

REGROW: Reimagining Global Crowdsourcing for Better Human-AI Collaboration

Andy Alorwu
University of Oulu
Oulu, Finland
andy.alorwu@oulu.fi

Dmitry Ustalov
Yandex
Saint Petersburg, Russia
dustalov@yandex-team.ru

Oliver Bates
Lancaster University
Lancaster, UK
o.bates@lancaster.ac.uk

Jorge Goncalves
University of Melbourne
Melbourne, Australia
jorge.goncalves@unimelb.edu.au

Saiph Savage
Northeastern University
Boston, USA
s.savage@northeastern.edu

Alexey Drutsa
Yandex
Moscow, Russia
adrutsa@yandex-team.ru

Danula Hettiachchi
RMIT University
Melbourne, Australia
danula.hettiachchi@rmit.edu.au

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

Niels van Berkel
Aalborg University
Aalborg, Denmark
nielsvanberkel@cs.aau.dk

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

Ujwal Gadiraju
TU Delft
Delft, Netherlands
u.k.gadiraju@tudelft.nl

ABSTRACT

Crowdworkers silently enable much of today's AI-based products, with several online platforms offering a myriad of data labelling and content moderation tasks through convenient labour marketplaces. The HCI community has been increasingly interested in investigating the worker-centric issues inherent in the current model and seeking for potential improvements that could be implemented in the future. This workshop explores how a reimagined perspective on crowdsourcing platforms could provide a more equitable, fair, and rewarding experience. This includes not only the workers but also the platforms, who could benefit e.g. from better processes for worker onboarding, skills-development, and growth. We invite visionary takes in various formats on this topic to spread awareness of worker-centric research and developments to the CHI community. As a result of interactive ideation work in the workshop, we articulate a future direction roadmap for research centred around crowdsourcing platforms. Finally, as a specific interest area, the workshop seeks to study crowdwork from the context of the Global South, which has been arising as an important but critically understudied crowdsourcing market in recent years.

ACM Reference Format:

Andy Alorwu, Saiph Savage, Niels van Berkel, Dmitry Ustalov, Alexey Drutsa, Jonas Oppenlaender, Oliver Bates, Danula Hettiachchi, Ujwal Gadiraju, Jorge Goncalves, and Simo Hosio. 2022. REGROW: Reimagining Global

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

CHI '22 Extended Abstracts, April 29-May 5, 2022, New Orleans, LA, USA

© 2022 Copyright held by the owner/author(s).

ACM ISBN 978-1-4503-9156-6/22/04.

<https://doi.org/10.1145/3491101.3503725>

Crowdsourcing for Better Human-AI Collaboration. In *CHI Conference on Human Factors in Computing Systems Extended Abstracts (CHI '22 Extended Abstracts)*, April 29-May 5, 2022, New Orleans, LA, USA. ACM, New York, NY, USA, 7 pages. <https://doi.org/10.1145/3491101.3503725>

1 BACKGROUND

Crowdsourcing refers to the division of work into smaller parallel tasks distributed to a workforce, with results being aggregated in some meaningful way. This often takes place via online platforms that provide the functionality needed to interact between the requesters of work and the available online workforce. Paid online crowdsourcing has rapidly become the backbone of industry applications in data enrichment and machine learning, and platforms such as Amazon's MTurk¹, Appen², or Toloka³ are now broadly used to fuel the numerous AI-based innovations that depend on high-quality training data. Collecting and enriching training data typically includes tasks such as image labelling, sentiment analysis, and different types of classification tasks.

The practice of crowdsourcing is not without issues. Workers on crowdsourcing platforms are often left without the attention needed in order to build a fair, sustainable future model for crowdsourcing. There are real people behind the unique participant identification numbers assigned by the platforms, with real-life needs and worries. To this end, HCI scholars have investigated aspects of worker mental health [24], invisible labour [29, 47], data management and reuse [1, 2], dystopian future forms of crowdwork [4], and worker attitudes toward e.g. creative work and skills-development [42]. Further, HCI researchers have recently become interested in leveraging the potential of these platforms for more nuanced and complex use

¹<https://www.mturk.com/>

²<https://appen.com/>

³<https://toloka.ai/>

cases, which might be accompanied by a more rewarding work experience. Here, examples include collaboration and work that requires multiple iterations [5, 46], creative work [42], or digital health and well-being solutions [6, 28].

