

Theoretical Underpinnings and Practical Challenges of Crowdsourcing as a Mechanism for Academic Study

António Correia
INESC TEC and UTAD and
University of Kent
antonio.g.correia@inesctec.pt

Shoaib Jameel
University of Kent
M.S.Jameel@kent.ac.uk

Daniel Schneider
Federal University of Rio de Janeiro
schneider@nce.ufrj.br

Benjamim Fonseca
INESC TEC and UTAD
benjaf@utad.pt

Hugo Paredes
INESC TEC and UTAD
hparedes@utad.pt

Abstract

Researchers in a variety of fields are increasingly adopting crowdsourcing as a reliable instrument for performing tasks that are either complex for humans and computer algorithms. As a result, new forms of collective intelligence have emerged from the study of massive crowd-machine interactions in scientific work settings as a field for which there is no known theory or model able to explain how it really works. Such type of crowd work uses an open participation model that keeps the scientific activity (including datasets, methods, guidelines, and analysis results) widely available and mostly independent from institutions, which distinguishes crowd science from other crowd-assisted types of participation. In this paper, we build on the practical challenges of crowd-AI supported research and propose a conceptual framework for addressing the socio-technical aspects of crowd science from a CSCW viewpoint. Our study reinforces a manifested lack of systematic and empirical research of the symbiotic relation of AI with human computation and crowd computing in scientific endeavors.

1. Introduction

The increasing application of crowdsourcing and Artificial Intelligence (AI) in science has attracted a great deal of research interest in recent decades [1]. IT-enabled work practices in scientific inquiry evolved from communication infrastructures deployed to share preliminary findings and resources (collaboratories) in the late 1980s, to citizen science platforms in the 1990s that engaged members of the public (citizen scientists) in data gathering, analysis, and interpretation [2]. More recently, researchers have adopted crowd-powered systems and human-in-the-loop machine learning for aiding scientific activities [3]. However, crowd-AI

supported work at scale is a problem that requires further examination since understanding, filtering, and processing the latest scientific advances is a difficult task for both humans and machines.

For humans, the massive volume of new research findings makes it impossible to grasp due to the human cognitive and memory limits. On the other hand, algorithms still cannot process efficiently the scientific unstructured data [4]. To bridge this gap, researchers have slightly explored the applicability of hybrid crowd-machine methods and techniques, mostly in small-scale institutional settings. Despite this interest, it fails to take into account the complexity of leveraging collective hybrid intelligence to produce large, high-accuracy training datasets on specific research topics, and there is still a need to understand how crowd-powered systems can be used “as a scaffold for AI” [5].

Drawing on review of extant literature on crowd science and related disciplines, the aim of this study is to develop a descriptive and analytical framework that enables a better understanding of the processes and mechanisms involved in scientific work settings. In particular, our aim is to describe how research might benefit from crowd-AI hybrids through a methodical approach to the conceptualization, design and deployment of new Computer-Supported Cooperative Work (CSCW) systems. We take a socio-technical view of human-centered systems design [6], which acknowledges that both human and technical aspects must be taken into account in the functioning of a system. Thus, this paper can be used to position the research already conducted on this subject in order to understand its relevance and coverage while articulating the research problem. Looking towards the future, we translate our findings into a set of design implications through the proposal of a conceptual framework. The results shed new light on the design of socio-technical systems for crowd-enabled science.

2. Related Work

The scientific activity and research in the field of CSCW has consistently sought for theoretical constructs and practice-oriented assumptions on the design and use of socio-technical systems, including key aspects of cooperative work arrangements. Such conceptualizations comprise a vast set of well-established theories, descriptive methods, conceptual frameworks, and hybrid forms [7]. However, “the field still suffers from the lack of an understanding of how people involve themselves, in larger, more complex socio-technical systems” [8]. That is, we need to understand how people plan, coordinate and manage their collective actions as stable social structures in the context of large-scale distributed work, more precisely ‘crowd work’ [9]. From a social analytic and system development position, a co-evolving relationship between hybrid human-machine interactions [48] constitutes a new class of socio-technical systems.

2.1. Managing Complex Problems in Crowd-Enabled Science: Scaling Up Scientific Work

It has now been suggested that crowdsourcing is a reliable method to evaluate research in a low-cost, rapid, and trustworthy manner despite the cost prohibitive nature of paid systems [3]. When applied to tackle scientific problems, crowdsourcing may be characterized by a ‘bottom up’ structure and openness to a large pool of scientific collaborators (researchers and other individuals such as students and citizen scientists) and their respective interactions (intentional or not) within or outside their institutions [11,12].

The use of crowd science in research design, data collection and analysis represents a reliable solution for research tasks that cannot be fully automated by means of specialized computer algorithms. For instance, crowdsourcing can be very useful to enhance “the quality, cost, and speed of a research project while engaging large segments of the public and creating novel science” [13]. In other words, crowdsourcing research activities can help students, researchers, and members of R&D projects in industry by providing some benefits such as scale, consensus and validation criteria, diversity of perspectives, and reduced time needed to complete a task [14].

Although this approach has received increasing attention by enabling thousands of people spatially distributed “to coordinate joint progress on an open-ended research effort” [4], some researchers are still reluctant to adopt crowdsourcing technology due to the highly uncertain and iterative nature of scientific discovery [3,10]. This breadth raises a number of new

challenges, including instruction ambiguity and worker honesty [15], motivation for participation in non-profit communities [16], and limited expertise and attention to cope with high-dimensional and ill-structured data [17]. Apparently it is difficult to assess the adoption of crowd-powered networked systems in scientific work contexts to make discoveries much faster and accurate while supporting scientists in keeping up to date.

There is a vast amount of studies involving crowdsourcing research tasks. For example, Good and co-authors [18] recruited non-scientists to recognize disease concepts in biomedical paper abstracts. Moreover, O’Leary and colleagues [19] proved that a crowd of over 600 non-experts could be as valuable as scientific experts when scoring phylogenetic matrices of phenotypes. This finding was also corroborated in other evaluations of scientific literature using crowdsourcing (e.g., [20]). Besides using existing crowdsourcing solutions, other researchers have also studied the feasibility of crowdsourcing for facilitating the systematic literature review process [1,14].

