatallah, B., & Allahbakhsh, M. (2018). Quality control in crowdsourcing: A survey of quality attributes, assessment techniques, and assurance actions. *ACM Computing Surveys (CSUR), 51*(1), 7.

Difallah, D. E., Catasta, M., Demartini, G., & Cudré-Mauroux, P. (2014). Scaling-up the crowd: Micro-task pricing schemes for worker retention and latency improvement. In *Second AAAI Conference on Human Computation and Crowdsourcing*.

Doan, A., Ramakrishnan, R., & Halevy, A. Y. (2011). Crowdsourcing systems on the world-wide web. *Communications of the ACM, 54*(4), 86–96.

Dong, Z., Lu, J., Ling, T. W., Fan, J., & Chen, Y. (2017). Using hybrid algorithmic-crowdsourcing methods for academic knowledge acquisition. *Cluster Computing, 20*(4), 3629–3641.

Dow, S., Kulkarni, A., Klemmer, S., & Hartmann, B. (2012). Shepherding the crowd yields better work. In *Proceedings of the ACM Conference on Computer Supported Cooperative Work* (pp. 1013–1022).

Franzoni, C., & Sauermann, H. (2014). Crowd science: The organization of scientific research in open collaborative projects. *Research Policy, 43*(1), 1–20.

Gaikwad, S. N. S., Morina, D., Ginzberg, A., Mullings, C., Goyal, S., Gamage, D., et al. (2016). Boomerang: Rebounding the consequences of reputation feedback on crowdsourcing platforms. In *Proceedings of the 29th ACM Symposium on User Interface Software and Technology* (pp. 625–637).

Garcia-Molina, H., Joglekar, M., Marcus, A., Parameswaran, A., & Verroios, V. (2016). Challenges in data crowdsourcing. *IEEE Transactions on Knowledge and Data Engineering, 28*(4), 901–911.

Geiger, D., Seedorf, S., Schulze, T., Nickerson, R. C., & Schader, M. (2011). Managing the crowd: Towards a taxonomy of crowdsourcing processes. In *Proceedings of the Proceedings of the 17th Americas Conference on Information Systems*.

Gil, Y., & Hirsh, H. (2012). Discovery informatics: AI opportunities in scientific discovery. In *Proceedings of the AAAI Fall Symposium: Discovery Informatics*.

Gil, Y., Greaves, M., Hendler, J., & Hirsh, H. (2014). Amplify scientific discovery with artificial intelligence. *Science, 346*(6206), 171–172.

Gil, Y., Honaker, J., Gupta, S., Ma, Y., D'Orazio, V., Garijo, D., et al. (2019). Towards human-guided machine learning. In *Proceedings of the 24th ACM International Conference on Intelligent User Interfaces*.

Good, B. M., Nanis, M., Wu, C., & Su, A. I. (2014). Microtask crowdsourcing for disease mention annotation in PubMed abstracts. In *Proceedings of the Pacific Symposium on Biocomputing* (pp. 282–293).

Haas, D., Ansel, J., Gu, L., & Marcus, A. (2015). Argonaut: Macrotask crowdsourcing for complex data processing. *Proceedings of the VLDB Endowment, 8*(12), 1642–1653.

Hansson, K., & Ludwig, T. (2018). Crowd dynamics: Conflicts, contradictions, and community in crowdsourcing. *Computer Supported Cooperative Work (CSCW)*, 1–4.

Hetmank, L. (2013). Components and functions of crowdsourcing systems – A systematic literature review. *Wirtschaftsinformatik*, 4.

Hochachka, W. M., Fink, D., Hutchinson, R. A., Sheldon, D., Wong, W. K., & Kelling, S. (2012). Data-intensive science applied to broad-scale citizen science. *Trends in Ecology & Evolution, 27*(2), 130–137.

Hosseini, M., Phalp, K., Taylor, J., & Ali, R. (2014). The four pillars of crowdsourcing: A reference model. In *Proceedings of the 2014 IEEE Eighth International Conference on Research Challenges in Information Science (RCIS)* (pp. 1–12).

Huang, S. W., & Fu, W. T. (2013). Don't hide in the crowd!: Increasing social transparency between peer workers improves crowdsourcing outcomes. In *Proceedings of the ACM CHI Conference on Human Factors in Computing Systems* (pp. 621–630).

Ikeda, K., Morishima, A., Rahman, H., Roy, S. B., Thirumuruganathan, S., Amer-Yahia, S., et al. (2016). Collaborative crowdsourcing with Crowd4U. *Proceedings of the VLDB Endowment, 9*(13), 1497–1500.

Kamar, E. (2016). Directions in hybrid intelligence: Complementing AI systems with human intelligence. In *IJCAI* (pp. 4070–4073).

Kittur, A., Smus, B., Khamkar, S., & Kraut, R. E. (2011). Crowdforge: Crowdsourcing complex work. In *Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology* (pp. 43–52).