Our aim with this workshop is to collectively explore what the future of these platforms could look like 10–20 years from now. With this goal, we will not only gather a community of researchers and industry experts around notions of work, but also build knowledge together. Some overarching questions that we will seek to address by drawing on the vast experiences and perspectives present at the workshop include: What are the platform affordances that could best support worker well-being and professional development, fundamental work rights, legal issues? An example starting point here could be *e.g.* the Fairwork⁴ principles on Gig Work and/or Cloudwork or through the lens of stopping Silicon Valley building a new global underclass of workers through crowdwork [26]. Would it be possible to envision data management models to help workers retain some degree of ownership of their personal data while still also benefiting the requesters? We encourage thinking about these questions in the context of identifying how the future of crowdsourcing platforms could look like especially in the Global South, particularly Latin America, where crowdsourcing platforms have started to become important key spaces to find labour⁵ [31, 50]. We acknowledge, however, that many issues to be discussed might be generic and not necessarily specific to this context, and submissions discussing more generic viewpoints are welcome as well.

To this end, related workshops across HCI and relevant venues have been organized, including many by the organizers of this one. Workshops in CHI, CSCW, and UbiComp have examined crowdsourcing from various different angles. For example, a CSCW '17 workshop explored the role of crowdsourcing in the context of law and policy [38]. At UbiComp '18, researchers explored smartphones and other mobile devices as a tool to support crowdworkers and increase data quality [54]. A CHI '19 workshop highlighted opportunities to support creativity in crowdsourcing platforms [41]. NeurIPS '20 and VLDB '21 workshops on trust, ethics and data excellence in crowdsourcing [52, 53]. Finally, at CSCW '21, a workshop was organised around the notion of the hidden labour that powers modern day's AI systems [7].

Building on such past workshops, we are now specifically interested in investigating the affordances of crowdsourcing platforms concerning the following thus far underrepresented or largely ignored topics, including, but not limited to:

- Worker onboarding into the platform ecosystems: How to facilitate a smooth start in a distributed, largely anonymous working environment?
- Offering meaningful work or enriching tasks in meaningful ways.
- Skills-development and personal growth opportunities of workers as a byproduct of work.
- Psychological support and feedback for workers.
- Trust between workers and requesters and how to engage the workers in quality assurance, while keeping the requesters in loop with them.

- Tooling-related support measures for workers, as offered by the platforms themselves (as opposed to being orchestrated externally).
- Novel data management solutions that can, for example, lead to novel incentive mechanisms through retained data ownership.
- New use cases for platform workers far beyond what is feasible currently, leveraging *e.g.* XR technologies, metaverse concepts, or any other futuristic technologies.
- Ensuring the fundamental work rights of crowd workers, for example, the legality of their contracts, their access to workplace benefits (*e.g.* sick leave, parental leave, pensions).

To foster viewpoint diversity we seek to accept visions and considerations from any other aspect relevant to the future potential of paid crowdsourcing, and from the perspective of any relevant stakeholders.

Impact statement: We are sensitive to the fact that often CHI workshops happen 'by researchers, for researchers', and it may be unrealistic to expect results that contribute concrete functionalities or instantly implementable policies. With this in mind, our intended impact is two-fold: 1) Spread awareness of worker-centric research and developments on the current platforms and 2) articulate a concrete research roadmap for crowdsourcing that integrates the identified global challenges. Doing so, we contribute our small part in redirecting the community's attention toward the workers' benefit – to a degree dictated ultimately by the workshop attendance and post-activities.