Literature suggests that crowd science can be established where tackling research tasks by the crowd is understood as a method of scientific discovery that researchers can use to solve complex problems [11]. MicroPasts [21] was established in 2013 by a group of researchers from the UCL Institute of Archaeology and the British Museum at UK as a heritage crowdsourcing platform that uses a participatory model in which volunteers can contribute by providing transcriptions of artifacts or sketching outlines of objects on figure collections. In the same vein, Science Gossip [22] is a crowd science project where volunteers can identify, categorize and correlate millions of pages of digitized texts from the Biodiversity Heritage Library. As it can be observed from Figure 1, researchers and members of the public can collaborate to further scientific discovery by generating high-quality research data that cannot be replicated by machines.



Figure 1. An example of a crowdsourced task in Science Gossip [22].

Crowd science projects such as Galaxy Zoo¹ have been considered to be the socio-technical ecosystems for important discoveries by means of large-scale resource acquisition through distributed cognitive effort. Researchers have also executed experiments using games with a scientific purpose [23]. At the same time, scheduling tools such as Frenzy [24] enable a crowd to manipulate, scan, and review a conference schedule with support for conflict resolution and multifaceted views. Previous experiments on massive authorship of academic papers reported some challenges related with coordination mechanisms, tool design, content handling, and task differences [25]. More recently, researchers engaged crowd members in advanced science tasks [26].

2.2. Hybrid Machine-Crowd Interaction: Understanding Symbiotic Human Interactions with AI

While human crowds can produce distinct ideas, annotate data, and classify information with high accuracy, algorithms are particularly useful in handling large volumes of data (with multiple criteria), providing new insights from complex decision rules of human intuition for further validation in a human-in-the-loop machine learning approach that was not yet captured by interactive, crowd-AI architectures and systems [27]. Hybrid human-machine computation has been addressed as an extension of the burgeoning field of human computation [28], pursuing the development of systems that “tightly integrate human computation and machine resources” [29]. This kind of mixed-initiative approach can be particularly fruitful in crowd science contexts to provide evidence on demand and to refine machine-extracted metadata [54] using semantic analysis and intelligent techniques for automatic classification enabled by crowd workers who can filter, process, and verify the information.

Most studies agree on the use of crowdsourcing as a reliable method for supervised and semi-supervised machine learning (e.g., active learning), from feature generation to prediction, deeper analysis, and classification of mass volumes of information [20]. For instance, the literature has documented systems that discover the institutional affiliations of authors in scholarly papers [31]. Concomitantly, intelligent systems integrating both machine resources and crowd-nominated (e.g., Flock [32], a system that asks the crowd to compare paired examples, containing a model for generating hybrid classifiers) can outperform those that use either in isolation. In a recent work, PANDA

[33] was presented as a hybrid intelligence system for academic knowledge discovery and acquisition, while SOLVENT [34] is a mixed-initiative system where humans can annotate aspects of research papers (e.g., findings) in order to find analogies through a computational model.

Research into solving research problems by means of crowd-AI integration is already underway. For instance, CrowdRev [35] is a crowd-powered system that uses AI for paper screening in systematic literature reviews. Furthermore, crowdsourcing and natural language processing have been also studied as valid approaches for extracting entity and relational phrases from sentences and ranking entities to obtain plausible hypotheses [33]. On the other hand, authors such as Quinn and others [28] assume that research to date has been limited in scope, and there is a large path of further improvement towards hybrid classifiers embedding crowds inside of machine learning systems and architectures.

3. Crowd-AI Supported Science: A Conceptual Framework

As previously noted by Hughes and co-authors [36], an analytic framework needs to be crafted ‘from the ground up’, comprising the details and diversity of work domains. A conceptual framework (Figure 2) was created as a result of a series of data-driven operations with the aim of providing a snapshot to guide future development of crowd-powered systems for supporting scientific work practices. We employed principles from human-centered computing research as a starting point to understand the complementary way in which human crowds interact with AI systems in scientific contexts.

3.1. Method

By documenting the breadth and variety of conceptual units from studies that seek to address the integration of crowd inputs into AI systems we try to make a new set of heuristics based on literature research found. The conceptual model follows the Kittur et al.’s [9] approach on complex crowd work and also earlier literature studies on the socio-technical dimensions of crowd-AI integration.

The taxonomy development method is derived from Nickerson and co-authors [37]. As the main goal of this study is to understand the integration of crowd- and AI-powered systems in scientific work settings through different levels of analysis, a scoping review was also performed following the guidelines of evidence-based software engineering [39].

