Quo Vadis, HCOMP? A Review of 12 Years of Research at the Frontier of Human Computation and Crowdsourcing

Jonas Oppenlaender
jonas.oppenlaender@oulu.fi
University of Oulu
Oulu, Finland

Ujwal Gadiraju
U.K.Gadiraju@tudelft.nl
Delft University of Technology
Delft, The Netherlands

Simo Hosio
simo.hosio@oulu.fi
University of Oulu
Oulu, Finland

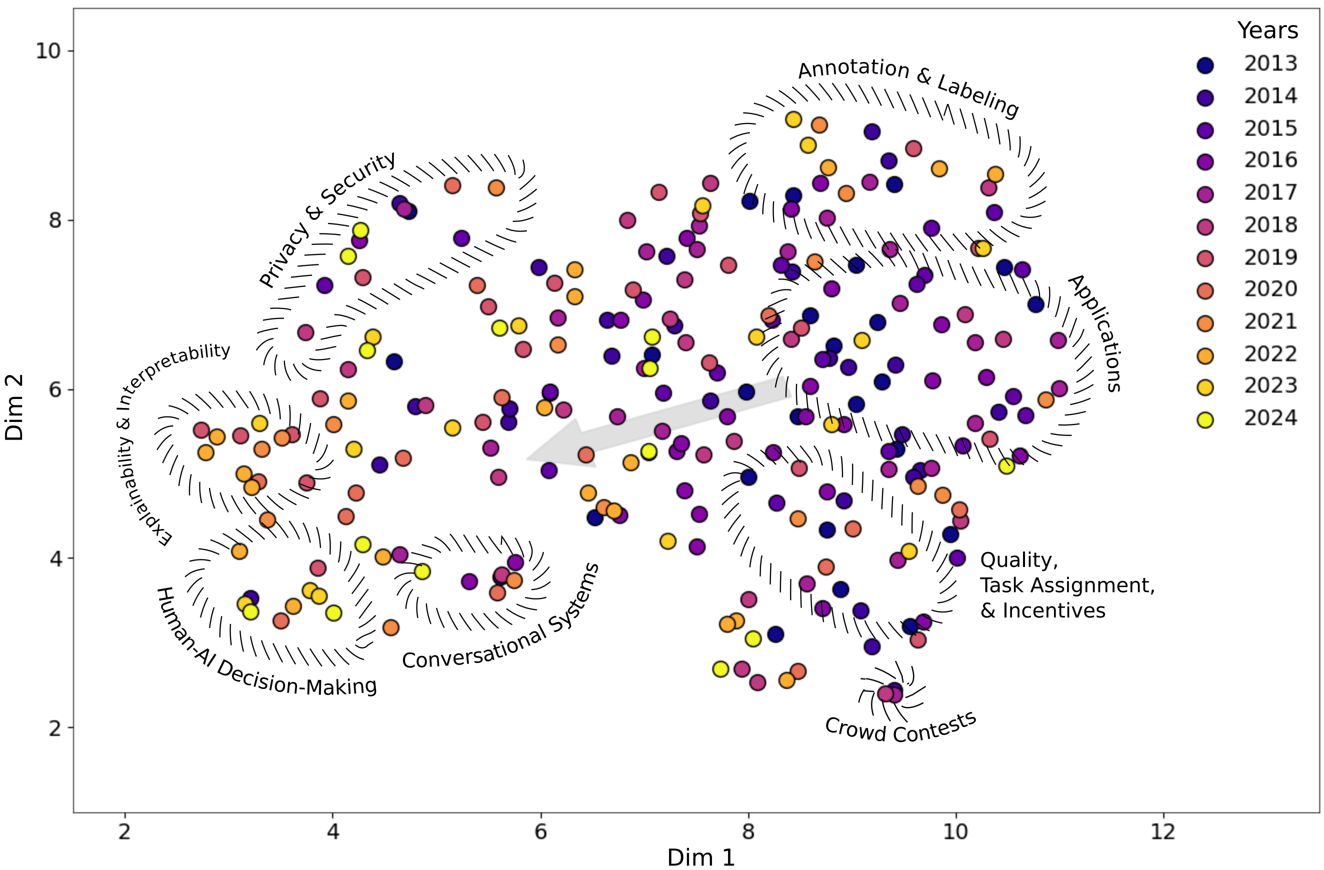


Figure 1: Research topics in articles published at the Conference on Human Computation and Crowdsourcing (HCOMP) between 2013 and 2024. The dots represent article titles embedded using sentence transformers and projected into two-dimensional space with a dimensionality reduction technique (UMAP). The arrow indicates the general direction of the HCOMP Conference from 2013 to 2024 (centroid to centroid). Key themes from 2013 and 2024 are annotated, demonstrating how many articles in HCOMP have migrated away from HCOMP's traditional key motor themes (such as annotation & labeling, quality, incentives & task assignment, and applications) toward the topics of explainable AI (XAI), conversational systems, and human-AI decision-making. An interactive visualization is available at https://hcomp-retrospective.github.io.

Abstract

The field of human computation and crowdsourcing has historically studied how tasks can be outsourced to humans. However, many tasks previously distributed to human crowds can today be completed by generative AI with human-level abilities, and concerns about crowdworkers using language models to complete tasks are surfacing. These developments undermine core premises of the

CI 2025, San Diego, CA, USA

ACM ISBN 979-8-4007-1489-4/2025/08
https://doi.org/10.1145/3715928.3737467

field. In this paper, we examine the evolution of the Conference on Human Computation and Crowdsourcing (HCOMP)—a representative example of the field as one of its key venues—through the lens of Kuhn's paradigm shifts. We review 12 years of research at HCOMP, mapping the evolution of HCOMP's research topics and identifying significant shifts over time. Reflecting on the findings through the lens of Kuhn's paradigm shifts, we suggest that these shifts do not constitute a paradigm shift. Ultimately, our analysis of gradual topic shifts over time, combined with data on the evident overlap with related venues, contributes a data-driven perspective to the broader discussion about the future of HCOMP and the field as a whole.

CCS Concepts

• **Information systems → Crowdsourcing**.

Keywords

crowdsourcing, human computation, HCOMP, science of science, meta-research

ACM Reference Format:

Jonas Oppenlaender, Ujwal Gadiraju, and Simo Hosio. 2025. Quo Vadis, HCOMP? A Review of 12 Years of Research at the Frontier of Human Computation and Crowdsourcing . In *Collective Intelligence Conference (CI 2025), August 4–6, 2025, San Diego, CA, USA.* ACM, New York, NY, USA, 14 pages. https://doi.org/10.1145/3715928.3737467

1 Introduction

The field of human computation and crowdsourcing has long relied on harnessing human ingenuity to address complex problems. Foundational work—such as Luis von Ahn's early contributions with projects such as the ESP Game and CAPTCHA [104, 105]—established a paradigm for leveraging human input online. Crowdsourcing [50] emerged as a productive research field, exploring the theoretical and practical dimensions of distributing tasks to a crowd. Over the years, the field evolved through what can be seen as a period of "normal science" [60], focused on solving fundamental issues in crowdsourcing—optimizing task design and incentives, ensuring data quality, exploring novel workflows, and refining models of human interaction—all while operating within a well-defined set of assumptions and methods [28, 56, 63].

All scientific fields evolve, adapting to new developments and emerging challenges. In recent years, however, the rapid progress of artificial intelligence has begun to shake the foundations of fields concerned with human input, labor, and cognition. Human computation and crowdsourcing is one such field. Tasks once assigned to human workers can now be performed at least partially by large language models, raising questions about the role of human input in crowdsourcing [111]. Concerns have also emerged that crowdworkers may be increasingly relying on automated tools to complete tasks, potentially undermining core premises of human computation [35, 103]. Other related developments, such as data-labeling firms rebranding as AI companies, further continue to disrupt the established framework of crowdsourcing. These developments challenge foundational assumptions in the field: that tasks must be decomposed for distributed human labor; that quality requires redundancy and human judgment; and that diverse worker perspectives are a core asset. LLMs offer an alternative model that scales at near-zero marginal cost, performs end-to-end task pipelines without human oversight, and simulates diverse linguistic and cultural inputs. As hybrid systems emerge—using LLMs to filter, generate, or evaluate crowd outputs—they expose the inadequacy of existing task taxonomies and evaluation methods. Established concerns around worker fairness, motivation, and error variance may give way to new challenges around prompt design, model hallucination, and synthetic bias. These developments offer a timely opportunity to revisit the field's scope, assumptions, and to envision its possible future directions.

In this paper, we investigate shifts at the Conference on Human Computation and Crowdsourcing (HCOMP) as a proxy into the wider field of research on human computation and crowdsourcing. We adopt Kuhn's notion of paradigm shifts [60] as a lens to examine the evolution of the HCOMP conference. Kuhn's model characterizes scientific progress as a series of distinct phases. In the pre-science phase, a field lacks consensus, and diverse, often conflicting theories coexist. This is followed by a period of normal science, during which a dominant paradigm emerges and research focuses on solving puzzles within that established framework. As anomalies and unexpected findings accumulate, the field may enter a crisis, challenging the core assumptions of the prevailing paradigm. If these challenges cannot be reconciled, a revolutionary phase occurs, leading to a paradigm shift in which the old framework is replaced by a new one that redefines the discipline.

One could argue that the role of human input being redefined already constitutes a revolutionary phase. The apparent challenges—brought about by the disruptive influence of generative AI—further suggest a form of incommensurability between the established paradigm and the new realities imposed by large language models, potentially leading to non-linear progress that defies past metrics of evaluation. Moving from "normal science" to revolutionary science requires questioning the fundamental aspects of the field (such as the need for human input) and an exploration of alternatives. And a true paradigm shift involves the fundamental reconceptualization of a field's underlying principles rather than merely a cumulative improvement of existing methods. It is worth questioning whether there has been a paradigm shift at HCOMP, or whether we are merely witnessing a gradual, natural shift in topics.

Our work undertakes a detailed analysis of research published at the HCOMP conference with a multi-method approach, to capture both the historical evolution and emerging trends within the community. We begin by employing embedding techniques and clustering algorithms to map research topics and identify shifts over time. Further, we compare the HCOMP conference with six related conferences by measuring the cosine similarity of article title embeddings, which allows us to speculate on the future trajectory of the field. This is complemented by a co-word analysis that examines the relationships among key terms at HCOMP and across conferences. Finally, we measure shifts in research topics at HCOMP with the aim of identifying whether a paradigm shift has taken place at HCOMP. Together, our analysis provides a comprehensive view of the evolution of HCOMP, illuminating both the enduring strengths of the traditional paradigm during the period of "normal science" and the recent disruptive challenges introduced by generative AI.