We will offer a low barrier of entry concerning submission types, accepting *e.g.* small-scale empirical studies, written essays, concept videos, and pictorials. We encourage creativity and futurism in form of visionary work. All types of visions are welcome: utilitarian, dystopian, utopian, or logical and practical. The organizers will use the submissions to curate an initial set of possible future affordances for the platforms to be briefly presented during the workshop and used during the workshop as a discussion stimulant. As a concrete output of the collaborative workshop, we will contribute a curated set of reflections, open research questions, and a concrete roadmap to *'what can the future of paid crowdsourcing platforms look like 10–20 years from now?'*

Integrating Critical Theory to Design the Future of Global Crowdsourcing. For CHI it can be important to question whether new proposed designs for the future of crowdsourcing are truly engaging and addressing critical societal problems that current crowdsourcing platforms are generating, or if the new designs are simply starting to touch on the mere symptoms of a problem [3]. The latter case can result when we create designs that make a societal problem more bearable. Without addressing the root of the problem, it is, however, possible that our proposed designs could actually reinforce critical structural issues, causing more abuse and harm to workers.

To address this problem, our workshop will draw from theorist Hebert Marcuse from the Frankfurt School of Critical Theory [30], who has argued that it is extremely challenging to engage in critical analysis of the structures and processes that exist in society [3, 37]. But a way to start practising more critical analysis is via artistic creativity [36]. According to Marcuse, artistic creativity facilitates

⁴<https://fair.work/en/fw/homepage/>

⁵<https://restofworld.org/2021/refugees-machine-learning-big-tech/>

creating designs that can truly challenge the current reality of what is possible and allows us to consider directions we might have been blind to consider otherwise. Based on these ideas, we plan to engage participants of our workshop in 'creative artistic co-design sessions' where they use fictional narratives to design 'alternative realities' to contemporary digital labour platforms and tools [20, 35].

Designing Crowdsourcing Tools and Platforms for Wider Audiences. Our goal is to focus one part of the design of solutions for the future of crowdsourcing within the context of Latin America as a prominent example. We focus on this population and culture given that studies have shown that a majority of the A.I. industry is relying on a global workforce located in the Global South, particularly Latin America, to complete their data labelling tasks [8, 27, 31, 32, 44, 45, 55]. For example more than 75% of the labelled training data that Tesla used for its self-driving car came from Venezuelan crowd workers [9, 50]. Additionally, we take advantage of the fact that some of the organisers of the workshop are leaders designing for Latin America [10, 21, 22, 39, 48].

Note that within our workshop, we do acknowledge that it can be difficult for participants to be able to analyse how certain crowdsourcing designs might be affecting Latin America and design adequate solutions, especially if these individuals are not from the region or familiar with the related culture [12, 49]. For this purpose, we will include workshop activities where we help participants to adopt the "*circuit of culture framework*" [18]. This framework provides culture probes that can help outsiders to Latin America to understand how a particular technology is currently being interpreted and understood by people of that region. Such cultural understanding facilitates creating better interventions and positive design changes for Latin America [33]. We will also include hybrid activities in our workshop that include outsiders and insiders to Latin America, e.g., co-design sessions with locals from Latin America and workshop participants from other parts of the world. Through these co-design sessions and by adopting the circuit of culture framework, we will be able to empower all participants to better understand the type of representation and identity that Latin Americans attribute to crowdsourcing platforms, as well as the type of production, consumption, and regulation that locals relate to crowd work [25, 51]. Through this, we will all start to collectively understand how Latin America is encoding meaning and lifestyles into crowdsourcing platforms. We will then use the understanding to collectively design better artefacts that power a better future of work for the region.