¹ <http://zoo1.galaxyzoo.org/>

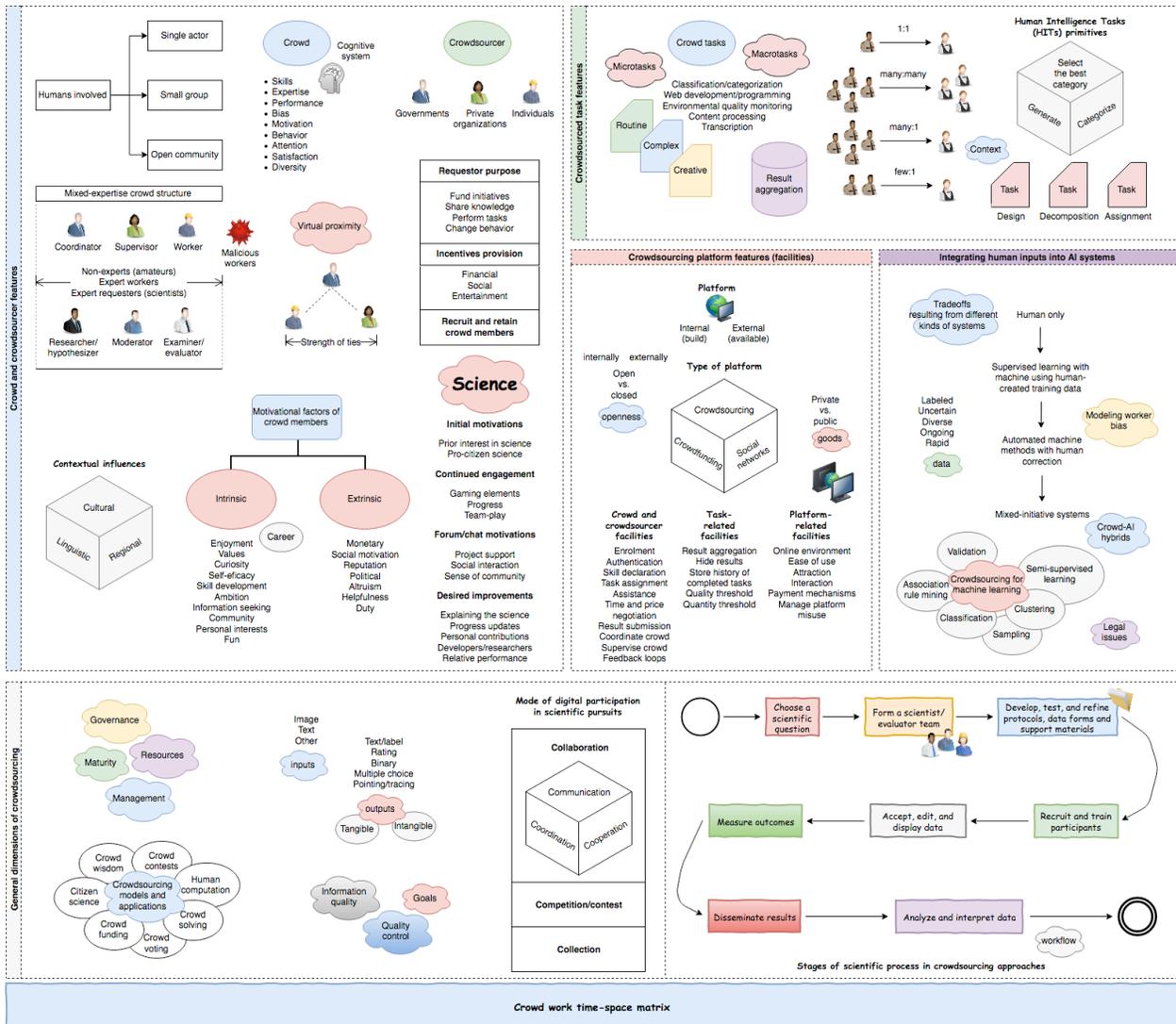


Figure 2. Conceptual model of crowd-AI science.

3.1.1. Data collection. The first stage of our study was aligned with a descriptive literature analysis [38] to identify key characteristics behind the socio-technical infrastructure of crowd-AI supported research. This approach seeks to systematically portray and examine insights based on categorizing research into a literature classification scheme that is used as a foundation for the taxonomy development process. Researchers have been applying this method as a procedure that involves systematic processes such as comprehensive literature search, filtering, and classification.

We limited our search to peer-reviewed papers in English (published in 2006 or later). In addition, we only considered studies that survey or describe features and models related with the ability of leveraging crowdsourcing and massively collaborative science. This excludes introductions to special issues and

minitracks, tutorials, dissertations and theses, etc. The authors performed searches on central scientific literature databases such as ACM Digital Library, IEEE Xplore, and ScienceDirect by applying two Boolean expressions using the following search string:

(crowd science OR citizen science OR crowdsourcing OR human computation OR collective intelligence OR mass collaboration) AND (artificial intelligence OR machine learning OR mixed-initiative systems) AND (survey OR taxonomy OR systematic literature review OR science OR research OR scientific discovery)

The review followed a complementary process in which the authors analyzed the reference list in the resulting papers by means of a snowballing strategy [39]. The data retrieved were stored using an extraction form created to gather all relevant information from the primary studies. After a de-duplication process, our assessment of research papers' eligibility yielded a final sample of unique articles from which we retrieved the main taxonomic units discussed in this paper.

3.1.2. Iterative taxonomy development. The second stage of our research relies on the application of Nickerson et al.'s [37] method to create a classification scheme. As argued by Schmidt-Kraepelin et al. [40], taxonomies are “important tools in most scientific disciplines” since they help to understanding and structuring knowledge in a particular domain. For this study, we defined a meta-characteristic concerned with the *attributes/features of crowdsourcing for scientific purposes*. As noted above, we followed the guidelines provided in previous studies (i.e., [37,40]) concerning the predefined ending conditions to create a valid taxonomy. Finally, an initial version of our conceptual model was formed and iteratively created by grouping characteristics and dimensions retrieved from the final list of papers reviewed during the first stage. The taxonomy development stages are shown in Figure 3.

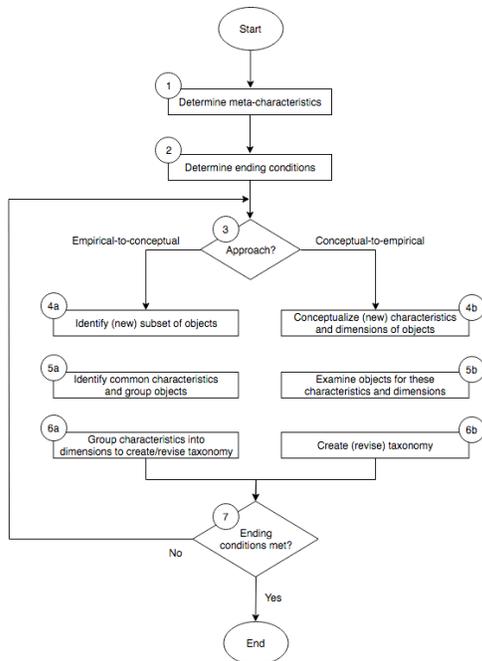


Figure 3. Taxonomy development process (adapted from Nickerson et al. [37]).

