

Opinion

Crowdsourcing or AI Sourcing?

Considering the impact of generative artificial intelligence on data annotation tasks.

THE NEED FOR human-generated labels to train machine learning (ML) models is a key driver behind the rise of crowdsourcing⁵: outsourcing simple data annotation tasks to non-experts, who are available 24/7 and are paid per task, via dedicated platforms. During the last 15 years, much research on crowdsourcing and human computation has been conducted, including methods to control for quality, cost, and speed of data labeling. Large corporations make use of crowdsourcing for tasks such as image annotation, relevance judgments, as well as content moderation. After many years of crowdsourcing being used in production, researchers envisioned a future of sustainable crowd work with long-term opportunities and managerial roles within hierarchical structures.⁷ In hindsight, crowd work has not changed much over the years, and has increasingly received criticism as being an exploitative process, in which the labor behind modern ML-driven AI applications is hidden.¹

Crowd work has always been technology-supported, with professional workers making use of scripts and tools to improve productivity.⁶ Now, with the wide availability of generative artificial intelligence (GenAI) tools such as large language models (LLMs), we ask a crucial question: How will workers use them, and will crowd work be made obsolete by GenAI? Will this disruptive technology



eliminate the need for crowd work or will it fundamentally reshape its nature? Considering, in particular, the case of data annotation, we take the stance that the incorporation of GenAI into crowd work is unavoidable, and that platforms, as well as requesters, must take steps to promote its responsible use. We explore data annotation specifically because it is the very process through which GenAI models have been built and continue to learn. As will be discussed, while both requesters and workers can benefit from the use of GenAI for annotation, unintended consequences (for example, la-

bels that are too homogeneous or low quality) must always be anticipated and treated with care.

Can GenAI Replace All Crowdsourced Annotations?

Internet corporations currently employ armies of annotators. One experimental study suggested that a large proportion of workers on Amazon Mechanical Turk (MTurk) are leveraging ChatGPT in text summarization tasks.¹⁰ Furthermore, MTurkers self-reported using GenAI in tasks.² While this is not surprising, it poses questions as to whether crowdsourc-

ing is still needed. Perhaps GenAI will replace humans in tasks such as moderating online content (especially when potentially harmful to people) and creating labeled data to train ML models, resulting in widespread *AI-Sourcing*.

Current LLMs work well for tasks such as text summarization, but it is not yet clear if they are ready to replace humans for more subjective tasks (for example, sentiment classification, hate speech detection, or image labeling).⁴ With respect to using ChatGPT broadly in annotation tasks, Kocon and colleagues described it as a “Jack of all trades ... but master of none.”⁸ Evaluating it on 25 tasks, they concluded that it performs decently overall, but can be hit-or-miss on particular tasks relative to bespoke solutions. Beyond the discussion of LLM capabilities, it is important to explore scenarios in which humans work together with GenAI to complete tasks more effectively. But then, requesters may be unaware of annotators using these tools as a co-pilot. Do they need to know? Should there be an option to collect *natively* human labels rather than hybrid human-machine annotations? Or, is what matters the quality of the resulting labels or of the predictions made by the models trained with such labels? The research strongly suggests a need for humans in the loop; however, the nature of human computation and crowdsourcing has fundamentally changed.

Banning or Enabling the Use of GenAI in Crowdsourcing?

Early stances toward workers using GenAI have not been aligned. We have noticed MTurk workers have started to use ChatGPT.² On the other hand, Prolific does not allow workers to use LLMs, informing workers: “Please don’t use AI

Beyond the discussion of LLM capabilities, it is important to explore scenarios in which humans work together with GenAI to complete tasks more effectively.

assistance tools or large language models (LLMs) when completing a study.”^a In the end, the market may decide on the “best approach,” with requesters choosing the platform that suits their needs. However, when we have diverse approaches to the management of GenAI in crowd work, we risk creating a digital divide. Workers who are not allowed to use it may, over time, become disadvantaged. Also, even if a platform bans GenAI, enforcing this is likely to prove futile. Automatic detection of AI-generated work is an “arms race,” as the technology is improving rapidly.