Designing the Future of Crowdsourcing with Workers, Requesters, and Platform Owners For crowdsourcing platforms to truly bring positive futures, we need the platforms to enable success for the different stakeholders: those providing the labour, those requesting it, and the platforms who benefit from productive commerce. In our workshop we will focus on having conversations on how can we imagine, design, and structure the data supply ecosystem to benefit all three parties. We adopt value-sensitive design as a principled framework to integrate the diverse needs of multiple stakeholders in the overall data labelling ecosystem [23]. To put this into practice in our workshop, we will have guided activities using the value sensitive design framework to help participants identify the different needs of the different stakeholders, as well as points where there are conflicts between needs. Through this, we

will empower participants to start to design solutions that aim to empower all stakeholders [14]. We will aim to include also all stakeholders as participants in our workshop. Currently the workshop organisers already encompass some of these different stakeholders. Notice that this approach also differentiates our workshop from previous efforts that focused primarily on workers' needs instead of the whole ecosystem.

2 ORGANISERS

The organisers of the workshop come from a variety of backgrounds and all have expertise and have been involved in the field of crowdsourcing, data management, and human computation over the years. With representation also from the microwork industry, the team is in a good position to disseminate and put forward the workshop's contributions. Many of the organizers have also been involved in the organization of a series of workshops and tutorials on crowdsourcing and human computation over the years [13, 15–17, 41, 43, 52–54, 56].

Andy Alorwu is a Doctoral Researcher in the Crowd Computing research group at the Center for Ubiquitous Computing, University of Oulu, Finland. His research interests include personal data management, m-health, and mobile computing.

Saiph Savage is an Assistant Professor at the College of Computer Science at Northeastern University in Boston. Dr. Savage is also a Research Collaborator at the Civic Innovation Lab at the Universidad Nacional Autonoma de Mexico (UNAM) in Mexico City. Dr. Savage has vast experience designing tools to empower crowd workers. She also has experience designing technology for the global south, especially Latin America.

Niels van Berkel is an Associate Professor at the Human-Centered Computing Group at Aalborg University. His research interests lay in Human-Computer Interaction, Social Computing, and Ubiquitous Computing.

Dmitry Ustalov is the Head of Research at the Toloka crowdsourcing platform. His research interests are focused on Crowdsourcing, Natural Language Processing, and Evaluation.

Alexey Drutsa is the Head of Efficiency and Growth Division at the Toloka crowdsourcing platform. His research interests are focused on Machine Learning, Data Analysis, Auction Theory.

Jonas Oppenlaender is an alumni of the Crowd Computing research group at the Center for Ubiquitous Computing, University of Oulu, Finland. His research interests include crowdsourcing, crowd feedback systems, and crowd-powered creativity support systems.

Oliver Bates is a Research Fellow at the School of Computing and Communications, Lancaster University, UK. He is broadly interested in labour rights especially in the gig economy.

Danula Hettiachchi Research Fellow at the ARC Centre of Excellence on Automated Decision Making and Society and RMIT School of Computing, Australia. His research interests include crowdsourcing, social computing and human-computer interaction.

Ujwal Gadiraju is an Assistant Professor at the Web Information Systems (WIS) group of the Software Technology department at TU

Delft in the Netherlands. He is a Director of the Delft Design@Scale AI Lab and leads a research line on Human-Centered AI and Crowd Computing at the WIS group.

Jorge Goncalves is a Senior Lecturer in the School of Computing and Information Systems at the University of Melbourne. He has conducted extensive research on improving crowd data quality, and bringing crowdsourcing beyond the desktop by using ubiquitous technologies. He has also served as Workshops Co-Chair for CHI'19 and CHI'20.

Simo Hosio is an Associate Professor at the Center for Ubiquitous Computing, University of Oulu, Finland. He leads the Crowd Computing Research Group and is affiliated with the Center for Life Course Health Research. His research interests include crowd-sourced wellbeing solutions, social computing and ubiquitous computing.

3 LINK TO WEBSITE

The workshop's website will be hosted on <https://crowdscience.ai/>, as part of the Crowd Science Initiative, and will be finalised upon acceptance of this proposal. The website will be used to spread the word about the workshop and act as a central repository for all important information to attendees. This includes the submission date, workshop date, submission modality, and links to related material so that candidates can get familiar with the scope of the subject and the goals of the workshop. A detailed schedule of the workshop and requirements for participants will be provided, as well as, the background of each organiser. Further, the website will host the primary output of the workshop; the future vision statement and the related Miro board with condensed versions of the teams' work, and accepted position papers will be available for download on the dedicated website.