3.2. Dimensions

In the literature, there are many descriptive studies primarily concerned with providing conceptualizations of social and technical aspects of crowdsourcing (e.g., [29]). Nevertheless, it has been argued that there is a lack of practical frameworks to distinguish crowd science “as a designable system whose boundaries encompass social, scientific and technical elements that are interconnected” [41]. A description of the main dimensions of our classification scheme for crowd-AI supported research is given in the ensuing sections.

3.2.1. Crowd and crowdsourcer features. We are influenced by the biological manifestations of the nature. Human crowds are not an exception. The nature’s ecosystem is a highly complex adaptive, self-regulating system where living organisms interact within the environment to sustain and even adapt to its changes [48]. That is, the biological properties of individuals influence the way as everything change and organisms’ observation is an important part of socio-technical research. In a crowd scenario, each organism follows a set of simple rules but the collective acts as a whole in an efficient way.

Over time, crowd members evolve their “hidden” networks and relationships using alternative forms of communication (e.g., subreddits) [45]. This evidence further highlight several potential opportunities for CSCW research to better understand how and why crowds are communicating in these subcommunities as well as the potential impact of these interactions (e.g., crowd bias). In crowd science, open projects may need a higher degree of collaboration [11], and the nature and roles of crowd members in research settings can vary from a core scientist/principal researcher (who advises the research project) to an academic expert from an external institution, local case actors (e.g., residents affected by the issue under consideration), individual scientists performing preliminary research work, and volunteers or members of the crowd who have different capabilities and use computing devices to interact, coordinate, and execute tasks [4].

An examination of prior work suggests that crowd participants involved in scientific research are usually (intrinsically or extrinsically) motivated by factors like personal interest/curiosity, recognition, involvement in research (including learning about science), working with an institution, skill building/career development, helping a project, and enjoyment [44]. Researchers following this path have also examined the effect of the context in which each member is situated on his/her motivation to participate in crowdsourcing. In general, contextual factors can range from cultural aspects to crowd size and diversity, virtual proximity, location, contribution and response time, task features, and domain expertise [10]. In addition to these aspects, Shirk et al. [65] found that the level of engagement of participants in crowd science can be measured by the time spent on each task, periodicity, relative activity duration, and research effort required to complete a scientific task.

3.2.2. Crowdsourced task features. In participatory research efforts, individual participants interact with researchers and therefore help crowdsourcing projects by providing their experience and skill sets in solving tasks that can vary in terms of complexity, modularity,

type and properties, and granularity. Such tasks must be carefully designed taking into account the socio-technical environment and its contextual or situational factors (e.g., location) as well as the agent(s) involved. Buecheler et al. [30] go even further by claiming that a crowd can be particularly suitable for conducting parts of the scientific process. This includes data collection (e.g., identifying and retrieving literature), analysis of the collected data (e.g., finding patterns and analogies), and interpretation of results. As a result, the traditional scientific processes being crowdsourced can be enhanced with novel insights based on the amount and diversity of ideas and views.

There is a rich body of research around microtask crowdsourcing for scientific purposes (e.g., [18]). As pointed out in our previous study [43], such types of tasks do not involve specialized skills and can be ephemeral and relatively simple for crowd workers. Although crowdsourcing literature has focused largely on microtasking and independent work, crowd science tasks are complex by nature and usually require higher degrees of cooperation [9]. That is, crowd members can be leveraged as an inexpensive and highly scalable means to perform tasks involving dependencies and advanced levels of coordination [26]. Examples include but are not limited to writing a scientific paper, designing a module to be incorporated into a scientific software system, or analyzing an academic field of research through the lenses of scientometrics.

Unlike microtask crowdsourcing, macrotasks can involve creativity and highly specialized expertise [43]. Consistently, a lot of work has also drawn attention to the design of workflow systems intended to support the task decomposition process [9]. This implies that crowd work applications and interfaces need to be designed taking into consideration the requisites for managing and aggregating interdependent and usually more extensive pieces of work that characterize macrotask crowdsourcing.

3.2.3. Crowdsourcing platform features (facilities).

Despite some investments in developing IT-integrated system infrastructures as massive collaborative spaces, designing crowd science platforms often face a large gap between the social requirements made by crowd workers and requesters and what system designers are able to support at a technical level [59]. This builds on discussions in the field of CSCW claiming for critical debate on the role of fieldwork in the construction of computational artifacts. Moreover, this is also in line with the Schmidt and Bannon's [47] claim for a wider conceptualization of the design space of CSCW in order to encompass highly volatile and complex social arrangements and practices that are open and diverse by nature and include a large (and possibly unknown)

number of participants with conflicting motives and incompatible strategies [8].

From the point of view of the dual nature of socio-technical systems, we need to provide functionalities for aiding scientific investigation with different levels of complexity. To do this, we require different forms of participant engagement (e.g., gamification [40]) in order to maintain the research interest while attaining new users and retaining the old ones. Among the design claims discussed by Tinati and colleagues [50], we draw attention to the task specificity that can vary from simple data labeling to rigorous laboratory studies. As argued by the authors, one of the challenges with such systems is that it is often unclear how to ensure timely support from expert scientists. That is, the quality and periodicity of feedback along with other related task-related aspects such as in-task user interface guidance and discussion of observations can be of critical importance in reducing the wasted efforts. At a technical level, some prior works (e.g., [49]) have stressed the need for incorporating a set of mechanisms for authentication (assuring security and access rules), task design and decomposition, workflow design, user management, aggregation, and payment.

3.2.4. Integrating human inputs into AI systems.

A large and growing body of literature (e.g., [27,48]) has attempted to investigate the socio-technical outcomes of collective hybrid intelligence. A key challenge for a system integrating AI and crowdsourcing for scientific purposes relies on modeling intelligent behavior [30] while providing support for vast amounts of users with diverse background and dispersed geographically contributing over multidimensional data fields. That is, thousands of unconnected outputs can be aggregated in order to train algorithms for providing better decisions [54]. In a similar way, the crowd becomes more skilled and accurate as a result of this co-evolving relationship [9,48]. Consistent with this suggestion, Dong et al. [33] came up with more or less the same aspects by arguing that the integration of crowd inputs into AI systems requires a comprehensible translation of system states and requirements to humans to elicit high quality outputs from the crowd.