To date, platforms have followed various approaches to managing GenAI tools. However, ultimately, embracing human-AI collaboration in crowd work is the only solution. Rather than a future in which annotation tasks will be taken over by GenAI, we envision a future collaborative environment where workers may be more empowered and efficient; however, this may signify a shift in

the nature of the tasks and the effort required. We can consider a spectrum in between the extremes of fully AI-driven and fully human annotation tasks. For instance, text summarization tasks might be completed by GenAI alone, common-sense knowledge assessment tasks could be done collaboratively, while the validation of AI annotations should be completed by humans.

If this collaborative approach is widely adopted, it is then critical to provide training to humans on the risks and opportunities of using these tools as a co-pilot. Increasing literacy on the impact of GenAI on human decisions, beyond in-depth knowledge of annotation guidelines, will be essential. It is not obvious what approach should be taken. Effort toward increasing GenAI literacy in the community of human annotators, as well as data requesters, is needed.

Embracing Human-AI Collaboration

We anticipate humans will move toward the role of *auditors* of AI-generated annotations, without this signifying the end of human-oriented tasks. On the contrary, we envision a human-AI collaborative spectrum³ that encompasses both AI-supported human annotations (HA-AI) and human-audited AI annotations (AIA-HV). Figure 1 outlines the potential forms of collaboration.

Humans can assume the role of annotators, providing pure human annotations (HA). As GenAI advances, humans will shift to auditors of AI-generated annotations, verifying labels and/or validating their correctness. The final step in this evolution is for humans to become observers, shifting to pure AI annotations (AIA), and even AI validations/verifications (AIV). Cur-

^a See <https://bit.ly/4k62Cd8>

Figure 1. Left: Abbreviations used to describe tasks (columns) and entities performing the task (rows). Directional arrows convey possible combinations of the entity in charge of the task, as explained in the figure on the right. Right: The possible roles of human crowd workers as GenAI advances.

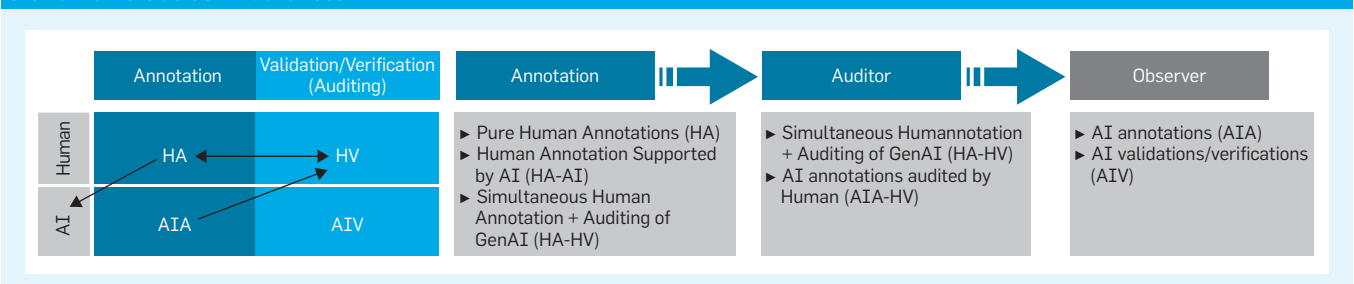


Figure 2. A discussion on the possible shift and/or intensification of the human effort (see arrows), either toward annotation or auditing on various examples of data annotation tasks.

Task	Fluctuation in the type of human effort as GenAI advances		Discussion
	AI-Supported Human Annotation	Human-Audited GenAI Annotations	
Content Creation	↘	↗	<ul style="list-style-type: none">▶ Human annotation effort decreases for tasks previously requiring HA.▶ Tasks that required AIA-HV, we envision an increase in the human auditing effort with the introduction of GenAI.
Sentiment detection/ classification	↗	↗→	<ul style="list-style-type: none">▶ We envision a shift to tasks that require HA-HV.▶ Tasks needing AIA-HV, we foresee continued intense human auditing efforts, regardless of GenAI reliance.
Verification and Validation	—	↗	<ul style="list-style-type: none">▶ We envision a relative increase in the human auditing effort involving AIA-HV.
Information Finding	↘	↗	<ul style="list-style-type: none">▶ Human annotation (HA) may decrease, transitioning to intense auditing, eliminating the need for direct human annotation, thus moving towards AIA-HV.
Image labeling	↗	↗	<ul style="list-style-type: none">▶ Given the level of dependence from GenAI we envision image labeling tasks to either need an intense human effort in HA-AI and AIA-HV tasks as well as increase HA-HV tasks.