Kittur, A., Khamkar, S., André, P., & Kraut, R. (2012). CrowdWeaver: Visually managing complex crowd work. In *Proceedings of the ACM Conference on Computer Supported Cooperative Work* (pp. 1033–1036).

Kittur, A., Nickerson, J. V., Bernstein, M., Gerber, E., Shaw, A., Zimmerman, J., et al. (2013). The future of crowd work. In *Proceedings of the ACM Conference on Computer-Supported Cooperative Work and Social Computing* (pp. 1301–1318).

Krivosheev, E., Casati, F., Caforio, V., & Benatallah, B. (2017). *Crowdsourcing paper screening in systematic literature reviews*. arXiv:1709.05168.

Krivosheev, E., Casati, F., & Benatallah, B. (2018). Crowd-based multi-predicate screening of papers in literature reviews. In *Proceedings of the World Wide Web Conference* (pp. 55–64).

Kulkarni, A., Gutheim, P., Narula, P., Rolnitzky, D., Parikh, T., & Hartmann, B. (2012). Mobileworks: Designing for quality in a managed crowdsourcing architecture. *IEEE Internet Computing, 16*(5), 28–35.

Kulkarni, A., Narula, P., Rolnitzky, D., & Kontny, N. (2014). Wish: Amplifying creative ability with expert crowds. In: Second AAAI Conference on Human Computation and Crowdsourcing.

Lasecki, W. S. (2014). Crowd-powered intelligent systems. *Human Computation Journal*.

Lasecki, W. S., Teevan, J., & Kamar, E. (2014). Information extraction and manipulation threats in crowd-powered systems. In *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work and Social Computing* (pp. 248–256).

Law, E., Gajos, K. Z., Wiggins, A., Gray, M. L., & Williams, A. C. (2017). Crowdsourcing as a tool for research: Implications of uncertainty. In *Proceedings of the ACM Conference on Computer-Supported Cooperative Work and Social Computing* (pp. 1544–1561).

Li, G., Wang, J., Zheng, Y., & Franklin, M. J. (2016). Crowdsourced data management: A survey. *IEEE Transactions on Knowledge and Data Engineering, 28*(9), 2296–2319.

Lofi, C., & El Maarry, K. (2014). Design patterns for hybrid algorithmic-crowdsourcing workflows. CBI, 1 (pp. 1–8).

Luz, N., Silva, N., & Novais, P. (2015). A survey of task-oriented crowdsourcing. *Artificial Intelligence Review, 44*(2), 187–213.

Marcus, A., & Parameswaran, A. (2015). Crowdsourced data management: Industry and academic perspectives. *Foundations and Trends in Databases, 6*(1–2), 1–161.

Morishima, A., Shinagawa, N., Mitsuishi, T., Aoki, H., & Fukusumi, S. (2012). CyLog/Crowd4U: A declarative platform for complex data-centric crowdsourcing. *Proceedings of the VLDB Endowment, 5*(12), 1918–1921.

Mortensen, M. L., Adam, G. P., Trikalinos, T. A., Kraska, T., & Wallace, B. C. (2017). An exploration of crowdsourcing citation screening for systematic reviews. *Research Synthesis Methods, 8*(3), 366–386.

Nebeling, M., Guo, A., To, A., Dow, S., Teevan, J., & Bigham, J. (2015). WearWrite: Orchestrating the crowd to complete complex tasks from wearables. In *Adjunct Proceedings of the 28th Annual ACM Symposium on User Interface Software and Technology* (pp. 39–40).

Nguyen, A. T., Wallace, B. C., & Lease, M. (2015). Combining crowd and expert labels using decision theoretic active learning. In *Proceedings of the Third AAAI Conference on Human Computation and Crowdsourcing*.

Niu, X. J., Qin, S. F., Vines, J., Wong, R., & Lu, H. (2018). Key crowdsourcing technologies for product design and development. *International Journal of Automation and Computing*, 1–15.

Nov, O., Arazy, O., & Anderson, D. (2014). Scientists@Home: What drives the quantity and quality of online citizen science participation? *PLoS ONE, 9*(4), e90375.

Parshotam, K. (2013). Crowd computing: A literature review and definition. In *Proceedings of the South African Institute for Computer Scientists and Information Technologists Conference* (pp. 121–130).

Peer, E., Brandimarte, L., Samat, S., & Acquisti, A. (2017). Beyond the Turk: Alternative platforms for crowdsourcing behavioral research. *Journal of Experimental Social Psychology, 70,* 153–163.

Quinn, A. J., Bederson, B. B., Yeh, T., & Lin, J. (2010). Crowdflow: Integrating machine learning with Mechanical Turk for speed-cost-quality flexibility. *Better Performance over Iterations*.

Ramirez, J., Krivosheev, E., Baez, M., Casati, F., & Benatallah, B. (2018). CrowdRev: A platform for crowd-based screening of literature reviews. arXiv:1805.12376.

Ranard, B. L., Ha, Y. P., Meisel, Z. F., Asch, D. A., Hill, S. S., Becker, L. B., et al. (2014). Crowdsourcing—Harnessing the masses to advance health and medicine, a systematic review. *Journal of General Internal Medicine, 29*(1), 187–203.