There are specific challenges inherent in deploying human-machine systems based on adaptivity [62]. In view of this fact, machine agents must be adaptive to dynamic domains and self-organized crowds. Recent studies have shown that designing workflows that leverage hybrid intelligence for complex tasks are difficult to achieve due to the fact that the current mechanisms are usually deployed in an ad-hoc manner and tailored for specific situations or use cases [51]. In light of these findings, remarkable progress has been made towards more flexible workflows. One recent

study shed light on the possible approaches for fixing the rigid (and often preprogrammed) nature of current workflow design [62]. As pointed out by the authors, a strategy to solve such problem relies on putting humans in the loop of decision process through a dynamic model in which human crowds are allowed to self-organize and determine the way of execution of workflow tasks. In addition, crowd-AI hybrid systems intended to support research pursuits must be fully able to enhance synergized interaction through explainable AI and crowd-nominated features [32]. Nevertheless, the literature also highlights important ethical risks and considerations linked to the potential influences of these technologies [58].

3.2.5. General dimensions of crowdsourcing. Crowd science has been deployed using direct engagement and (explicit) peer feedback to train crowd participants [4]. The level of activeness in the crowdsourcing system and the relationship between researchers (as requesters) and crowd workers (or volunteers) are crucial for a research project's success [52]. Moreover, quality control is considered as a critical factor to the success of the crowdsourced research work [12]. The mechanisms for improving data quality in crowd science range from decomposing and simplifying tasks to adapting the research questions, cross-checking for consistency, and training/tight supervision [55]. Some strategies used by researchers to reduce the dangers of malicious work in crowdsourcing include governance, active control, incentive mechanisms, and monitoring of crowd activities through review systems [57].

In CSCW, there is agreement that coordination is a central function in highly complex, interdependent activities. Coordination theory [53] considers the set of dependencies occurring between tasks and resources. For instance, performance losses can occur if there is a mismatch between dependencies and available coordination mechanisms. Extrapolating to the specific context of crowd science projects, volunteers may face a 'shared-output dependency' by agreeing on the outputs (e.g., image labels) that resulted from consensus building [26]. Further expanding the scope, there are several potential opportunities for future research to better understand scientific activities where there is an absence of a central control mechanism and management function [46]. Although there has been long-standing interest in studying stigmergic forms of crowd work as situated cognition arrangements in which participants are able to "build on one another's contributions without explicit coordination" [42], many aspects regarding 'stigmergic coordination' in crowd science platforms have not been investigated by the CSCW community intensively so far.

4. Discussion

4.1. General Implications for CSCW Research

Combining human and machine intelligence poses several challenges and creates new paths for further examination within the CSCW community-wide intellectual agenda. As we have outlined before, many of the methodological advances in the study of crowd behavior and complex socio-technical phenomena have remained outside the umbrella of CSCW. Social relations occurring in crowd-machine hybrid scenarios constitute a raw material for the construction of models that can explain the crowd behavior and its co-evolving social structures as highly interdependent organisms with recurring transformations in their systems and subsystems.

Consequently, novel forms of participatory design and ethnography inquiry are required to obtain insights from crowd workers' experiences and their evolving relationships [2]. When considering the in-depth study of crowd science and its intersection with AI systems, several problems arise from simple crowd management to errors due to the lack of scientific domain expertise, insufficient training data, data quality and verification, task coordination/modulation, and uncertain reliability of the machine's decisions [10]. The characteristics of this integration have not been dealt with in depth from a CSCW perspective, so a research agenda can be extended to this domain to inform the development of human-centered AI systems and interfaces 'from the ground up' [27]. That is, previous work has focused mainly on general crowdsourcing aspects and failed to address the boundaries and constraints of crowd-AI hybrids for scientific research.

4.2. Open Issues in Crowd-AI Science

Crowd-AI supported research aims to explore the role of AI and IT-mediated crowds on achieving high impact scientific discoveries while generating large amounts of scientific information (e.g., observation statements) to overcome the difficulties experienced by researchers and non-professionals in science. It is clear that much additional work will be needed before a complete understanding of the crowd science projects' success. As future work, we are interested in how to deal with the undisclosed and decentralized nature of crowd work. A lens into the contribution patterns of some crowdsourcing systems has revealed that the highest amount of crowd work is often performed by a small portion of the community members [61]. In this regard, we see continuation of ongoing debates over

crowd-related aspects like motivation, quality control, expertise, intrusiveness, diversity, and crowd bias.

To make crowd science trustworthy, scientists must interact actively and respectfully with crowd workers and volunteers. As a result, new avenues of research stemming directly from these observational practices can be achieved by measuring the potential impact of crowd-AI supported research on institutional science [12]. This is in line with what Watson and Floridi [63] called by ‘socio-technical turn’ in crowd science. So far there have been few attempts to explore complex research tasks (e.g., correlating the contents of scientific papers from different disciplines) through a combination of humans and machines performing complementary activities.

By addressing crowdsourcing and AI as a socio-technical construct, some limitations point to the lack of infrastructural support for crowd workers [64], training machine learning models [52], cheating and manipulating task outputs [56], and query optimization [66]. Researchers have been trying to understand how crowd science projects are deployed and the strategies used to interact with the volunteers [12]. However, there is an increasing difficulty for attaining high degrees of participation and retaining workers in crowd science platforms over time.

On reading the literature, we also come to the view that there are still some barriers concerning the attitudes (i.e., hesitancy) of scientists when confronted with the use and collection of data by non-experts [3]. This supports previous experiments addressing aspects like data handling and preservation, responsibility (e.g., intellectual property), trust (perceptions of data quality), and lack of a shared disciplinary background [60]. Another set of contributions regards the study of complex work in crowd science [26]. We suggest further avenues of research in CSCW to (re)design intelligent systems able to manipulate and make sense of highly complex, unstructured and heterogeneous datasets. To this end, crowd-AI hybrid systems must be augmented with collaboration functionalities to achieve higher levels of participation and coordination.