We contribute:

- An empirical investigation into the evolution of the HCOMP conference, the key venue for research on human computation and crowdsourcing. We highlight recent developments and fundamental shifts in the conference's research topics and analyze co-occurring words.
- An investigation of shifts at the HCOMP conference in relation to six related conferences—Collective Intelligence (CI), CSCW, FAccT, IUI, UMAP, and AAMAS—providing valuable insights to inform the future of HCOMP.
- A discussion of these findings through the lens of Kuhn's model of paradigm shifts. Our work can help inform others wishing to analyze the evolution of research at scientific venues in a similar way.

By framing the discussion in terms of a potential paradigm shift, we explore the critical juncture in the evolution of human computation and crowdsourcing, marked by the transformative impact of generative AI. This perspective highlights the crisis of reconciling traditional methods with new technological capabilities and invites a broader discussion on the future direction of the field of human computation and crowdsourcing.

2 Related Work

2.1 Kuhn's Paradigm Shifts

Kuhn's model of scientific progress [60] offers a framework for understanding how disciplines evolve through four distinct phases:

(1) *Pre-science*: In this initial phase, a field lacks a unified theoretical framework. Researchers pursue diverse and often conflicting approaches without a shared set of standards or observational criteria. This period is characterized by debates over fundamentals, where as many theories exist as there are theorists.
(2) *Normal science*: Once a dominant paradigm is established, the field enters a phase of normal science. Researchers work within this established framework, addressing puzzles and refining existing methods rather than challenging the core assumptions. Anomalies—observations that do not easily fit the paradigm—are typically treated as challenges to be solved within the current structure, rather than reasons to question it.
(3) *Crisis*: Over time, if anomalies accumulate and prove resistant to resolution, confidence in the established paradigm begins to wane. This phase is marked by a growing sense of crisis, as the foundational assumptions of the field are increasingly questioned. Researchers start to explore alternative explanations, and competing theories emerge to address the persistent anomalies.
(4) *Revolution*: Should the crisis remain unresolved, the field may undergo a revolutionary shift. In this phase, a new paradigm emerges—one that redefines the field's basic principles and methods. The new framework is not simply an extension of the old one but represents a fundamental change in how problems are understood and approached. Kuhn emphasizes that this shift is driven by both empirical findings and sociological factors, making the transition complex and non-linear.

Kuhn's model has been applied in computer science. For instance, the model was used to frame developments in the field of computer vision, where researchers eagerly adopted advances in deep learning [58]. In other fields and scientific disciplines, deep learning has also had a strong impact, enabling new ways of science [13]. Another example is prompt-based learning (i.e., prompting large pre-trained models), which brought paradigm shifts in the fields of AI and Natural Language Processing [67]. In Human-Computer Interaction, using synthetic participants (e.g., for usability testing) is a growing trend [75]. Simulating users with generative AI is a new frontier that fundamentally challenges the traditional assumption that HCI studies must involve human participants [5, 75, 98]. Another example is, arguably, education which is undergoing a shift brought about by generative AI [39].

Kuhn's model provides a useful lens through which to view the evolution of research in crowdsourcing and human computation. In the following section, we provide a retrospective on HCOMP's phase of "normal science."

2.2 A Retrospective on HCOMP's Period of "Normal Science"

During HCOMP's phase of normal science, several research topics served as key motor themes for the field. Early work focused on quality control as a fundamentally important aspect of human computation and crowdsourcing. The quality of crowdsourced responses was found to be a critical bottleneck in many applications. As early as in the year 2008, Kittur et al. noted in their crowdsourced user studies that almost 50% of the responses on Amazon Mechanical Turk "consisted of uninformative responses including semantically empty [...], non-constructive [...], or copy-and-paste responses" [55]. This wasteful ratio of good to bad responses persisted over the years, with quality-control methods being proposed to overcome existing challenges. This paved way for the use of gold-standard questions [80], post-hoc filtering [25], statistical and algorithmic methods to control for quality [9], collusion detection [54], pre-task worker selection and behavior-based quality control methods [38] emerging as key methods for improving response quality. Approaches from psychology and survey research, such as Instructional Manipulation Checks (IMC) [81], were adopted by the field of crowdsourcing. Over time, crowd workers adapted to the evolving quality control measures and were found to be more attentive to IMC than other human subject pools [46], and Checco et al. later demonstrated how gold questions can be gamed [21]. There was also a growing interest in quality control within citizen science initiatives, exploring a different set of intrinsic incentives for participation [16, 51, 109].

Leveraging crowdsourcing methods to address real-world problems and use cases was another strong research stream at HCOMP during the period of "normal science." Applications included, for instance, the synthesis of information [69], paper screening for literature reviews [59], augmenting video [94], conference scheduling [12], and a genomics game [100]. Annotation and labeling was another large area of focus at HCOMP during this period (see Figure 1), with research on methods and algorithms for aggregating labels to fuel training of computer vision models. Notably, work by Sheshadri and Lease to improve response aggregation methods in crowdsourcing was impactful [99]. The authors presented an open source shared task framework including benchmark datasets, defined tasks, standard metrics, and reference implementations with empirical results for popular methods at the time.

While there has been a strong focus on microtasks at HCOMP, applications in alternative areas were also explored, such as citizen science [112], crowdfunding [49], and crowd contests [20, 96]. We also observed geo-enabled applications such as spatial crowdsourcing and crowdsensing, with notable examples such as earthquake detection using citizen science [70], local crowds for event reporting [4], and participatory sensing [118]. Real-time applications started becoming a topical focus at HCOMP in 2016, including works on real-time question-answering [97], real-time disease information [77], and real-time assistance in real environments [2, 42].

Workflow and task design have also received strong attention in the HCOMP community [41, 71, 110]. Cost-quality-time optimization [40], predicting label quality [52], or aggregation mechanisms [106] were some objectives pursued in this direction. Task routing and incentive design have received keen interest, too. For instance, parallelization of tasks [17], skill and stress aware task assignment [61], and dynamic task assignment to crowd workers versus AI [57] have been explored. Different pricing schemes [29] or incentives to increase engagement and counter bias [36] have been explored. Monetary interventions were utilized to prevent task switching [116] and to predict work quality [115].

In the years that followed, researchers explored agreement and disagreement mechanisms [22], linguistic frame disambiguation [32], and the use of dummy events to improve worker engagement [33]. Others explored training workers and leveraging worker skills in different contexts, such as providing stress management support [3], or music annotation [95], and developed methods to ensure fair wages [108] or support novice workers [90]. Over the years, efforts have also been invested to understand crowd worker behavior—including workers' strategies to maximize earnings [53], their goal-setting behavior [1]—and improve worker experiences in different contexts [26, 48] and worker communities [114, 117]. Others explored alternative input modalities to lower the barrier for participation in crowd work [6, 101].

In this paper, we investigate how research in the field of human computation and crowdsourcing has shifted from "normal science" to a new phase over the past twelve years. To do so, we adopt a multi-method approach, which we detail in the following section.

3 Method

We analyze shifts at the HCOMP conference from multiple perspectives through the lens of Kuhn's model. In the following, we describe our data collection and analysis.

3.1 Data Collection

We collected the titles and abstracts of all research articles (N = 250) published at the HCOMP conference from 2013 to 2024. Each year's proceedings include between 14 and 27 articles (Mean = 20.8, SD = 4.3). The data was scraped from the website of the Association for the Advancement of Artificial Intelligence (AAAI). Title lengths range from 3 to 28 tokens (Mean = 10.7, Median = 10). The collected data was analyzed using multiple methods, as described below.

3.2 Data Analysis

3.2.1 Initial exploration. The lead author started to explore the proceedings of the HCOMP conference to develop an overall understanding of the venue by using Voyant Tools [92]. To this end, titles and abstracts were merged and, using Voyant, several visualizations were created to initially explore the corpus. We then proceeded to review all works published at HCOMP, focusing on titles and abstracts, to identify research themes and topics at the conference. This exploration and review informed sections 2.2 and 4.1.

3.2.2 Topic analysis. To identify relationships between topics, we encoded the article titles into embeddings using Sentence Transformers [91] (all-mpnet-base-v2) and used UMAP [72] to project the embeddings into a two-dimensional space. UMAP is a dimensionality reduction technique which preserves local and global structures better than t-SNE and PCA [24, 72]. The sentence transformer captures contextual relationships, word order, and deeper semantics. As a result, the embedding space reflects semantic similarity: titles with similar meaning are positioned closer together. We used clustering to identify the approximate locations of topics in embedding space by iteratively applying HDBSCAN [19], a density-based clustering method, with different parameters. The exploration of different clustering solutions allowed us to get an overview of the structure of the embedding space and the trends within. We manually annotated the clusters and indicate the general trend with an arrow, which we calculated from the embedding centroids of the 2013 and 2024 HCOMP proceedings. The centroid is the 'mean embedding' (i.e., the point in space that, on average, is closest to all other data points in a given year). Further, we mapped how HCOMP topics, as identified by the clustering algorithm, have evolved over time (see figures 1, 3, and 7).

3.2.3 Paradigm shift. We use the notion of a Gestalt-shift in the context of Kuhn's framework to measure whether a sudden shift in research topics has taken place at the HCOMP conference. The idea is to measure the cosine distance between the embedding centroids of article titles across consecutive years. This allows us to assess whether a shift in research topics has occurred, and when it took place. A sharp increase in cosine distance between centroids from one year to the next would suggest a sudden shift in research focus. Of course, whether a detected shift constitutes a paradigm shift is arguable, since it is not clear what magnitude of shift would constitute a paradigm shift. Or in other words, how far would the HCOMP conference need to move away from its traditional research topics to constitute a paradigm shift? Given how the HCOMP conference is affected by recent developments in AI, we expect there to be a notable shift in research topics in recent years. The results are depicted in figures 1, 4, 7, and 9.

3.2.4 Conference analysis. To inform decision-making on the future of the HCOMP conference and trigger reflection in the HCOMP community, we used the same approach as in Section 3.2.2 and encoded the titles of articles published at six related conferences from 2013–2024: ACM Collective Intelligence Conference (CI; N = 220), ACM SIGCHI Conference on Computer-Supported Cooperative Work & Social Computing (CSCW; $N = 3,081$), ACM Conference on Fairness, Accountability, and Transparency (FAccT; $N = 657$), ACM Conference on Intelligent User Interfaces (IUI; $N = 612$), ACM Conference on User Modeling, Adaptation and Personalization (UMAP; $N = 232$), and the Conference on Autonomous Agents and Multiagent Systems (AAMAS; $N = 2,203$). This set of conferences was selected for comparison with HCOMP for several reasons. All conferences have, in part, some topical overlap with HCOMP, and CI and UMAP are of similar size. CSCW was chosen for it having a strong representation of crowdsourcing research in the past (around 2012–2014), with many HCOMP authors also publishing at CSCW. FAccT, UMAP, IUI, and AAMAS were selected for, potentially, being relevant to recent research at HCOMP. We decided not to include ACM CHI, because it is a large and very diverse conference, with only a tiny fraction of the published articles relating to crowdsourcing and human computation. Note that for ACM CI, some older proceedings were no longer accessible. We plot the resulting embeddings into twodimensional space using UMAP. Since embeddings are numeric vector representations of semantic meaning encoded in text, the plots give us a topical overview of HCOMP's relation to other related conferences and the direction of the recent shift in HCOMP, in terms of centroid cosine distance of conference proceedings. The results are depicted in Table 1, Figure 3, and Figure 7.