rently, workers are far from reaching the role of observers, but do experience different forms of collaboration with GenAI. They can receive support during annotation from AI tools (HA-AI) (for example, a shorter list of possible tags for an image annotation task) or they might act as both annotators and auditors during the same task (HA-HV) (for example, performing sentiment analysis on a paragraph and verifying an AI-generated sentiment analysis for the same paragraph) or, finally, they might move to a role of verifying/validating AI annotations (AIA-HV) as pure auditors (for example, validating the correctness of AI labels). This collaborative spectrum requires two types of effort from workers: producing annotations or auditing AI annotations (see Figure 2). We consider the roles of validator and verifier similar, different from the current role of annotator, and not yet at the stage of observer. For this reason, we discuss these two roles together, under the general role of auditor.

Figure 2 analyzes how the balance between human and AI effort may evolve over time, showing several types of annotation tasks. In particular, the human effort dedicated to either (original) annotation or auditing when facilitated by GenAI on common tasks is considered. We envision a transition toward GenAI as a co-pilot for humans and then, a gradual shift of the role of the human toward having less original work to complete, but more responsibility as auditors.

Tasks such as sentiment classification or image annotation traditionally

require high human effort. For these tasks, it is possible that humans are required to both create original annotations, as well as to audit samples of AI-generated annotations (HA-HV). Humans may need to audit low-confidence annotations, or intervene on potentially biased AI outcomes, and on higher-risk topics and decisions. Even in cases where human involvement is only required for auditing (AIA-HV), it is expected that the effort/responsibility of the auditors will be substantial, as shown in Figure 2. For tasks such as content creation, the human effort is drastically reduced with the effective use of GenAI content. This however, underscores the necessity for human auditing tasks that require increased responsibility on the side of the auditor. The actors enabling this type of human-AI collaborations

There will always be a role for humans in AI pipelines, although GenAI is disrupting the crowdsourcing environment as we know it.

must be mindful, ensuring training and worker qualifications on the use or auditing of AI-generated annotations. As GenAI advances, humans will assume more intricate tasks in terms of responsibility. This requires reassessment of the training of workers, as well as of the design of tasks to include those where GenAI alone does not perform well enough.

Crowdsourcing or AI Sourcing?

There will always be a role for humans in AI pipelines, although GenAI is disrupting the crowdsourcing environment as we know it. Platforms need to set standards, supported by guidelines and templates, for its use. Workers have used tools/scripts for more than 10 years; thus, GenAI can be viewed as a new tool they will add to their stack. This disruption does not mean the end of crowdsourcing but there will be adjustments, including reevaluating quality control mechanisms and payment schemes. While auditing AI output might seem like an “easy” task requiring a yes or no answer, it may actually be a time-consuming process. These changes will have implications for stakeholders in the current ecosystem: platforms, crowd workers, infrastructure, and customers. As GenAI becomes deployable locally rather than in the cloud, it will be easier to perform annotation without the use of crowd platforms. While running internal annotation pipelines will be an option, the need for human oversight and hardware will still make platform-based data annotation pipelines viable.

Involving humans helps us avoid the vicious loop of “AI training AI,” which poses significant risks. In fact, commercial search engines have started to use LLMs to generate relevance judgments and to evaluate and train retrieval models.

Even now, many cloud providers offer fully managed services that include foundational models, MLOps, and human label collection to build, train, and deploy custom ML models at scale.

Involving humans helps us avoid the vicious loop of “AI training AI,” which poses significant risks. In fact, commercial search engines have started to use LLMs to generate relevance judgments and to evaluate and train retrieval models.⁹ After some iterations of training AI with AI-generated labels, the quality of the model’s predictions could quickly drop (for example, due to lack of variability), but also, there is a risk of reinforcing existing bias, if human oversight is no longer required. The adoption of GenAI in annotation tasks will disrupt not only the way these tasks are completed, but also the labeling outcome, by resulting in less diverse labels. The expertise and social knowledge of humans should be harnessed to ensure quality and appropriate diversity, to produce labels that reflect what a broad spectrum of humans “knows.” Training AI on GenAI labels creates risks for the future of AI, where systems may be optimized for accuracy, but carry on invisible issues of fairness and impartiality. Thus, there is a continued need for human oversight.