4 PRE-WORKSHOP PLANS

We will first finalise the workshop website, to act as a simple information hub about the workshop and, later on, as a hub that links to all workshop submissions which are stored online.

We require submissions from all workshop participants in the format of short visions that may deal with one or more of the topics of the workshop, as listed earlier. Prior to the workshop, we mostly focus on spreading awareness of the workshop and distributing the CfP across all relevant academic as well as industry mailing lists (e.g., chi-announcements). We will further use our own distribution lists (based on a variety of workshops and conferences we have held in the past). Workshop organisers will also use social media channels at their disposal to publicise the CfP.

Each submission will be checked for topicality and reviewed by at least two of the organisers. While we are not planning to reject any submissions based on empirical contribution depth, we reserve the right to reject submissions with no connection to any of the workshop themes.

The submissions will be curated into a publicly available Notion dashboard (<https://www.notion.so/>) that will be made available to the workshop participants a week before the workshop date. This way, participants can already familiarise themselves with other

participants as well as their interests and positions concerning the future of crowdsourcing platforms.

5 WORKSHOP STRUCTURE

The workshop emphasises discussion, teamwork, and collaboration in teams of 2–4 members. We aim for 15–20 participants (excluding the organisers). As we have a fairly diverse and numerous organiser board, we hope to attract a lot of expertise around the subject matter and from various different backgrounds to facilitate different viewpoints. There will be no traditional paper presentations; instead we will focus on ideation and collaborative activities that are supported by the potentially informal and hopefully provocative submissions by participants.

We start the workshop with a brief keynote presentation by an industry representative, followed by rapidly moving into break out groups. All participants have access to the workshop submissions through a shared Notion workspace online. In these breakout discussions and informal groupings in person participants will receive fiction prompts to start co-designing with workers, academics, and industry actors new crowdsourcing interfaces. In specific, we use the context and concepts of design fictions to enable workshop participants to conceptualise, explore, and critique new design ideas for crowdsourcing platforms [34, 35, 57]. Here we will build off our research studying and organising hybrid events [19],⁶ especially with minorities, to now leverage the crowd in different ways (the people in-person could build and hack things, while the audience online could search for information, such as needs and facts of the region to help motivate the design. We will also utilise the hybrid aspect to facilitate the participation from workers and researchers from Latin America, who might not be able to travel to the conference due to costs.

Following a 30-minute coffee break during which we will organize an informal panel discussion, the participants continue the morning's tasks in groups, focusing specifically on novel platform affordances. Doing so, we hope the different groups can bring along their unique expertise and viewpoints to each of the themes that were derived from the initial ideas. Each group will work in a shared Miro board.

The session after lunch starts with team presentations and discussion. We generously reserve time for this activity, as experience has shown that HCI workshops typically run out of time when participants get to present and discuss newly-created work and ideas. This will also help participants refine and rethink some of their ideas from the earlier stage. In the final stage, we ask for the groups to condense the key objectives, research questions and other ideas to the main Miro board of the workshop. This way, the main board acts as a live memo for the organisers who will use it to create the overarching vision statement on the workshop website. Finally, we will have one panel discussion around the key shared concepts emerging during the day and close the workshop.