3.2.5 Co-word analysis. To complement our analysis of topics and conferences, we analyzed co-words in the titles and abstracts of HCOMP articles (excluding stopwords). Co-words are co-occurring words that are frequently used together in a sentence. We counted the frequency of co-word pairs, treating them as unordered (i.e., ignoring the order of terms in a co-word pair). We then plotted the frequency of these co-words at the HCOMP conference over time, including only those that appeared in more than one year (see Figure 2). Further, we compared shared keywords in the titles of articles at HCOMP and the six related conferences (see figures 5 and 6).

4 Results

4.1 Recent Shift in Topics

The initial years of HCOMP, as discussed earlier, were focused on optimizing and addressing issues around crowd work, but also applications of crowdsourcing. Since 2018, we can observe a gradual shift of research topics studied at HCOMP. Since then, HCOMP shifted toward tackling problems at the intersection of humans and AI systems (as represented in the bottom-left of Figure 1). With the growing advances in machine learning and recognizing important societal implications, the HCOMP community began to address challenges around bias and fairness [14, 30, 79, 83, 84], interpretability [62, 74], explainability [47, 64, 78, 89], privacy, trust and reliance on AI systems [7, 11, 34], human-AI decision making [43, 68, 79, 88, 113], human-AI team performance [11], collaborative human-AI methods [65, 119], and AI risks [15].

This shift in research focus is also evident in our co-word analysis (see Figure 2), where the co-word pairs *task–worker* and *crowd–worker* ceased to be present in the HCOMP titles and abstracts in 2021 and 2023, respectively. Instead, the HCOMP community moved to using more human-centered co-word pairs, such as *human–behavior*, *AI–crowd*, and *human–AI*. Our analysis also shows that this broadening in perspectives does not coincide with the introduction of OpenAI's popular ChatGPT language model in 2022. Instead, a reorientation is notable as early as 2019, with clear changes in co-word pairs becoming notable in 2021, one year before OpenAI introduced ChatGPT (cf. Figure 2). In that year, OpenAI released GPT-3 [18], a language model that with 175 billion parameters had over 100 times the size of its predecessor GPT-2. Perhaps it was this new model that raised both interest in AI but also heightened concerns in the HCOMP community.

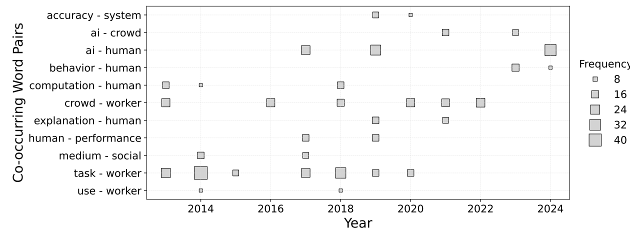


Figure 2: Co-word occurrences at the HCOMP conference over time (order-insensitive, considering only co-words that appear in more than one year)

Already in 2018, the HCOMP community started to demonstrate concerns raised by increasingly intelligent automation tools. One notable incident occurred in mid-2018, when researchers outside the HCOMP community reported a decline in the quality of crowdsourced data, along with responses that appeared to be generated by "bots," speculating that fraudulent activity and potentially automation was at play [10, 23, 31, 93, 102]. In their blog posts, Moss and Litman later concluded that this incident was likely due to "farmers"—i.e., workers using 'server farms' for submitting HITs [66, 76]. In the same year, Kaplan et al. studied work strategies and tool use among crowd workers [53]. Automation tools have, of course, been used by workers for long already, but more prominently for task management than data generation. In the hands of crowd workers, the use of automated tools for generating answers to tasks is a threat to the validity of data collected on crowdsourcing platforms. These developments highlighted growing tensions between human labor and automation on crowdsourcing platforms, coinciding with a broader shift in the HCOMP community's focus.

Since then, the commoditization of AI has drawn interest from some members of the HCOMP community to research topics that fall within the focus of other venues. Specifically, some recent research at HCOMP now strongly relates to topics studied at ACM FAccT (see Figure 3), a conference focusing on issues such as algorithmic transparency, fairness in machine learning, explainability and interpretability, bias, and ethics. By cosine similarity of embedded article titles, ACM FAccT is, on average, most similar to HCOMP today (see Table 1). Examples of works published at HCOMP include the work by Lage et al. on factors that make machine learning models interpretable by humans [62], Ray et al.'s work on evaluating the efficacy of explanations in human-AI collaborative tasks [89], and Hase et al.'s work on interpretability of vision models

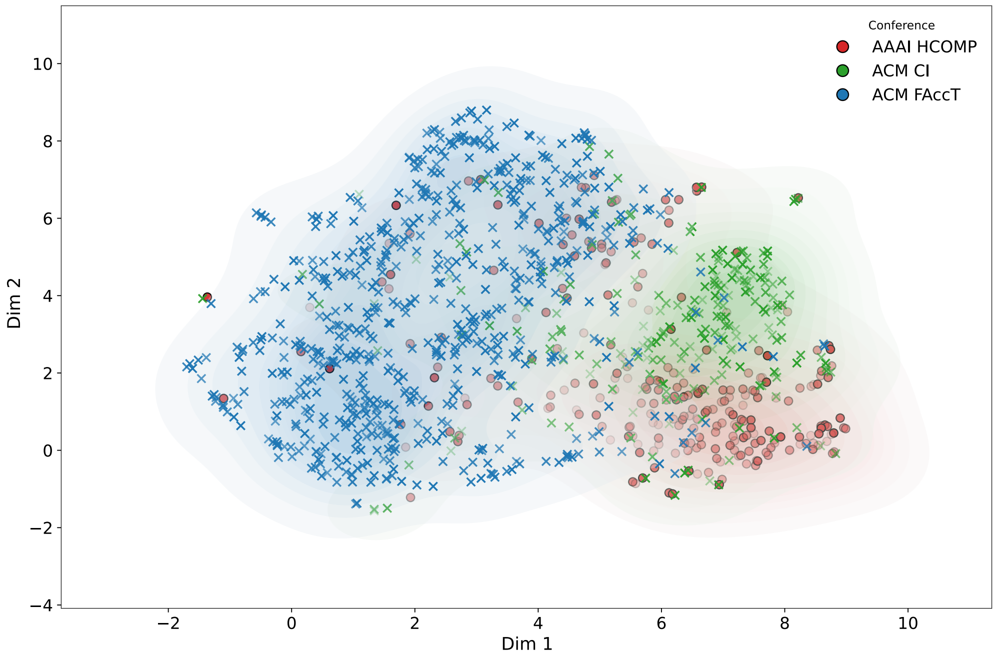


Figure 3: Comparison of article titles in HCOMP, ACM Collective Intelligence (CI) and ACM Conference on Fairness, Accountability, and Transparency (FAccT). Titles of recent articles are more opaque.

with hierarchical prototypes [45]. These examples suggest that interpretability and explainability have emerged as novel themes at HCOMP, reflecting a broadening of the community's scope beyond traditional crowdsourcing paradigms.

As the field of HCOMP evolved over the years, a growing similarity can also be noted with other conferences, in terms of centroid distances of article title embeddings (see Table 1) and shared keywords (see figures 5 and 6). There is also a strong overlap in relevant keywords in article titles between HCOMP and ACM CSCW (see Figure 5). Recently, HCOMP has also moved closer to the research spaces of ACM IUI, with its intelligent user interfaces providing a point of interaction between humans and AI, and ACM UMAP, which explores user modeling and personalization as a foundation for adaptive human-AI systems (see Figure 7). However, on average, HCOMP remains closely related to ACM CI (see Figure 3).

In all fairness, some overlap exists between all investigated conferences, as depicted in Figure 8. Our analysis of centroids can only be an approximation of how conferences, as a whole, have developed over time. It is interesting to note that some conferences—in particular CSCW, AAMAS, but also FAccT—demonstrate very little year-by-year centroid movements (see Figure 4), which may speak to the stability of research topics at these conferences. In the following section, we investigate whether a "Gestalt-shift" has taken place at the HCOMP conference.

4.2 HCOMP's Gestalt-Shift

Since around 2018, the HCOMP conference has been gradually moving away from its original research topics (see figures 1 and 7). However, no sudden "Gestalt-shift" can be noticed in our analysis of centroid movements for the HCOMP conference (see figures 4 and 9). In these two figures, we would expect to see a large "jump" if a paradigm shift had taken place, yet the year-by-year movement of centroids is evolving gradually. This may be evidence for the field still being in a phase of gradual transition, where some authors clearly switch to different topics, seeking alternatives to the persistent anomalies, while others continue to conduct "normal science."

Nevertheless, the fundamental assumptions of the field are increasingly being questioned. This could indicate that the field has moved from "normal science" into the crisis phase of Kuhn's model, where anomalies and disturbances (e.g., the long-standing issues of quality in crowdsourced work, but also external shocks such as the introduction of large language models and technological advances in generative AI) accumulate, and the fundamental assumptions of the field are upended. While the introduction of large language models has accelerated this for some authors, leading them to explore alternatives in topical areas that have traditionally not received much attention at HCOMP, the HCOMP conference seems to have, on average, not yet entered a revolutionary paradigm shift.

5 Discussion

5.1 Topical Shifts at HCOMP

The HCOMP conference originated from the Human Computation Workshop before evolving into a stand-alone conference in 2013. Since then, the HCOMP conference has gradually evolved and broadened its focus in recent years, to include critical perspectives at the intersection of humans and technology, touching on key

Table 1: Mean similarity between HCOMP (250 articles) and related conferences, by cosine similarity of centroids of article title embeddings. Conferences that are most similar to HCOMP in a given year are highlighted in bold.

Conference	N	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023	2024
CSCW	3081	**0.671**	0.670	0.623	0.660	0.706	**0.743**	0.657	0.748	0.603	0.688	0.637	0.750
FAccT	657	–	–	–	–	–	–	0.714	0.623	0.618	0.791	0.723	**0.805**
CI	220	–*	–*	**0.737**	–*	**0.807**	0.736	0.654	0.611	**0.699**	0.733	**0.972**	0.760
IUI	612	0.566	0.611	0.553	**0.664**	0.617	0.635	**0.771**	**0.752**	0.679	**0.811**	0.744	0.742
UMAP	232	–	–	–	0.631	0.538	0.529	0.668	0.690	0.577	0.768	0.729	0.790
AAMAS	2203	0.620	**0.716**	0.595	0.574	0.601	0.603	0.611	0.546	0.605	0.585	0.585	0.514

* Website no longer available.