Platforms should embrace GenAI as co-pilots and thus, they should: in-

corporate these tools into tasks when possible; take responsibility for training workers and requesters, in their responsible use; provide a qualification system to recognize worker expertise in the use of these tools; and promote transparency in the use of GenAI, so that trusted workers are identifiable. Such an approach will preserve the benefits of GenAI. In summary, we see a future where the data annotation ecosystem is significantly different, becoming mostly automated, but still requiring humans in the loop, to ensure the quality and diversity of AI-generated labels. **C**

References

1. Altenried, M. The platform as factory: Crowdwork and the hidden labour behind artificial intelligence. *Capital and Class* 44, 2 (2020).
2. Christoforou, E., Demartini, G., and Otterbacher, J. Generative AI in crowdwork for web and social media research: A survey of workers at three platforms. In *Proceedings of the 18th Intern. AAAI Conf. on Web and Social Media (ICWSM)* (2024).
3. Faggioli, G. et al. Perspectives on large language models for relevance judgment. In *Proceedings of the 9th ACM SIGIR/The 13th Intern. Conf. on the Theory of Information Retrieval* (2023).
4. Fan, S. et al. Socio-economic diversity in human annotations. In *Proceedings of the 14th ACM Web Science Conf.* 2022 (2022).
5. Howe, J. et al. The rise of crowdsourcing. *Wired* 14, 6 (2006).
6. Kaplan, T. et al. Striving to earn more: a survey of work strategies and tool use among crowd workers. In *Proceedings of the AAAI Conf. on Human Computation and Crowdsourcing* 6 (2018).
7. Kittur, A. et al. The future of crowd work. In *Proceedings of the 2013 Conf. on Computer Supported Cooperative Work* (2013).
8. Kocoń, J. et al. ChatGPT: Jack of all trades, master of none. *Information Fusion* 99, (2023).
9. Thomas, P. et al. Large language models can accurately predict searcher preferences. In *Proceedings of the 47th Intern. ACM SIGIR Conf. on Research and Development in Information Retrieval* (2024).
10. Veselovsky, V., Ribeiro, M.H., and West, R. *Artificial Artificial Intelligence: Crowd Workers Widely Use Large Language Models for Text Production Tasks*. (2023); 10.48550/arXiv.2306.07899

Evgenia Christoforou (e.christoforou@cyens.org.cy) is a research associate at CYENS Centre of Excellence Ltd, Nicosia, Nicosia, Cyprus.

Gianluca Demartini (g.demartini@uq.edu.au) is a professor at the University of Queensland School of Electrical Engineering and Computer Science, Saint Lucia, Queensland, Australia.

Jahna Otterbacher (jahna.otterbacher@ouc.ac.cy) is an associate professor at the Open University of Cyprus, Cyprus Center for Trustworthy AI (CyCAT) Latsia, Nicosia, Cyprus.

This project has received funding from the Cyprus Research and Innovation Foundation under grant EXCELLENCE/0421/0360 (KeepA(n)I), the European Union’s Horizon 2020 Research and Innovation Programme under Grant Agreement No. 739578 (RISE), the Government of the Republic of Cyprus through the Deputy Ministry of Research, Innovation and Digital Policy, from the Australian Research Council (ARC) Training Centre for Information Resilience (Grant No. IC200100022), and from an ARC Future Fellowship (Grant No. FT240100022).

© 2025 Copyright held by the owner/author(s).



Association for
Computing Machinery

Career & Job Center

**The #1 Career Destination
to Find Computing Jobs.**



**Connecting you with
top industry employers.**

**Check out these new features
to help you find your next
computing job.**



Access to new and exclusive career resources, articles, job searching tips and tools.



Gain insights and detailed data on the computing industry, including salary, job outlook, 'day in the life' videos, education, and more with our new Career Insights.



Receive the latest jobs delivered straight to your inbox with **new exclusive Job Flash™** emails.



Get a free resume review from an expert writer listing your strengths, weaknesses, and suggestions to give you the best chance of landing an interview.



Receive an alert every time a job becomes available that matches your personal profile, skills, interests, and preferred location(s).

Visit <https://jobs.acm.org/>