5. Final Considerations

In this paper, we presented a review of prior work towards a reference model on the socio-technical aspects that should be considered in crowd science. Such work shed a light to the theory and practice of innovative interactive systems, and the results achieved act as a foundation for more complex evaluation exercises to be undertaken. Scientific investigation requires more than technology and there is little knowledge about theoretical frameworks for helping institutions and researchers analyzing concrete

situations and identifying requirements before designing crowd-AI hybrid systems. Such approach can be particularly fruitful for making scientific work more accessible while enriching educational programs and disseminating results.

A possible benefit of crowdsourcing is on a closer integration between human and machine intelligence and we need to deal with the question of what parts of scientific work to crowdsource and how to support these processes with AI. While we are aware of very little work that tries to characterize this burgeoning field, there is a rich area of inquiry for design when exploring the pitfalls and opportunities of crowd-AI supported research to shape scientific activities and socio-technical infrastructures.

Acknowledgements

This research has been sponsored by National Funds through FLAD – Luso-American Development Foundation and FCT – Portuguese Foundation for Science and Technology, in the scope of the project CMU/CS/0012/2017. The work of António Correia is supported by FCT grant SFRH/BD/136211/2018.

References

- [1] N. Nama, M. Sampson, N. Barrowman, R. Sandarage, K. Menon, G. Macartney, K. Murto, J. P. Vaccani, S. Katz, R. Zemek, A. Nasr, and J. D. McNally, “Crowdsourcing the citation screening process for systematic reviews: Validation study”, *Journal of Medical Internet Research*, 21(4), e12953, 2019.
- [2] P. C. Shih, *Beyond human-in-the-loop: Empowering end-users with transparent machine learning*, in Jianlong Zhou and Fang Chen, eds., *Human and Machine Learning: Visible, Explainable, Trustworthy and Transparent*, Springer, Cham, pp. 37-54, 2018.
- [3] E. Law, K. Z. Gajos, A. Wiggins, M. L. Gray, and A. C. Williams, “Crowdsourcing as a tool for research: Implications of uncertainty”, *Proceedings of the ACM Conference on Computer-Supported Cooperative Work and Social Computing*, pp. 1544-1561, 2017.
- [4] R. Vaish, S. N. S. Gaikwad, G. Kovacs, A. Veit, R. Krishna, I. A. Ibarra, C. Simoiu, M. Wilber, S. Belongie, S. Goel, J. Davis, and M. S. Bernstein, “Crowd research: Open and scalable university laboratories”, *Proceedings of the ACM Symposium on User Interface Software and Technology*, pp. 829-843, 2017.
- [5] W. S. Lasecki, “Crowd-powered intelligent systems”, *Human Computation Journal*, 2014.
- [6] W. J. Orlikowski and S. V. Scott, “Sociomateriality: Challenging the separation of technology, work and organization”, *The Academy of Management Annals*, 2(1), pp. 433-474, 2008.
- [7] D. Shapiro, “The limits of ethnography: Combining social sciences for CSCW”, *Proceedings of the ACM Conference*