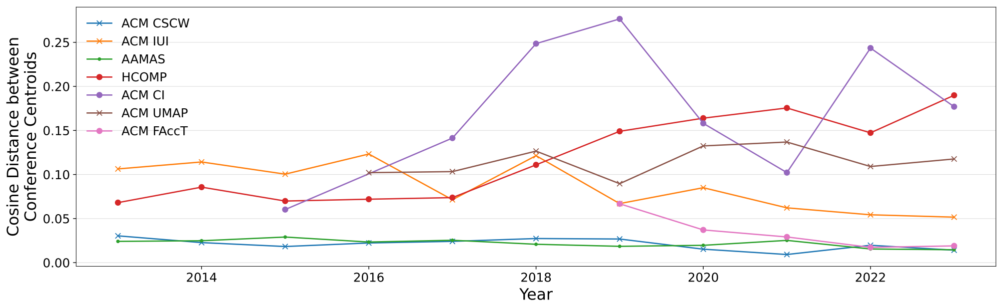


Figure 4: Cosine distances between centroids of article title embeddings in subsequent conference years at the HCOMP Conference (red) and six related conferences. One can note that older venues, such as CSCW and AAMAS, display little centroid movement. This is also evident at ACM FAccT oin recent years. HCOMP and CI, on the other hand, display the largest centroid movement between subsequent years. This could be indicative of shifts in research topics at the conferences.

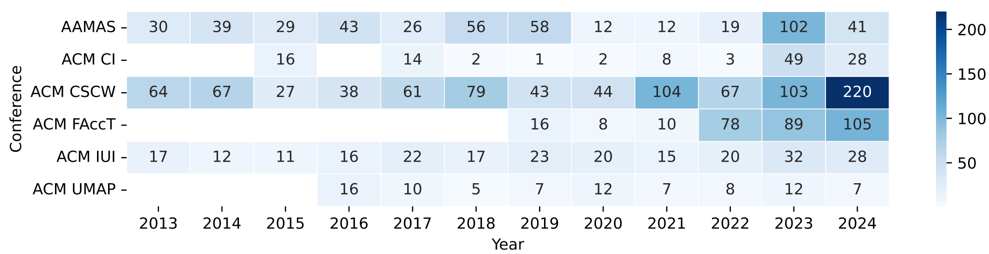


Figure 5: Shared keywords between the HCOMP conference and six related conferences over time (based on article titles and after removal of stopwords)

topics from other conferences, such as ACM FAccT, IUI, and UMAP. This shift in focus is reflected in the types of problems studied and in the terminology and conceptual framings that have become more prominent at the HCOMP conference. While earlier years were focused on studies optimizing crowd workflows and task design, the trend since about 2018 highlights a renewed focus on integrating AI into socio-technical systems and the implications this has for human agency, fairness, and transparency. Notably, research at HCOMP now engages with the design and evaluation of systems that involve humans and AI as collaborative agents, with a new emphasis on trust, explainability & interpretability, and the responsible use of automation. For instance, the theme for the 2024 HCOMP conference was '*Responsible Crowd Work for Better AI*.' The disappearance of traditional co-word pairs, such as task–worker and crowd–worker, in favor of combinations such as human–AI and AI–crowd also reflects this conceptual broadening. Rather than viewing the crowd as a passive labor pool, the field now increasingly investigates humans as active collaborators in systems shaped by algorithmic logic.

This reorientation suggests a departure from purely instrumental framings of human computation toward richer, more nuanced understandings of human–AI configurations. In addition, the overlap with neighboring conferences illustrates a changing scientific landscape at HCOMP. Our findings suggest that HCOMP is undergoing a gradual redefinition of its intellectual boundaries. Rather than abandoning its roots in crowd work and human computation, the community appears to be integrating these origins into a broader agenda that reflects contemporary concerns around AI ethics, collaboration, and human-centered design. In the following section, we discuss whether a paradigm shift has taken place at HCOMP.

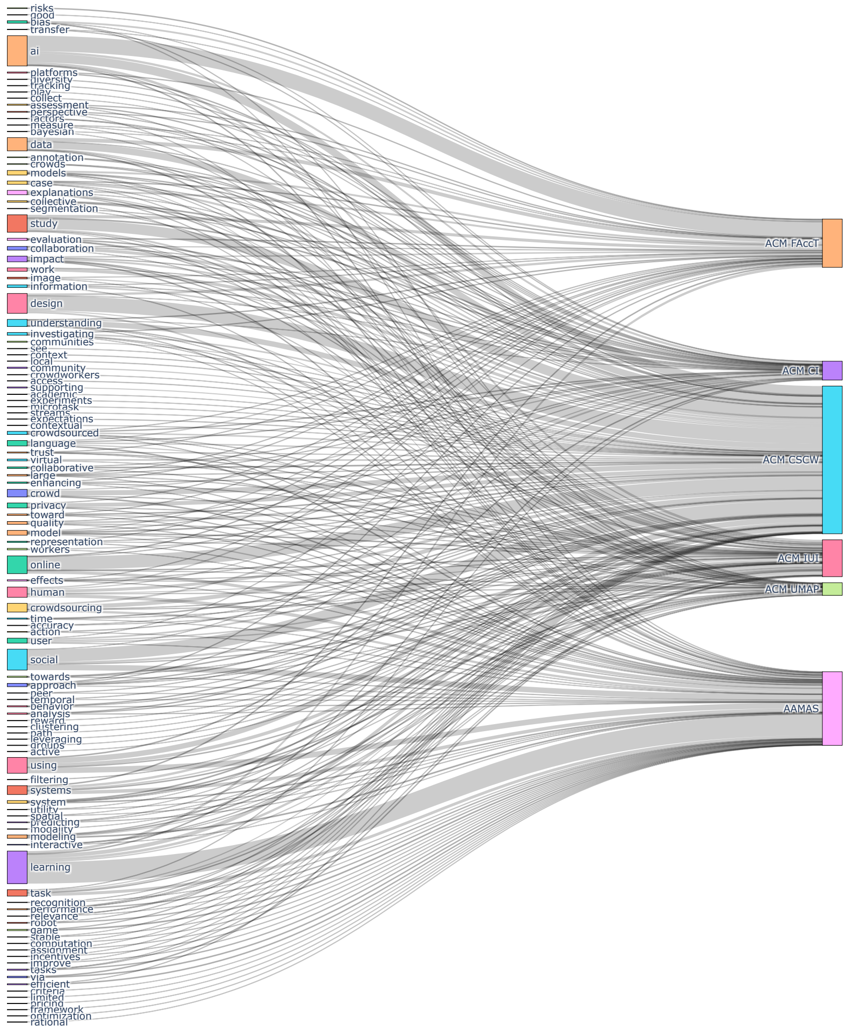
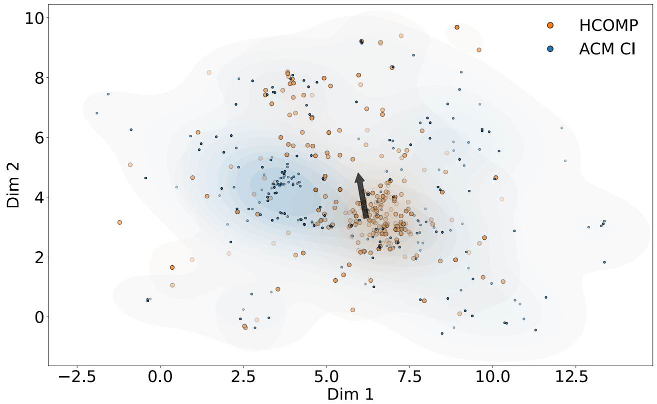


Figure 6: Shared keywords in article titles at HCOMP (left) and six related conferences (ACM CI, ACM FAccT, ACM CSCW, ACM IUI, ACM UMAP, and AAMAS)

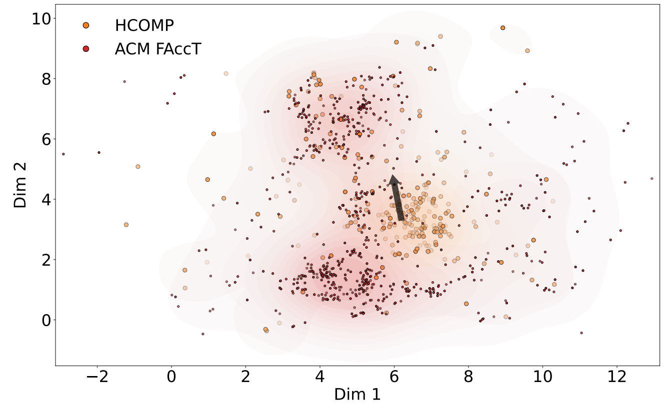
5.2 A Paradigm Shift at HCOMP?

Has there been a revolutionary paradigm shift at HCOMP? In recent years, HCOMP has observed an application and reinvention of methods and concepts to adapt to the new advances that generative AI has brought about. Examples include the application of workflow design—an area well-studied and refined in the crowdsourcing research community—to new human-AI configurations, and the increased focus on conversational agents, hybrid human-AI decision making, and human-AI teaming. If a paradigm shift had indeed already taken place, there would be some "incommensurability" between paradigms (i.e., they could not be directly compared since they use different methods or metrics for evaluation). The adoption of the new paradigm would resemble a "Gestalt-switch" (i.e., a rather sudden perceptual switch in what the community identifies with rather than a gradual one). Our investigation shows that shifts at HCOMP have been gradually occurring since about 2018, and there is likely no incommensurability between HCOMP's period of normal science and its current research. We conclude that recent changes in HCOMP do not (at least yet) constitute a paradigm shift as per Kuhn's model.