Retelny, D., Robaszkiewicz, S., To, A., Lasecki, W. S., Patel, J., Rahmati, N., & Bernstein, M. S. (2014). Expert crowdsourcing with flash teams. In *Proceedings of the 27th Annual ACM Symposium on User Interface Software and Technology* (pp. 75–85).

Rigby, J. (2009). Comparing the scientific quality achieved by funding instruments for single grant holders and for collaborative networks within a research system: Some observations. *Scientometrics, 78*(1), 145–164.

Salehi, N., Teevan, J., Iqbal, S., & Kamar, E. (2017). Communicating context to the crowd for complex writing tasks. In *Proceedings of the ACM Conference on Computer Supported Cooperative Work and Social Computing* (pp. 1890–1901).

Schmitz, H., & Lykourentzou, I. (2016). *It's about time: Online macrotask sequencing in expert crowdsourcing*. arXiv:1601.04038.

Schmitz, H., & Lykourentzou, I. (2018). Online sequencing of non-decomposable macrotasks in expert crowdsourcing. *ACM Transactions on Social Computing, 1*(1), 1.

Schneider, D., Moraes, K., De Souza, J. M., & Esteves, M. G. P. (2012). CSCWD: Five characters in search of crowds. In *Proceedings of the IEEE International Conference on Computer Supported Cooperative Work in Design* (pp. 634–641).

Sieg, J. H., Wallin, M. W., & von Krogh, G. (2010). Managerial challenges in open innovation: A study of innovation intermediation in the chemical industry. *R&D Management, 40*(3), 281–291.

Stonebraker, M., Bruckner, D., Ilyas, I. F., Beskales, G., Cherniack, M., Zdonik, S. B. et al. (2013). Data curation at scale: The data tamer system. In *CIDR*.

Talia, D. (2019). A view of programming scalable data analysis: From clouds to exascale. *Journal of Cloud Computing, 8*(1), 4.

Tsueng, G., Nanis, M., Fouquier, J., Good, B., & Su, A. (2016). Citizen science for mining the biomedical literature. *BioRxiv*, 038083.

Vaish, R., Davis, J., & Bernstein, M. (2015). Crowdsourcing the research process. *Collective Intelligence*.

Vaish, R., Gaikwad, S. N. S., Kovacs, G., Veit, A., Krishna, R., Arrieta Ibarra, I.,… & Davis, J. (2017). Crowd research: Open and scalable university laboratories. In: Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology (pp. 829–843).

Valentine, M. A., Retelny, D., To, A., Rahmati, N., Doshi, T., & Bernstein, M. S. (2017). Flash organizations: Crowdsourcing complex work by structuring crowds as organizations. In *Proceedings of the ACM CHI Conference on Human Factors in Computing Systems* (pp. 3523–3537).

Vaughan, J. W. (2018). Making better use of the crowd: How crowdsourcing can advance machine learning research. *Journal of Machine Learning Research, 18*(193), 1–46.

Vukovic, M. (2009). Crowdsourcing for enterprises. In *IEEE Congress on Services-I* (pp. 686–692).

Walsh, B., Maiers, C., Nally, G., Boggs, J., & Team, Praxis Program. (2014). Crowdsourcing individual interpretations: Between microtasking and macrotasking. *Literary and Linguistic Computing, 29*(3), 379–386.

Wang, N. C., Hicks, D., & Luther, K. (2018). Exploring trade-offs between learning and productivity in crowdsourced history. In *Proceedings of the ACM on Human-Computer Interaction* (CSCW) (Vol. 2, p. 178).

Weiss, M. (2016). Crowdsourcing literature reviews in new domains. *Technology Innovation Management Review, 6*(2), 5–14.

Whiting, M. E., Gamage, D., Gaikwad, S. N. S., Gilbee, A., Goyal, S., Ballav, A., et al. (2017). Crowd guilds: Worker-led reputation and feedback on crowdsourcing platforms. In *Proceedings of the ACM Conference on Computer Supported Cooperative Work and Social Computing* (pp. 1902–1913).

Xie, H., & Lui, J. C. (2018). Incentive mechanism and rating system design for crowdsourcing systems: Analysis, tradeoffs and inference. *IEEE Transactions on Services Computing, 11*(1), 90–102.

Yan, X., Ding, X., & Gu, N. (2016). Crowd work with or without crowdsourcing platforms. In: *Proceedings of the IEEE 20th International Conference on Computer Supported Cooperative Work in Design (CSCWD)* (pp. 56–61).

Zakaria, N. A., & Abdullah, C. Z. H. (2018). Crowdsourcing and library performance in digital age. Development, 7(3).

Zyskowski, K., Morris, M. R., Bigham, J. P., Gray, M. L., & Kane, S. K. (2015). Accessible crowdwork? Understanding the value in and challenge of microtask employment for people with disabilities. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work and Social Computing* (pp. 1682–1693).