⁶<https://www.milenio.com/cultura/hackaton-une-mujeres-crear-casas-inteligentes>, <https://www.criptonoticias.com/educacion/unam-anfriton-hackaton-blockchain-mexico/>, <https://devday4w.com/blog/mujeres-conectadas/>

Table 1: Proposed structure of the workshop

| First Session | |
|----------------|--|
| 30 min | Workshop introduction and keynote |
| 60 min | Break out in teams, start co-designing new futures (value-sensitive methods) |
| 30 min | Coffee break, informal panel discussion |
| 60 min | Teamwork continues, focusing on further conceptualising new platform affordances |
| 60 min | Lunch break |
| Second Session | |
| 60 min | Team presentations and discussion |
| 15 min | Teams create a synthesis of their work in a shared workspace online |
| 30 min | Panel discussion focusing on the created synthesis |
| 15 min | Final thoughts, closing the workshop |

6 POST-WORKSHOP PLANS

One of the primary goals of the workshop is to create a vision with a set of research questions and engineering challenges – a roadmap for future. These will be summarised on the workshop website, as a central repository of research ideas to explore. The workshop results will be communicated to a larger audience through the extensive network of the organizers. We will also ask for volunteers among the organisers, as well as from the participating platform workers and industry participants, to act as a task force to publish the workshop results in an academic format in a suitable venue (*e.g.*, HCOMP or related workshops).

7 REMOTE/ONSITE/ASYNCHRONOUS PLANS

We will organise this workshop as a hybrid event, with most participants encouraged to participate physically in the event as part of the CHI conference. Selected virtual attendees will be requested to fill out an online form regarding their timezones, platform restrictions, and accessibility requirements, and interests in topics related to crowdsourcing. This information will be used to schedule the workshop, arranging an online communication format, and creation of discussion groups. Most importantly, all practical work takes place on a shared central Miro board, where we can facilitate also asynchronous work. Short presentations can be posted there as video messages using a free tool such as Loom, but for the most part the ideation work is shaped and edited in written form, uploads of sketches, and diagrams directly in Miro. Based on current research in hybrid events that has shown that shorter, more focused sessions keep attendees' attention better [11, 40]. We will organise different activities within the workshop to keep delegates moving between activities and create a flow for the day. As mentioned in the workshop structure, some of the activities will include: panels with global south crowd workers, academics, industry actors who design the crowdsourcing platforms. We will also have collective discussions on the type of labour conditions and power dynamics present in current crowdsourcing platforms. The audience of the panel will participate via fireside chats.

All materials will be distributed as fully digital using Miro and Notion, which will further help remote participation. The presentations will be streamed online for remote participants to view by using a tabletop microphone and a web camera.

8 250-WORD CALL FOR PARTICIPATION

This hybrid workshop reimagines crowdsourcing to better answer to the modern global worker-centric challenges and specific issues in the Global South.

Following a juried process, we invite anyone interested in the future of crowdsourcing to submit vision statements with an overarching question of what could these platforms look like in 10-20 years from now. To this end, we focus on the following topics (for the full list of topics, please see the workshop website):

- Platform affordances: onboarding issues, skills development, personal growth opportunities
- Psychological support and feedback for workers
- Trust between workers and requesters: Fair quality assurance and good communication
- Tooling-related support measures for workers
- Novel data management solutions that can, for example, lead to new incentive mechanisms
- New use cases for platforms beyond what is feasible currently, leveraging *e.g.* XR technologies, metaverse concepts, or other futuristic technologies
- Ensuring the fundamental work rights of crowd workers, for example, the legality of their contracts, access to workplace benefits (*e.g.* sick leave, parental leave, pensions)

We solicit submissions through the workshop website in the ACM Primary Article Template, with links to *e.g.* online videos or other necessary resources. We do not impose page limits, but we encourage the authors to keep submissions as concise as possible. At least one author of each accepted submission must attend the workshop and that all participants must register for both the workshop and for at least one day of the conference. For details please visit https://crowdsourcing.ai/conference_events/chi22.

ACKNOWLEDGMENTS

This workshop is connected to the GenZ strategic profiling project at the University of Oulu, supported by the Academy of Finland (project number 318930), CRITICAL (Academy of Finland Strategic Research, 335729), EPSRC Grants EP/R044929/1 and EP/S027726/1, the TU Delft AI Labs Programme and the Delft Design@Scale AI Lab. Part of the work was also carried out with the support of Biocenter Oulu, spearhead project ICON.