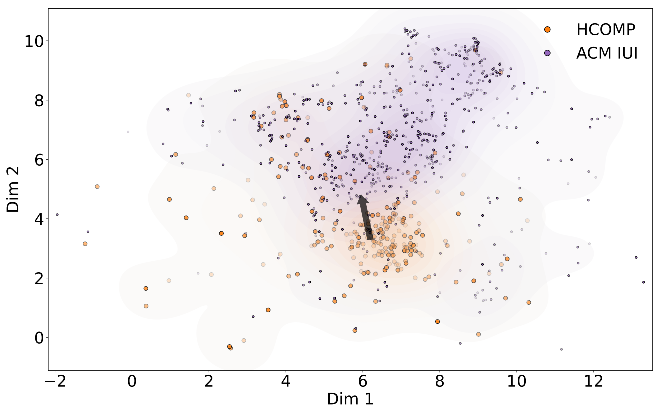
The follow-up question, then, is whether HCOMP is in a phase of crisis. The traditional HCOMP paradigm focused on effectively integrating and optimizing human computation, which provided ample opportunities for the field to make progress during a period of "normal science." However, recent developments in AI can be argued to mean the field of human computation and crowdsourcing has entered the crisis stage in Kuhn's model, in which people have begun to fundamentally question and even undermine the role of "human input" in the age of generative AI. Examples include the generation of data that would traditionally be collected from humans [44, 86, 87, 107] or commentaries on how language models can augment or replace human labor [27, 98]. These recent



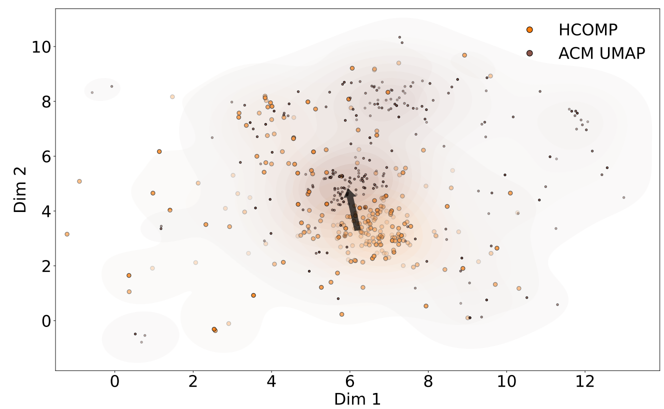
(a) HCOMP and ACM CI



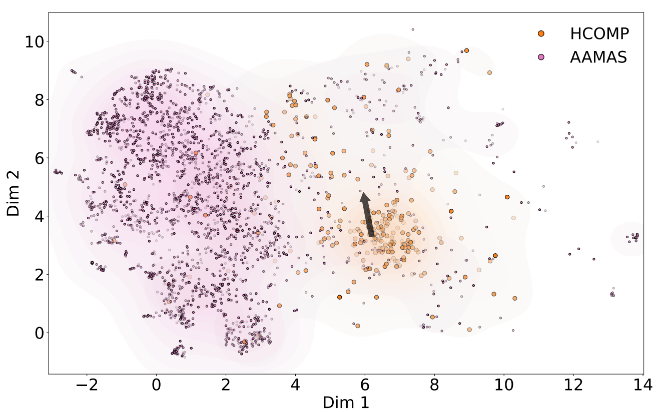
(b) HCOMP and ACM FAccT



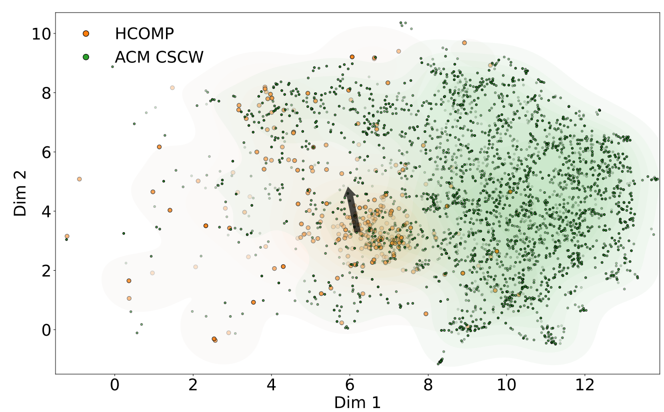
(c) HCOMP and ACM IUI



(d) HCOMP and ACM UMAP



(e) HCOMP and AAMAS



(f) HCOMP and ACM CSCW

Figure 7: The relation of HCOMP to six related conferences. Recent articles are more opaque. The direction of the HCOMP conference from 2013 to 2024, as depicted in Figure 1, is annotated with an arrow.

challenges—accelerated by the disruptive influence of generative AI—indicate that HCOMP is no longer operating within a stable period of "normal science." Instead, the field appears to be entering the crisis phase of Kuhn's model. Core assumptions that have long underpinned human computation—such as the unique value of human-generated data—are being actively questioned or undermined. Researchers have begun to reorient their work toward issues of fairness, interpretability, and human-AI collaboration, and are increasingly publishing research that aligns more closely with the agendas of neighboring conferences, such as ACM FAccT, IUI, and

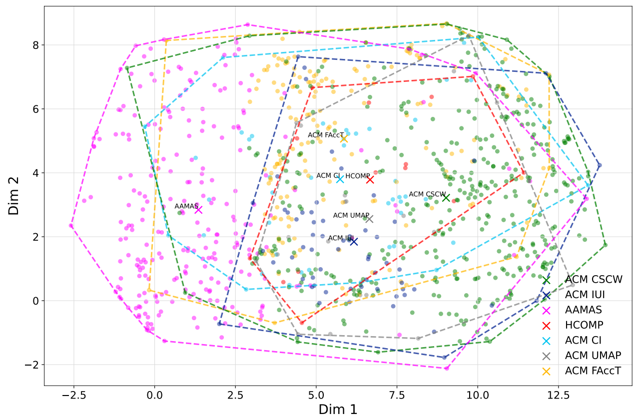


Figure 8: The HCOMP conference's relation to six related conferences in 2024. Dots represent embeddings of article titles published in 2024. Convex hulls enclose all articles of a conference. The conference centroids are marked with X.

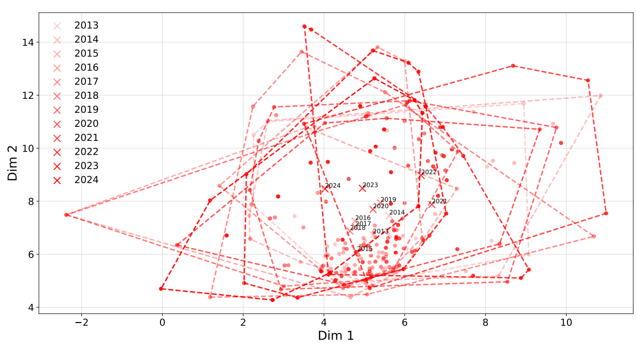


Figure 9: Shifts at the HCOMP conference (2013–2024), as expressed through article title embeddings (represented as dots) and centroids for each year (marked with X). Convex hulls enclose articles in a given year.

UMAP. While a full paradigm shift may not yet have occurred, the evidence suggests that HCOMP is now in the midst of a profound transformation in its epistemic foundations and research priorities. Related to this is the question of whether the community is shedding its old identity in this process. In the following section, we speculate on the future of HCOMP in relation to other fields.

5.3 The Future of HCOMP

We believe that by looking at the past, we can understand the present and critically inform decision-making for the future. We have empirically identified that research at the HCOMP conference has gradually shifted focus since around 2018, with an innovative reorientation from 'workers' to 'humans' taking place since 2021. The research area has evolved, with notable shifts in research topics toward the intersection of AI and humans. The field of HCOMP seems to have moved on from some of its past motor themes and is in a process of reorienting itself in terms of topics studied.

Paid crowdsourcing was enormously important and instrumental to the revolution of artificial intelligence that we bask in today. For instance, early computer vision models relied on crowd workers labeling images and instruction fine-tuning via reinforcement learning from human feedback (RLHF) contributed to the flourishing of large language models. Increasingly, however, companies turn to collect data with alternative, more cost-effective means—often for free—by spinning their own "data flywheels" without the need for outsourcing human labor. Tesla, for instance, collects massive amounts of data from Tesla vehicles in-the-wild. OpenAI uses conversations and user feedback for training future generations of their chatbots. And social media companies—at least until recently—maintained large numbers of in-house content moderators [37, 73]. However, with the emergence of now ubiquitous free large language models and technological advances in automation, the human contribution in crowdsourcing is called into question.

One question during this phase of transition is whether HCOMP should remain its own research field, or merge with another conference. Given the strong interest of researchers in studying large language models—some even advocating for replacing human participants [8, 86, 87, 98, 107]—there is an increasing overlap of topics studied at HCOMP and other conferences. We found empirical evidence that some researchers have moved closer to topics studied at ACM FAccT (Figure 3), and HCOMP—as a whole—has moved closer to conferences such as ACM IUI and UMAP (Figure 7). However, one of the closest conferences, in terms of topic similarity, remains the ACM Collective Intelligence Conference (see Figure 3 and Figure 7). With its broadening topical focus, the HCOMP conference fits well together with CI. This is, to some extent, no surprise, given that the two conferences were co-located and held jointly in the past. We argue there is still a place at HCOMP for research on crowdsourcing. However, as evident in our work, some introspection and reflection on the past is needed to inform HCOMP's future. In the age of generative AI, the purpose of human labor may need to shift from data generation to verification and oversight. Our work contributes data-driven insights to this discussion.

5.4 Limitations and Future Work

We acknowledge a number of limitations to our work. First, with a mean of 20.8 articles published each year, HCOMP is a small conference, and there are only a limited number of data points each year. This affects our analysis, in particular Figure 9. Second, we acknowledge that the field of human computation and crowdsourcing research is much larger than just the HCOMP conference. However, we use the HCOMP conference as a proxy for the wider research field on human computation and crowdsourcing. Future work could extend the analysis to develop a deeper understanding of the impact of recent technological developments on the field of HCOMP. A related limitation is our use of embeddings, which encode semantic meaning of text. There are limitations to interpreting these plots (see [24, 82]), and the complexity of human decision-making in research cannot fully be captured by embedding titles of published articles. Further, in the interpretation of our findings, one needs to consider that HCOMP is a highly specialized venue, rich in generic domain terms, such as 'crowdsourcing' and 'crowd work'. This may have influenced the results. Future work could involve HCOMP researchers in a qualitative investigation to address these limitations. Last, a limitation to our approach is that recent advances in AI can also be used to study existing research topics, by simply replacing existing methods. The Conference on Human Factors in Computing Systems (CHI), for instance, has seen a strong uptake in both studying and using large language models [85]. Such shifts are much harder to measure because in this case, research topics stay the same, and only methods change. Also, a natural drift in topics can be expected, moving fields away from their original topics. This would, however, not constitute a sudden incommensurable paradigm shift.

6 Conclusion

The Conference on Human Computation and Crowdsourcing is at a crossroads. We found that research at the HCOMP conference has gradually shifted away from its traditional motor themes toward artificial intelligence, explainability & interpretability, conversational systems, and human-AI decision-making. This could mean that HCOMP has transitioned—prompted by anomalies brought about by generative AI, challenging and undermining fundamental assumptions in the field—from a period of "normal science" into a new phase. However, we argue this shift cannot be called a revolutionary paradigm shift, according to Kuhn's framework, as of yet. Instead, the field's research focus has gradually broadened to include critical perspectives at the intersection of humans and technology, incorporating topics from other conferences, such as ACM FAccT, IUI, and UMAP. Ultimately, the fate of any given venue hinges on many factors outside the evolution of its topics, for instance funding and community spirit. With our work, we contribute a meaningful and data-informed piece to this broader discussion.

