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

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

Ujwal Gadiraju
Delft University of Technology
Delft, Netherlands
u.k.gadiraju@tudelft.nl

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

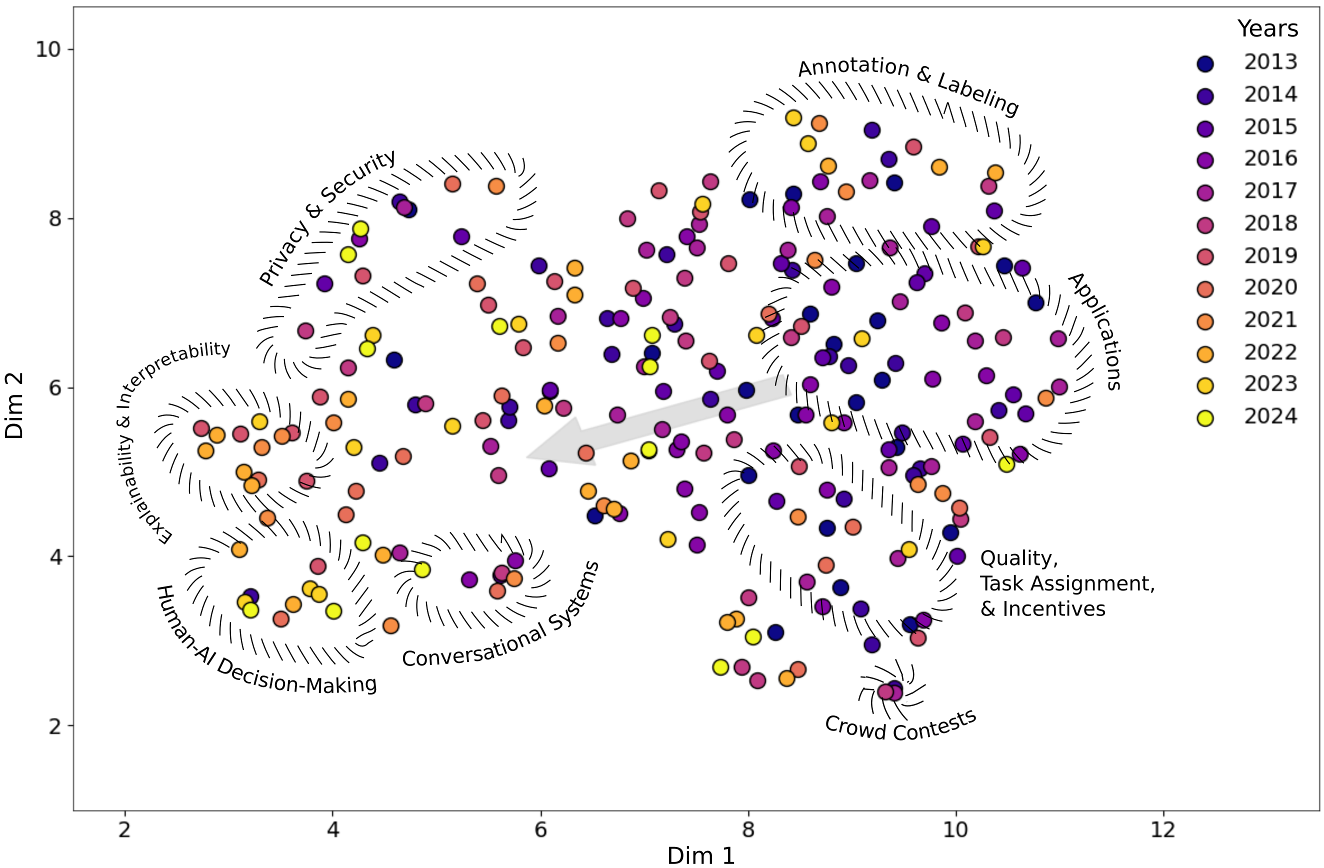

Figure 1: Research topics in articles published at the Conference on Human Computation and Crowdsourcing (HCOMP) between 2013 and 2024. The dots represent article titles embedded using sentence transformers and projected into two-dimensional space with a dimensionality reduction technique (UMAP). The 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 979-8-4007-1489-4/25/08
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 04–06, 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

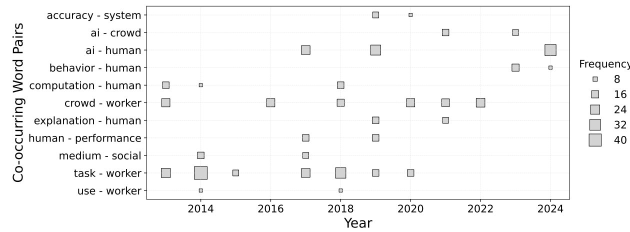

Figure 2: Co-word occurrences at the HCOMP conference over time (order-insensitive, considering only co-words that appear in more than one year)

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.

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