- on Computer Supported Cooperative Work, pp. 417-428, 1994.
- [8] C. P. Lee and D. Paine, "From the matrix to a model of coordinated action (MoCA): A conceptual framework of and for CSCW", *Proc. of the ACM Conf. on Computer Supported Cooperative Work and Social Computing*, pp. 179-194, 2015.
- [9] A. Kittur, J. V. Nickerson, M. S. Bernstein, E. Gerber, A. Shaw, J. Zimmerman, M. Lease, and J. Horton, "The future of crowd work", *Proceedings of the ACM Conference on Computer-Supported Cooperative Work and Social Computing*, pp. 1301-1318, 2013.
- [10] A. Correia, D. Schneider, B. Fonseca, and H. Paredes, "Crowdsourcing and massively collaborative science: A systematic literature review and mapping study", *Proceedings of the International Conference on Collaboration and Technology*, pp. 133-154, 2018.
- [11] C. Franzoni and H. Sauermaun, "Crowd science: The organization of scientific research in open collaborative projects", *Research Policy*, 43(1), pp. 1-20, 2014.
- [12] K. Scheliga, S. Friesike, C. Puschmann, and B. Fecher, "Setting up crowd science projects", *Public Understanding of Science*, 27(5), pp. 515-534, 2018.
- [13] B. L. Ranard, Y. P. Ha, Z. F. Meisel, D. A. Asch, S. S. Hill, L. B. Becker, A. K. Seymour, and R. M. Merchant, "Crowdsourcing – Harnessing the masses to advance health and medicine, a systematic review", *Journal of General Internal Medicine*, 29(1), pp. 187-203, 2014.
- [14] M. Weiss, "Crowdsourcing literature reviews in new domains", *Technology Innovation Management Review*, 6(2), pp. 5-14, 2016.
- [15] J. Chandler, G. Paolacci, and P. Mueller, *Risks and rewards of crowdsourcing marketplaces*, in Pietro Michelucci, ed., *Handbook of Human Computation*. Springer, New York, pp. 377-392, 2013.
- [16] O. Nov, O. Arazy, and D. Anderson, "Scientists@Home: What drives the quantity and quality of online citizen science participation?", *PloS One*, 9(4), e90375, 2014.
- [17] P. André, A. Kittur, and S. P. Dow, "Crowd synthesis: Extracting categories and clusters from complex data", *Proceedings of the ACM Conference on Computer-Supported Cooperative Work and Social Computing*, pp. 989-998, 2014.
- [18] B. M. Good, M. Nanis, C. Wu, and A. I. Su, "Microtask crowdsourcing for disease mention annotation in PubMed abstracts", *Proceedings of the Pacific Symposium on Biocomputing*, pp. 282-293, 2014.
- [19] M. A. O'Leary, K. Alphonse, A. H. Mariangeles, D. Cavaliere et al., "Crowds replicate performance of scientific experts scoring phylogenetic matrices of phenotypes", *Systematic Biology*, 67(1), pp. 49-60, 2017.
- [20] A. W. Brown and D. B. Allison, "Using crowdsourcing to evaluate published scientific literature: Methods and example", *PloS One*, 9(7), e100647, 2014.
- [21] C. Bonacchi, A. Bevan, D. Pett, A. Keinan, R. Sparks, J. Wexler, and N. Wilkin, "Crowd-sourced archaeological research: The MicroPasts project", *Archaeology International*, 17, pp. 61-68, 2014.
- [22] G. Dawson, C. Lintott, and S. Shuttleworth, "Constructing scientific communities: Citizen science in the nineteenth and twenty-first centuries", *Journal of Victorian Culture*, 20(2), pp. 246-254, 2015.
- [23] B. M. Good and A. I. Su, "Games with a scientific purpose", *Genome Biology*, 12, Article 135, 2011.
- [24] L. B. Chilton, J. Kim, P. André, F. Cordeiro, J. A. Landay, D. S. Weld, S. P. Dow, R. C. Miller, and H. Zhang, "Frenzy: Collaborative data organization for creating conference sessions", *Proc. of the ACM CHI Conference on Human Factors in Computing Systems*, pp. 1255-1264, 2014.
- [25] B. Tomlinson, J. Ross, P. André, E. Baumer, D. Patterson, J. Corneli et al., "Massively distributed authorship of academic papers", *CHI'12 Extended Abstracts on Human Factors in Computing Systems*, pp. 11-20, 2012.
- [26] K. Crowston, E. Mitchell, and C. Østerlund, "Coordinating advanced crowd work: Extending citizen science", *Proceedings of the Hawaii International Conference on System Sciences*, pp. 1681-1690, 2018.
- [27] D. Dellermann, A. Calma, N. Lipusch, T. Weber, S. Weigel, and P. Ebel, "The future of human-AI collaboration: A taxonomy of design knowledge for hybrid intelligence systems", *Proceedings of the Hawaii International Conference on System Sciences*, 2019.
- [28] A. J. Quinn, B. B. Bederson, T. Yeh, and J. Lin, "Crowdfow: Integrating machine learning with Mechanical Turk for speed-cost-quality flexibility", *Better Performance Over Iterations*, 2010.
- [29] A. J. Quinn and B. B. Bederson, "Human computation: A survey and taxonomy of a growing field", *Proceedings of the ACM CHI Conference on Human Factors in Computing Systems*, pp. 1403-1412, 2011.
- [30] T. Buecheler, J. H. Sieg, R. M. Fuchslin, and R. Pfeifer, "Crowdsourcing, open innovation and collective intelligence in the scientific method: A research agenda and operational framework", *Proceedings of the International Conference on the Synthesis and Simulation of Living Systems*, pp. 679-686, 2010.
- [31] H. H. N. Do, M. K. Chandrasekaran, P. S. Cho, and M. Y. Kan, "Extracting and matching authors and affiliations in scholarly documents", *Proceedings of the ACM/IEEE-CS Joint Conference on Digital Libraries*, pp. 219-228, 2013.
- [32] J. Cheng and M. S. Bernstein, "Flock: Hybrid crowd-machine learning classifiers", *Proceedings of the ACM Conference on Computer Supported Cooperative Work and Social Computing*, pp. 600-611, 2015.
- [33] Z. Dong, J. Lu, T. W. Ling, J. Fan, and Y. Chen, "Using hybrid algorithmic-crowdsourcing methods for academic knowledge acquisition", *Cluster Computing*, 20(4), pp. 3629-3641, 2017.
- [34] J. Chan, J. C. Chang, T. Hope, D. Shahaf, and A. Kittur, "SOLVENT: A mixed initiative system for finding analogies between research papers", *Proceedings of the ACM Human-Computer Interaction*, 2, CSCW, Article 31, 2018.
- [35] J. Ramirez, E. Krivosheev, M. Baez, F. Casati, and B. Benatallah, "CrowdRev: A platform for crowd-based screening of literature reviews", *arXiv:1805.12376*, 2018.
- [36] J. Hughes, W. W. Sharrock, T. Rodden, J. O'Brien, M. Rouncefield, and D. Calvey, *Perspectives on the social organization of work*, in Field Studies and CSCW, COMIC Deliverable 2.2, pp. 129-160, Esprit Basic Research Project 6225: Lancaster University, pp. 129-160, 1994.
- [37] R. C. Nickerson, U. Varshney, and J. Muntermann, "A method for taxonomy development and its application in