Acknowledgments

The first author gratefully acknowledges funding from the Center for Ubiquitous Computing (UBICOMP) and the University of Oulu's Noste development programme for education. The work was further funded by the Strategic Research Council (SRC), established within the Academy of Finland (grants 356128, 335625, 335729), and Academy Research Fellow funding by Research Council of Finland (formerly Academy of Finland) (grants 356128, 349637 and 353790). This work was also partially supported by the TU Delft AI Initiative, the Convergence Flagship 'ProtectMe', and the Model-driven Decisions Lab (MoDDL).

References

[1] Tahir Abbas and Ujwal Gadiraju. 2022. Goal-setting behavior of workers on crowdsourcing platforms: An exploratory study on MTurk and Prolific. In *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*, Vol. 10. AAAI, Washington, DC, USA, 2–13.

[2] Tahir Abbas, Ujwal Gadiraju, Vassilis-Javed Khan, and Panos Markopoulos. 2021. Making time fly: Using fillers to improve perceived latency in crowd-powered conversational systems. In *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*, Vol. 9. AAAI, Washington, DC, USA, 2–14.

[3] Tahir Abbas, Vassilis-Javed Khan, Ujwal Gadiraju, and Panos Markopoulos. 2020. Trainbot: A conversational interface to train crowd workers for delivering on-demand therapy. In *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*, Vol. 8. AAAI, Washington, DC, USA, 3–12.

[4] Elena Agapie, Jaime Teevan, and Andrés Monroy-Hernández. 2015. Crowdsourcing in the field: A case study using local crowds for event reporting. In *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*, Vol. 3. AAAI, Washington, DC, USA, 2–11. doi:10.1609/hcomp.v3i1.13235

[5] William Agnew, A. Stevie Bergman, Jennifer Chien, Mark Díaz, Seliem El-Sayed, Jaylen Pittman, Shakir Mohamed, and Kevin R. McKee. 2024. The illusion of artificial inclusion. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems (CHI '24)*. ACM, New York, NY, USA, Article 286, 12 pages. doi:10.1145/3613904.3642703

[6] Garrett Allen, Andrea Hu, and Ujwal Gadiraju. 2022. Gesticulate for health's sake! Understanding the use of gestures as an input modality for microtask crowdsourcing. In *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*, Vol. 10. AAAI, Washington, DC, USA, 14–26.

[7] Abdullah Alshaibani, Sylvia Carrell, Li-Hsin Tseng, Jungmin Shin, and Alexander Quinn. 2020. Privacy-preserving face redaction using crowdsourcing. In *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*, Vol. 8. AAAI, Washington, DC, USA, 13–22.

[8] Jacy Reese Anthis, Ryan Liu, Sean M. Richardson, Austin C. Kozlowski, Bernard Koch, James Evans, Erik Brynjolfsson, and Michael Bernstein. 2025. LLM Social Simulations Are a Promising Research Method. arXiv:2504.02234 https://arxiv.org/abs/2504.02234

[9] Yukino Baba and Hisashi Kashima. 2013. Statistical quality estimation for general crowdsourcing tasks. In *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, New York, NY, USA, 554–562.

[10] Hui Bai. 2018. Evidence that a large amount of low quality responses on MTurk can be detected with repeated GPS coordinates. Blog post. https://www.maxhuibai.com/blog/evidence-that-responses-from-repeating-gps-are-random

[11] Gagan Bansal, Besmira Nushi, Ece Kamar, Walter S. Lasecki, Daniel S. Weld, and Eric Horvitz. 2019. Beyond accuracy: The role of mental models in human-AI team performance. In *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*, Vol. 7. AAAI, Washington, DC, USA, 2–11.

[12] Anant Bhardwaj, Juho Kim, Steven Dow, David Karger, Sam Madden, Rob Miller, and Haoqi Zhang. 2014. Attendee-sourcing: Exploring the design space of community-informed conference scheduling. In *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*, Vol. 2. AAAI, Washington, DC, USA, 2–10. doi:10.1609/hcomp.v2i1.13163

[13] Stefano Bianchini, Moritz Müller, and Pierre Pelletier. 2020. Deep learning in science. arXiv:2009.01575 https://arxiv.org/abs/2009.01575

[14] Shreyan Biswas, Ji-Youn Jung, Abhishek Unnam, Kuldeep Yadav, Shreyansh Gupta, and Ujwal Gadiraju. 2024. "Hi. I'm Molly, your virtual interviewer!" Exploring the impact of race and gender in AI-powered virtual interview experiences. In *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*, Vol. 12. AAAI, Washington, DC, USA, 12–22.

[15] Edyta Bogucka, Sanja Šćepanović, and Daniele Quercia. 2024. Atlas of AI risks: Enhancing public understanding of AI risks. In *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*, Vol. 12. AAAI, Washington, DC, USA, 33–43.

[16] Alex Bowyer, Chris Lintott, Greg Hines, Campbell Allan, and Ed Paget. 2015. Panoptes, A project building tool for citizen science. In *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*. AAAI, Washington, DC, USA, 1–2.

[17] Jonathan Bragg, Andrey Kolobov, Mausam Mausam, and Daniel Weld. 2014. Parallel task routing for crowdsourcing. In *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*, Vol. 2. AAAI, Washington, DC, USA, 11–21. doi:10.1609/hcomp.v2i1.13170

[18] Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel Ziegler, Jeffrey Wu, Clemens Winter, Chris Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. 2020. Language models are few-shot learners. In *Advances in Neural Information Processing Systems*, H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin (Eds.), Vol. 33. Curran Associates, Inc., 1877–1901. https://proceedings.neurips.cc/paper_files/paper/2020/file/1457c0d6bfcb4967418bfb8ac142f64a-Paper.pdf

[19] Ricardo J. G. B. Campello, Davoud Moulavi, and Joerg Sander. 2013. Density-based clustering based on hierarchical density estimates. In *Advances in Knowledge Discovery and Data Mining*, Jian Pei, Vincent S. Tseng, Longbing Cao, Hiroshi Motoda, and Guandong Xu (Eds.). Springer, Berlin, Heidelberg, 160–172.

[20] Ruggiero Cavallo and Shaili Jain. 2013. Winner-take-all crowdsourcing contests with stochastic production. In *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*, Vol. 1. AAAI, Washington, DC, USA, 34–41. doi:10.1609/hcomp.v1i1.13090

[21] Alessandro Checco, Jo Bates, and Gianluca Demartini. 2018. All that glitters is gold – An attack scheme on gold questions in crowdsourcing. In *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*, Vol. 6. AAAI, Washington, DC, USA, 2–11. doi:10.1609/hcomp.v6i1.13332

[22] Alessandro Checco, Kevin Roitero, Eddy Maddalena, Stefano Mizzaro, and Gianluca Demartini. 2017. Let's agree to disagree: Fixing agreement measures for crowdsourcing. In *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*, Vol. 5. AAAI, Washington, DC, USA, 11–20. doi:10.1609/hcomp.v5i1.13306

[23] Michael Chmielewski and Sarah C. Kucker. 2020. An MTurk crisis? Shifts in data quality and the impact on study results. *Social Psychological and Personality Science* 11, 4 (2020), 464–473. doi:10.1177/1948550619875149

[24] Andy Coenen and Adam Pearce. n.d.. Understanding UMAP. https://pair-code.github.io/understanding-umap/

[25] Florian Daniel, Pavel Kucherbaev, Cinzia Cappiello, Boualem Benatallah, and Mohammad Allahbakhsh. 2018. Quality control in crowdsourcing: A survey of quality attributes, assessment techniques, and assurance actions. *ACM Comput. Surv.* 51, 1, Article 7 (2018), 40 pages. doi:10.1145/3148148

[26] Anubrata Das, Brandon Dang, and Matthew Lease. 2020. Fast, accurate, and healthier: Interactive blurring helps moderators reduce exposure to harmful content. In *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*, Vol. 8. AAAI, Washington, DC, USA, 33–42.

[27] Chameera De Silva and Thilina Halloluwa. 2025. Augmenting human potential: The role of LLMs in shaping the future of HCI. *Interactions* 32, 2 (2025), 42–45. doi:10.1145/3712068

[28] Gianluca Demartini, Djellel Eddine Difallah, Ujwal Gadiraju, Michele Catasta, et al. 2017. An introduction to hybrid human-machine information systems. *Foundations and Trends in Web Science* 7, 1 (2017), 1–87.

[29] Djellel Difallah, Michele Catasta, Gianluca Demartini, and Philippe Cudré-Mauroux. 2014. Scaling-up the crowd: Micro-task pricing schemes for worker retention and latency improvement. In *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*, Vol. 2. AAAI, Washington, DC, USA, 50–58. doi:10.1609/hcomp.v2i1.13154

[30] Tim Draws, Alisa Rieger, Oana Inel, Ujwal Gadiraju, and Nava Tintarev. 2021. A checklist to combat cognitive biases in crowdsourcing. In *Proceedings of the AAAI conference on human computation and crowdsourcing*, Vol. 9. AAAI, Washington, DC, USA, 48–59.

[31] Emily Dreyfuss. 2018. A bot panic hits Amazon's Mechanical Turk. *Wired* (2018). https://www.wired.com/story/amazon-mechanical-turk-bot-panic/

[32] Anca Dumitrache, Lora Aroyo, and Chris Welty. 2018. Capturing ambiguity in crowdsourcing frame disambiguation. In *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*, Vol. 6. AAAI, Washington, DC, USA, 12–20. doi:10.1609/hcomp.v6i1.13330

[33] Avshalom Elmalech, David Sarne, Esther David, and Chen Hajaj. 2016. Extending workers' attention span through dummy events. In *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*, Vol. 4. AAAI, Washington, DC, USA, 42–51. doi:10.1609/hcomp.v4i1.13276

[34] Alexander Erlei, Franck Nekdem, Lukas Meub, Avishek Anand, and Ujwal Gadiraju. 2020. Impact of algorithmic decision making on human behavior: Evidence from ultimatum bargaining. In *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*, Vol. 8. AAAI, Washington, DC, USA, 43–52.

[35] Guglielmo Faggioli, Laura Dietz, Charles LA Clarke, Gianluca Demartini, Matthias Hagen, Claudia Hauff, Noriko Kando, Evangelos Kanoulas, Martin Potthast, Benno Stein, et al. 2023. Perspectives on large language models for relevance judgment. In *Proceedings of the 2023 ACM SIGIR International Conference on Theory of Information Retrieval*. ACM, New York, NY, USA, 39–50.

[36] Boi Faltings, Radu Jurca, Pearl Pu, and Bao Duy Tran. 2014. Incentives to counter bias in human computation. In *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*, Vol. 2. AAAI, Washington, DC, USA, 59–66. doi:10.1609/hcomp.v2i1.13145

[37] Hibaq Farah. 2023. Diary of a TikTok moderator: 'We are the people who sweep up the mess'. https://www.theguardian.com/technology/2023/dec/21/diary-of-a-tiktok-moderator-we-are-the-people-who-sweep-up-the-mess

[38] Michael Feldman and Abraham Bernstein. 2014. Behavior-based quality assurance in crowdsourcing markets. In *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*, Vol. 2. AAAI, Washington, DC, USA, 14–15.

[39] Manuel Gentile, Giuseppe Città, Salvatore Perna, and Mario Allegra. 2023. Do we still need teachers? Navigating the paradigm shift of the teacher's role in the AI era. *Frontiers in Education* 8 (2023). doi:10.3389/feduc.2023.1161777

[40] Karan Goel, Shreya Rajpal, and Mausam Mausam. 2017. Octopus: A framework for cost-quality-time optimization in crowdsourcing. In *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*, Vol. 5. AAAI, Washington, DC, USA, 31–40. doi:10.1609/hcomp.v5i1.13311

[41] Shinsuke Goto, Toru Ishida, and Donghui Lin. 2016. Understanding crowdsourcing workflow: Modeling and optimizing iterative and parallel processes. In *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*, Vol. 4. AAAI, Washington, DC, USA, 52–58. doi:10.1609/hcomp.v4i1.13289

[42] Sai Gouravajhala, Jinyeong Yim, Karthik Desingh, Yanda Huang, Odest Jenkins, and Walter Lasecki. 2018. EURECA: Enhanced understanding of real environments via crowd assistance. In *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*, Vol. 6. AAAI, Washington, DC, USA, 31–40. doi:10.1609/hcomp.v6i1.13339

[43] Nina Grgić-Hlača, Claude Castelluccia, and Krishna P Gummadi. 2022. Taking advice from (dis)similar machines: The impact of human-machine similarity on machine-assisted decision-making. In *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*, Vol. 10. AAAI, Washington, DC, USA, 74–88.

[44] Perttu Hämäläinen, Mikke Tavast, and Anton Kunnari. 2023. Evaluating large language models in generating synthetic HCI research data: A case study. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23)*. ACM, New York, NY, USA, Article 433, 19 pages. doi:10.1145/3544548.3580688

[45] Peter Hase, Chaofan Chen, Oscar Li, and Cynthia Rudin. 2019. Interpretable image recognition with hierarchical prototypes. In *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*, Vol. 7. AAAI, Washington, DC, USA, 32–40. doi:10.1609/hcomp.v7i1.5265

[46] David J. Hauser and Norbert Schwarz. 2016. Attentive turkers: MTurk participants perform better on online attention checks than do subject pool participants. *Behavior Research Methods* 48, 1 (2016), 400–407. doi:10.3758/s13428-015-0578-z

[47] Gaole He, Agathe Balayn, Stefan Buijsman, Jie Yang, and Ujwal Gadiraju. 2022. It is like finding a polar bear in the savannah! Concept-level ai explanations with analogical inference from commonsense knowledge. In *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*, Vol. 10. AAAI, Washington, DC, USA, 89–101.

[48] Danula Hettiachchi, Senuri Wijenayake, Simo Hosio, Vassilis Kostakos, and Jorge Goncalves. 2020. How context influences cross-device task acceptance in crowd work. In *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*, Vol. 8. AAAI, Washington, DC, USA, 53–62.

[49] Emőke-Ágnes Horvát, Johannes Wachs, Rong Wang, and Anikó Hannák. 2018. The role of novelty in securing investors for equity crowdfunding campaigns. In *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*, Vol. 6. AAAI, Washington, DC, USA, 50–59. doi:10.1609/hcomp.v6i1.13336

[50] Jeff Howe. 2006. The rise of crowdsourcing. *Wired* (2006).

[51] Charlene Jennett and Anna Cox. 2014. Eight guidelines for designing virtual citizen science projects. In *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*, Vol. 2. AAAI, Washington, DC, USA, 16–17.

[52] Hyun Jung, Yubin Park, and Matthew Lease. 2014. Predicting next label quality: A time-series model of crowdwork. In *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*, Vol. 2. AAAI, Washington, DC, USA, 87–95. doi:10.1609/hcomp.v2i1.13165

[53] Toni Kaplan, Susumu Saito, Kotaro Hara, and Jeffrey Bigham. 2018. Striving to earn more: A survey of work strategies and tool use among crowd workers. In *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*, Vol. 6. AAAI, Washington, DC, USA, 70–78.

[54] Ashiqur KhudaBukhsh, Jaime Carbonell, and Peter Jansen. 2014. Detecting non-adversarial collusion in crowdsourcing. In *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*, Vol. 2. AAAI, Washington, DC, USA, 104–111. doi:10.1609/hcomp.v2i1.13157

[55] Aniket Kittur, Ed H. Chi, and Bongwon Suh. 2008. Crowdsourcing user studies with Mechanical Turk. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '08)*. ACM, New York, NY, USA, 453–456. doi:10.1145/1357054.1357127

[56] Aniket Kittur, Jeffrey V Nickerson, Michael Bernstein, Elizabeth Gerber, Aaron Shaw, John Zimmerman, Matt Lease, and John Horton. 2013. The future of crowd work. In *Proceedings of the 2013 Conference on Computer Supported Cooperative Work*. ACM, New York, NY, USA, 1301–1318.

[57] Masaki Kobayashi, Kei Wakabayashi, and Atsuyuki Morishima. 2021. Human+AI crowd task assignment considering result quality requirements. In *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*, Vol. 9. AAAI, Washington, DC, USA, 97–107.

[58] Andreas Kriegler. 2022. Paradigmatic revolutions in computer vision. In *I can't believe it's not better workshop: Understanding deep learning through empirical falsification*. https://openreview.net/forum?id=wMKHI8g0OVe

[59] Evgeny Krivosheev, Fabio Casati, Valentina Caforio, and Boualem Benatallah. 2017. Crowdsourcing paper screening in systematic literature reviews. In *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*, Vol. 5. AAAI, Washington, DC, USA, 108–117. doi:10.1609/hcomp.v5i1.13302

[60] Thomas S. Kuhn. 1997. *The structure of scientific revolutions*. Vol. 962. University of Chicago Press, Chicago.

[61] Katsumi Kumai, Masaki Matsubara, Yuhki Shiraishi, Daisuke Wakatsuki, Jianwei Zhang, Takeaki Shionome, Hiroyuki Kitagawa, and Atsuyuki Morishima. 2018. Skill-and-stress-aware assignment of crowd-worker groups to task streams. In *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*, Vol. 6. AAAI, Washington, DC, USA, 88–97. doi:10.1609/hcomp.v6i1.13328

[62] Isaac Lage, Emily Chen, Jeffrey He, Menaka Narayanan, Been Kim, Samuel J. Gershman, and Finale Doshi-Velez. 2019. Human evaluation of models built for interpretability. In *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*, Vol. 7. AAAI, Washington, DC, USA, 59–67. doi:10.1609/

hcomp.v7i1.5280

[63] Edith Law and Luis Von Ahn. 2011. *Human computation.* Morgan & Claypool Publishers.

[64] Q Vera Liao, Yunfeng Zhang, Ronny Luss, Finale Doshi-Velez, and Amit Dhurandhar. 2022. Connecting algorithmic research and usage contexts: A perspective of contextualized evaluation for explainable AI. In *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*, Vol. 10. AAAI, Washington, DC, USA, 147–159.

[65] Jude Lim, Vikram Mohanty, Terryl Dodson, and Kurt Luther. 2023. BackTrace: A human-AI collaborative approach to discovering studio backdrops in historical photographs. In *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*, Vol. 11. AAAI, Washington, DC, USA, 91–102.

[66] Leib Litman. 2018. Moving beyond bots: MTurk as a source of high quality data. Blog post. https://www.cloudresearch.com/resources/blog/moving-beyond-bots-mturk-as-a-source-of-high-quality-data/

[67] Pengfei Liu, Weizhe Yuan, Jinlan Fu, Zhengbao Jiang, Hiroaki Hayashi, and Graham Neubig. 2023. Pre-train, prompt, and predict: A systematic survey of prompting methods in natural language processing. *ACM Comput. Surv.* 55, 9, Article 195 (2023), 35 pages. doi:10.1145/3560815

[68] Zhuoran Lu, Syed Hasan Amin Mahmoo, Zhuoyan Li, and Ming Yin. 2024. Mix and match: Characterizing heterogeneous human behavior in AI-assisted decision making. In *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*, Vol. 12. AAAI, Washington, DC, USA, 95–104.

[69] Kurt Luther, Nathan Hahn, Steven Dow, and Aniket Kittur. 2015. Crowdlines: Supporting synthesis of diverse information sources through crowdsourced outlines. In *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*, Vol. 3. AAAI, Washington, DC, USA, 110–119. doi:10.1609/hcomp.v3i1.13239

[70] Jazmine Maldonado Flores, Jheser Guzman, and Barbara Poblete. 2017. A lightweight and real-time worldwide earthquake detection and monitoring system based on citizen sensors. 5 (2017), 137–146. doi:10.1609/hcomp.v5i1.13303

[71] V. K. Manam and Alexander Quinn. 2018. WingIt: Efficient refinement of unclear task instructions. In *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*, Vol. 6. AAAI, Washington, DC, USA, 108–116. doi:10.1609/hcomp.v6i1.13338

[72] Leland McInnes, John Healy, and James Melville. 2018. UMAP: Uniform manifold approximation and projection for dimension reduction. arXiv:1802.03426

[73] Liv McMahon, Zoe Kleinman, and Courtney Subramanian. 2025. Facebook and Instagram get rid of fact checkers. https://www.bbc.com/news/articles/cly74mpy8klo

[74] David Alvarez Melis, Harmanpreet Kaur, Hal Daumé III, Hanna Wallach, and Jennifer Wortman Vaughan. 2021. From human explanation to model interpretability: A framework based on weight of evidence. In *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*, Vol. 9. AAAI, Washington, DC, USA, 35–47.

[75] Meredith Ringel Morris. 2025. HCI for AGI. *Interactions* 32, 2 (2025), 26–32. doi:10.1145/3708815

[76] Aaron Moss. 2018. After the bot scare: Understanding what's been happening with data collection on MTurk and how to stop it. Blog post. https://www.cloudresearch.com/resources/blog/after-the-bot-scare-understanding-whats-been-happening-with-data-collection-on-mturk-and-how-to-stop-it/

[77] Daniel Mutembesa, Christopher Omongo, and Ernest Mwebaze. 2018. Crowdsourcing real-time viral disease and pest information: A case of nation-wide cassava disease surveillance in a developing country. In *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*, Vol. 6. AAAI, Washington, DC, USA, 117–125. doi:10.1609/hcomp.v6i1.13322

[78] Mahsan Nourani, Samia Kabir, Sina Mohseni, and Eric D Ragan. 2019. The effects of meaningful and meaningless explanations on trust and perceived system accuracy in intelligent systems. In *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*, Vol. 7. AAAI, Washington, DC, USA, 97–105.