- information systems”, *European Journal of Information Systems*, 22(3), pp. 336-359, 2013.
- [38] J. Webster and R. T. Watson, “Analyzing the past to prepare for the future: Writing a literature review”, *MIS Quarterly*, xiii-xxiii, 2002.
- [39] K. R. Felizardo, E. Mendes, M. Kalinowski, É. F. Souza, and N. L. Vijaykumar, “Using forward snowballing to update systematic reviews in software engineering”, *Proceedings of the ACM/IEEE International Symposium on Empirical Software Engineering and Measurement*, p. 53, 2016.
- [40] M. Schmidt-Kraepelin, S. Thiebes, M. C. Tran, and A. Sunyaev, “What’s in the game? Developing a taxonomy of gamification concepts for health apps”, *Proceedings of the Hawaii International Conference on System Sciences*, 2018.
- [41] R. Houghton, J. Sprinks, J. Wardlaw, S. Bamford, and S. Marsh, “A sociotechnical system approach to virtual citizen science: An application of BS ISO 27500:2016”, *Journal of Science Communication*, 18(1), 2019.
- [42] H. Xia, C. S. Østerlund, B. McKernan, J. E. Folkestad, P. G. C. Rossini, O. Boichak, J. L. Robinson, K. Kenski, R. Myers, B. A. Clegg, and J. Stromer-Galley, “TRACE: A stigmergic crowdsourcing platform for intelligence analysis”, *Proceedings of the Hawaii International Conference on System Sciences*, 2019.
- [43] A. Correia, S. Jameel, H. Paredes, B. Fonseca, and D. Schneider, *Hybrid machine-crowd interaction for handling complexity: Steps toward a scaffolding design framework*, in *Macrotask Crowdsourcing: Engaging the Crowds to Address Complex Problems*, Springer, Cham, pp. 149-161, 2019.
- [44] C. Bonacchi, A. Bevan, A. Keinan-Schoonbaert, D. Pett, and J. Wexler, “Participation in heritage crowdsourcing”, *Mus. Manag. Curatorship*, 34(2), pp. 166-182, 2019.
- [45] M. L. Gray, S. Suri, S. S. Ali, and D. Kulkarni, “The crowd is a collaborative network”, *Proceedings of the ACM Conference on Computer Supported Cooperative Work and Social Computing*, pp. 134-147, 2016.
- [46] A. Dipple, K. Raymond, and M. Docherty, “General theory of stigmergy: Modelling stigma semantics”, *Cognitive Systems Research*, 31, pp. 61-92, 2014.
- [47] K. Schmidt and L. Bannon, “Taking CSCW seriously”, *Computer Supported Cooperative Work*, 1(1-2), pp. 7-40, 1992.
- [48] D. A. Döppner, P. Derckx, and S. Schoder, “Symbiotic co-evolution in collaborative human-machine decision making: Exploration of a multi-year design science research project in the Air Cargo Industry”, *Proceedings of the Hawaii International Conference on System Sciences*, 2019.
- [49] L. Hetmank, “Components and functions of crowdsourcing systems – A systematic literature review”, *Wirtschaftsinformatik*, 4, 2013.
- [50] R. Tinati, M. Van Kleek, E. Simperl, M. Luczak-Rösch, R. Simpson, and N. Shadbolt, “Designing for citizen data analysis: A cross-sectional case study of a multi-domain citizen science platform”, *Proceedings of the ACM CHI Conference on Human Factors in Computing Systems*, pp. 4069-4078, 2015.
- [51] C. Lofi and K. El Maarry, “Design patterns for hybrid algorithmic-crowdsourcing workflows”, *Proceedings of the IEEE Conference on Business Informatics*, pp. 1-8, 2014.
- [52] J. W. Vaughan, “Making better use of the crowd: How crowdsourcing can advance machine learning research”, *Journal of Machine Learning Research*, 18(193), pp. 1-46, 2018.
- [53] T. W. Malone and K. Crowston, “What is coordination theory and how can it help design cooperative work systems?”, *Proceedings of the ACM Conference on Computer-Supported Cooperative Work*, pp. 357-370, 1990.
- [54] A. Correia, B. Fonseca, H. Paredes, D. Schneider, and S. Jameel, “Development of a crowd-powered system architecture for knowledge discovery in scientific domains”, *Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics*, 2019.
- [55] H. Riesch and C. Potter, “Citizen science as seen by scientists: Methodological, epistemological and ethical dimensions”, *Public Understanding of Science*, 23(1), pp. 107-120, 2014.
- [56] F. Daniel, P. Kucherbaev, C. Cappiello, B. Benatallah, and M. Allahbakhsh, “Quality control in crowdsourcing: A survey of quality attributes, assessment techniques, and assurance actions”, *ACM Computing Surveys*, 51(1), Article 7, 2018.
- [57] K. Täuscher, “Leveraging collective intelligence: How to design and manage crowd-based business models”, *Business Horizons*, 60(2), pp. 237-245, 2017.
- [58] S. Amershi, D. Weld, M. Vorvoreanu, A. Fournay, B. Nushi, P. Collisson, J. Suh, S. Iqbal, P. N. Bennett, K. Inkpen, J. Teevan, R. Kikin-Gil, and E. Horvitz, “Guidelines for human-AI interaction”, *Proceedings of the ACM CHI Conference on Human Factors in Computing Systems*, 2019.
- [59] M. S. Ackerman, “The intellectual challenge of CSCW: The gap between social requirements and technical feasibility”, *Human-Computer Interaction*, 15(2-3), pp. 179-203, 2000.
- [60] K. Williamson, M. A. Kennan, G. Johanson, and J. Weckert, “Data sharing for the advancement of science: Overcoming barriers for citizen scientists”, *Journal of the Association for Information Science and Technology*, 67(10), pp. 2392-2403, 2016.
- [61] A. Basiri, M. Haklay, and Z. Gardner, “The impact of biases in the crowdsourced trajectories on the output of data mining processes”, *Proceedings of the AGILE Conference on Geographic Information Science*, 2018.
- [62] A. Smirnov and A. Ponomarev, *Decision support based on human-machine collective intelligence: Major challenges*, in *Internet of Things, Smart Spaces, and Next Generation Networks and Systems*, Springer, Cham, pp. 113-124, 2019.
- [63] D. Watson and L. Floridi, “Crowdsourced science: Sociotechnical epistemology in the e-research paradigm”, *Synthese*, 195(2), pp. 741-764, 2018.
- [64] B. V. Hanrahan, D. Martin, J. Willamowski, and J. M. Carroll, “Investigating the Amazon Mechanical Turk market through tool design”, *Computer Supported Cooperative Work*, 27(3-6), pp. 1255-1274, 2018.
- [65] J. L. Shirk, H. L. Ballard, C. C. Wilderman, T. Phillips, A. Wiggins, R. Jordan, E. McCallie, M. Minarchek, B. V. Lewenstein, M. E. Krasny, and R. Bonney, “Public participation in scientific research: A framework for deliberate design”, *Ecology and Society*, 17(2), 2012.
- [66] G. Li, J. Wang, Y. Zheng, and M. J. Franklin, “Crowdsourced data management: A survey”, *IEEE Transactions on Knowledge and Data Engineering*, 28(9), pp. 2296-2319, 2016.