[79] Besmira Nushi, Ece Kamar, and Eric Horvitz. 2018. Towards accountable AI: Hybrid human-machine analyses for characterizing system failure. In *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*, Vol. 6. AAAI, Washington, DC, USA, 126–135. doi:10.1609/hcomp.v6i1.13337

[80] David Oleson, Alexander Sorokin, Greg P Laughlin, Vaughn Hester, John Le, and Lukas Biewald. 2011. Programmatic gold: Targeted and scalable quality assurance in crowdsourcing. *Human computation* 11, 11 (2011).

[81] Daniel M. Oppenheimer, Tom Meyvis, and Nicolas Davidenko. 2009. Instructional manipulation checks: Detecting satisficing to increase statistical power. *Journal of Experimental Social Psychology* 45, 4 (2009), 867–872. doi:10.1016/j.jesp.2009.03.009

[82] Jonas Oppenlaender and Joonas Hämäläinen. 2023. Mapping the challenges of HCI: An application and evaluation of ChatGPT for mining insights at scale. doi:10.48550/arXiv.2306.05036

[83] Jahna Otterbacher. 2018. Social cues, social biases: Stereotypes in annotations on people images. In *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*, Vol. 6. AAAI, Washington, DC, USA, 136–144. doi:10.1609/hcomp.v6i1.13320

[84] Jahna Otterbacher, Pınar Barlas, Styliani Kleanthous, and Kyriakos Kyriakou. 2019. How do we talk about other people? Group (un)fairness in natural language image descriptions. In *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*, Vol. 7. AAAI, Washington, DC, USA, 106–114.

[85] Rock Yuren Pang, Hope Schroeder, Kynnedy Simone Smith, Solon Barocas, Ziang Xiao, Emily Tseng, and Danielle Bragg. 2025. Understanding the LLM-ification of CHI: Unpacking the impact of LLMs at CHI through a systematic literature review. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*. ACM, New York, NY, USA. arXiv:2501.12557

[86] Joon Sung Park, Joseph O'Brien, Carrie Jun Cai, Meredith Ringel Morris, Percy Liang, and Michael S. Bernstein. 2023. Generative agents: Interactive simulacra of human behavior. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology (UIST '23)*. ACM, New York, NY, USA, Article 2, 22 pages. doi:10.1145/3586183.3606763

[87] Joon Sung Park, Carolyn Q. Zou, Aaron Shaw, Benjamin Mako Hill, Carrie Cai, Meredith Ringel Morris, Robb Willer, Percy Liang, and Michael S. Bernstein. 2024. Generative agent simulations of 1,000 people. arXiv:2411.10109 https://arxiv.org/abs/2411.10109

[88] Charvi Rastogi, Liu Leqi, Kenneth Holstein, and Hoda Heidari. 2023. A taxonomy of human and ML strengths in decision-making to investigate human-ML complementarity. In *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*, Vol. 11. AAAI, Washington, DC, USA, 127–139.

[89] Arijit Ray, Yi Yao, Rakesh Kumar, Ajay Divakaran, and Giedrius Burachas. 2019. Can you explain that? Lucid explanations help human-AI collaborative image retrieval. In *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*, Vol. 7. AAAI, Washington, DC, USA, 153–161.

[90] Amy Rechkemmer and Ming Yin. 2020. Motivating novice crowd workers through goal setting: An investigation into the effects on complex crowdsourcing task training. In *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*, Vol. 8. AAAI, Washington, DC, USA, 122–131.

[91] Nils Reimers and Iryna Gurevych. 2019. Sentence-BERT: Sentence embeddings using siamese BERT-networks. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, Kentaro Inui, Jing Jiang, Vincent Ng, and Xiaojun Wan (Eds.). Association for Computational Linguistics, 3982–3992. doi:10.18653/v1/D19-1410

[92] Geoffrey Rockwell and Stéfan Sinclair. 2016. *Hermeneutica: Computer-assisted interpretation in the humanities*. MIT Press. doi:10.7551/mitpress/9522.001.0001

[93] Timothy J. Ryan. 2018. Data contamination on MTurk. Blog post. https://timryan.web.unc.edu/2018/08/12/data-contamination-on-mturk/

[94] Elliot Salisbury, Sebastian Stein, and Sarvapali Ramchurn. 2015. CrowdAR: Augmenting live video with a real-time crowd. In *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*, Vol. 3. AAAI, Washington, DC, USA, 169–177. doi:10.1609/hcomp.v3i1.13220

[95] Ioannis Petros Samiotis, Sihang Qiu, Christoph Lofi, Jie Yang, Ujwal Gadiraju, and Alessandro Bozzon. 2021. Exploring the music perception skills of crowd workers. In *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*, Vol. 9. AAAI, Washington, DC, USA, 108–119.

[96] David Sarne and Michael Lepioshkin. 2017. Effective prize structure for simple crowdsourcing contests with participation costs. In *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*. AAAI, Washington, DC, USA, 167–176. doi:10.1609/hcomp.v5i1.13305

[97] Denis Savenkov and Eugene Agichtein. 2016. CRQA: Crowd-powered real-time automatic question answering system. In *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*, Vol. 4. AAAI, Washington, DC, USA, 189–198. doi:10.1609/hcomp.v4i1.13291

[98] Albrecht Schmidt, Passant Elagroudy, Fiona Draxler, Frauke Kreuter, and Robin Welsch. 2024. Simulating the human in HCD with ChatGPT: Redesigning interaction design with AI. *Interactions* 31, 1 (2024), 24–31.

[99] Aashish Sheshadri and Matthew Lease. 2013. Square: A benchmark for research on computing crowd consensus. In *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*, Vol. 1. AAAI, Washington, DC, USA, 156–164.

[100] Akash Singh, Faizy Ahsan, Mathieu Blanchette, and Jérôme Waldispühl. 2017. Lessons from an online massive genomics computer game. In *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*, Vol. 5. AAAI, Washington, DC, USA, 177–186. doi:10.1609/hcomp.v5i1.13309

[101] Aayush Singh, Sebastian Wehkamp, and Ujwal Gadiraju. 2022. SignUpCrowd: Using sign-language as an input modality for microtask crowdsourcing. In *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*, Vol. 10. AAAI, Washington, DC, USA, 184–194.

[102] C. Stokel-Walker. 2018. Bots on Amazon's Mechanical Turk are ruining psychology studies. *New Scientist* (2018). https://www.newscientist.com/article/2176436-bots-on-amazons-mechanical-turk-are-ruining-psychology-studies/

[103] Veniamin Veselovsky, Manoel Horta Ribeiro, and Robert West. 2023. Artificial artificial artificial intelligence: Crowd workers widely use large language models for text production tasks. arXiv:2306.07899

[104] Luis von Ahn. 2005. *Human computation*. Ph. D. Dissertation. Carnegie Mellon University.

[105] Luis von Ahn and Laura Dabbish. 2004. Labeling images with a computer game. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '04)*. ACM, New York, NY, USA, 319–326. doi:10.1145/985692.985733

[106] Bo Waggoner and Yiling Chen. 2014. Output agreement mechanisms and common knowledge. In *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*, Vol. 2. AAAI, Washington, DC, USA, 220–226. doi:10.1609/hcomp.v2i1.13151

[107] Zhilin Wang, Yu Ying Chiu, and Yu Cheung Chiu. 2023. Humanoid agents: Platform for simulating human-like generative agents. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, Yansong Feng and Els Lefever (Eds.). Association for Computational Linguistics, Singapore, 167–176. doi:10.18653/v1/2023.emnlp-demo.15

[108] Mark E Whiting, Grant Hugh, and Michael S Bernstein. 2019. Fair work: Crowd work minimum wage with one line of code. In *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*, Vol. 7. AAAI, Washington, DC, USA, 197–206.

[109] Andrea Wiggins, Carl Lagoze, Weng-Keen Wong, and Steve Kelling. 2014. A sensor network approach to managing data quality in citizen science. In *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*, Vol. 2. AAAI, Washington, DC, USA, 35–36.

[110] Meng-Han Wu and Alexander Quinn. 2017. Confusing the crowd: Task instruction quality on Amazon Mechanical Turk. In *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*, Vol. 5. AAAI, Washington, DC, USA, 206–215. doi:10.1609/hcomp.v5i1.13317

[111] Tongshuang Wu, Haiyi Zhu, Maya Albayrak, Alexis Axon, Amanda Bertsch, Wenxing Deng, Ziqi Ding, Bill Guo, Sireesh Gururaja, Tzu-Sheng Kuo, et al. 2023. LLMs as workers in human-computational algorithms? Replicating crowdsourcing pipelines with LLMs. arXiv:2307.10168

[112] Yexiang Xue, Bistra Dilkina, Theodoros Damoulas, Daniel Fink, Carla Gomes, and Steve Kelling. 2013. Improving your chances: Boosting citizen science discovery. In *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*, Vol. 1. Washington, DC, USA, 198–206. doi:10.1609/hcomp.v1i1.13070

[113] Yaniv Yacoby, Ben Green, Christopher L Griffin Jr, and Finale Doshi-Velez. 2022. "If it didn't happen, why would I change my decision?": How judges respond to counterfactual explanations for the public safety assessment. In *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*, Vol. 10. AAAI, Washington, DC, USA, 219–230.

[114] Jie Yang, Carlo van der Valk, Tobias Hoßfeld, Judith Redi, and Alessandro Bozzon. 2018. How do crowdworker communities and microtask markets influence each other? A data-driven study on Amazon Mechanical Turk. In *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*, Vol. 6. AAAI, Washington, DC, USA, 193–202. doi:10.1609/hcomp.v6i1.13335

[115] Ming Yin and Yiling Chen. 2016. Predicting crowd work quality under monetary interventions. In *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*, Vol. 4. AAAI, Washington, DC, USA, 259–268. doi:10.1609/hcomp.v4i1.13282

[116] Ming Yin, Yiling Chen, and Yu-An Sun. 2014. Monetary interventions in crowdsourcing task switching. In *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*, Vol. 2. AAAI, Washington, DC, USA, 234–241. doi:10.1609/hcomp.v2i1.13160

[117] Khobaib Zaamout and Ken Barker. 2018. Towards quantifying behaviour in social crowdsourcing communities. In *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*, Vol. 6. AAAI, Washington, DC, USA, 203–212. doi:10.1609/hcomp.v6i1.13323

[118] Alexandros Zenonos, Sebastian Stein, and Nicholas Jennings. 2017. A trust-based coordination system for participatory sensing applications. In *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*, Vol. 5. AAAI, Washington, DC, USA, 226–234. doi:10.1609/hcomp.v5i1.13297

[119] Ruohan Zong, Yang Zhang, Frank Stinar, Lanyu Shang, Huimin Zeng, Nigel Bosch, and Dong Wang. 2023. A crowd–AI collaborative approach to address demographic bias for student performance prediction in online education. In *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*, Vol. 11. AAAI, Washington, DC, USA, 198–210.