



Ethical Tensions, Norms, and Directions in the Extraction of Online Volunteer Work

Hanlin Li
lihanlin@u.northwestern.edu
Northwestern University
Evanston, IL, USA

Leah Ajmani
ajman004@umn.edu
University of Minnesota
Minneapolis, MN, USA

Moyan Zhou
zhou0972@umn.edu
University of Minnesota
Minneapolis, MN, USA

Nicholas Vincent
nickvincent@u.northwestern.edu
Northwestern University
Evanston, IL, USA

Sohyeon Hwang
sohyeonhwang@u.northwestern.edu
Northwestern University
Evanston, IL, USA

Tiziano Piccardi
tiziano.piccardi@epfl.ch
EPFL
Lausanne, Switzerland

Sneha Narayan
snarayan@carleton.edu
Carleton College
Northfield, MN, USA

Sherae Daniel
daniel@ucmail.uc.edu
University of Cincinnati
Cincinnati, OH, USA

Veniamin Veselovsky
veniamin.veselovsky@epfl.ch
EPFL
Lausanne, Switzerland

ABSTRACT

Online volunteer work such as moderating forums and participating in open source projects not only underpins today's digital infrastructures, but also helps companies generate immense profits. However, there remains a lack of ethical norms around using volunteer labor for corporate interests, opening opportunities for unchecked extraction of online volunteer work at scale. Early evidence suggests that the extraction of online volunteer work may have negative implications on the tech ecosystem and obfuscate the potential for exploitative labor practices. In this workshop, we invite participants to discuss 1) what ethical tensions exist in the current approaches to extracting online volunteer work, 2) what ethical norms should be followed or recommended and 3) what are the opportunities for social computing technologies to promote these norms. Furthermore, we open a dialogue around whether online platforms should be providing non-monetary compensation, such as education and resources, that is often promised in in-person volunteer settings. We plan to involve a diversity of roles beyond academic researchers, such as online volunteers and practitioners to discuss these questions.

CCS CONCEPTS

• **Human-centered computing** → HCI theory, concepts and models.

KEYWORDS

peer production, monetization, volunteer, data labor

ACM Reference Format:

Hanlin Li, Leah Ajmani, Moyan Zhou, Nicholas Vincent, Sohyeon Hwang, Tiziano Piccardi, Sneha Narayan, Sherah Daniel, and Veniamin Veselovsky. 2022. Ethical Tensions, Norms, and Directions in the Extraction of Online Volunteer Work. In *Computer Supported Cooperative Work and Social Computing (CSCW'22 Companion)*, November 8–22, 2022, Virtual Event, Taiwan. ACM, New York, NY, USA, 5 pages. <https://doi.org/10.1145/3500868.3560923>

1 INTRODUCTION

Many prominent technologies such as user-generated content platforms and open source software are built upon the backs of online volunteers who freely contribute their time and knowledge at an unprecedented scale. By producing content and writing code, online volunteers not only help to support key digital infrastructures but also contribute to for-profit entities, which increasingly repurpose the outputs of volunteer work for corporate interests. For example, in addition to supporting Wikipedia readers, Wikipedia editors' work is also featured in Google Search and benefits Google Search users [8, 12, 14], even though this may not be the intention of these editors. Similarly, the Linux project has been foundational to commercial computing services such as Amazon Web Services [3].

This workshop builds upon the literature on open source projects and peer production to call for increased attention to ethical issues in the extraction of online volunteer work. While CSCW researchers have a rich history of studying the processes underlying online volunteer work, the broad, societal implications of the extraction of online volunteer work remain under-investigated. In particular, unlike in the case of in-person volunteering, there is a lack of ethical norms around the extraction of online volunteer work for profit, which has been met with questions and critiques. As an increasing number of for-profit technology businesses benefit from or repurpose the outputs of online volunteer work, a form of public goods, recent research has started to examine volunteers' role in the tech ecosystem with a broader lens. Specifically, as online volunteer work becomes increasingly valuable to the tech industry, questions arise around whether for-profit companies are exploiting

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).

CSCW'22 Companion, November 8–22, 2022, Virtual Event, Taiwan

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

ACM ISBN 978-1-4503-9190-0/22/11.

<https://doi.org/10.1145/3500868.3560923>

volunteer labor and whether online volunteers should be compensated or have a say over the downstream re-appropriation of their work. For example, research on Wikipedia's role in the success of Google Search, Reddit, and StackOverflow suggests that the relationship between online volunteers and these commercial products is highly unilateral: while Wikipedia, a volunteer-powered platform, improves the click through rates of Google Search and user engagement, it receives little benefits (i.e. edits) in return [8, 13].

In light of online volunteer work's broader role in the technology landscape, we aim to invite diverse stakeholders including researchers, practitioners, and volunteers to catalyze a research agenda investigating the ethical tensions stemming from the extraction of online volunteer work for profit. Additionally, we aim to discuss what ethical norms should be followed when companies utilize online volunteer work for their benefits. We will also facilitate discussion and ideation on how research communities could better support these ethical norms and empower online volunteers to protect their work if they wish to do so.

2 KEY THEMES

Our workshop will focus on three key themes around the extraction of online volunteer work: 1) creating a research agenda for future work to better understand ethical tensions in the extraction of online volunteer work, 2) constructing ethical norms around the extraction, i.e. how could companies ethically use online volunteer work for profit and self-interests? and 3) proposing a roadmap for the research community to promote such norms.

2.1 Catalyzing a Research Agenda on Ethical Tensions in the Extraction of Online Volunteer Work

Several instances of ethical tensions when online volunteer work is monetized have emerged, particularly when this work is repurposed by third-party entities. Below we provide some examples of such tensions. These will additionally serve as prompts for participants to construct a research agenda to gain a more comprehensive understanding of the ethical tensions in this space.

- **Unfair exchange:** While companies use or repurpose online volunteer work for profits, they often provide few benefits to the volunteer organization or online volunteers in return. For example, research has shown there exists a somewhat one-sided relationship between Wikipedia and those that benefit from Wikipedia content, including Google Search, Reddit, and StackOverflow [8, 13]. While Wikipedia contributes valuable content to these other for-profit, prominent companies, it receives benefits in return that could be seen as disproportionately small in comparison. We seek to involve the research community to map out what other unfair transactions occur to online volunteers and opportunities to make such transactions transparent to online volunteers.
- **Lack of consent:** Volunteers contributing to open source software and technologies often are unable to express non-consent when their work is used to harmful ends [10]. Similarly, volunteers producing content in online spaces are unable to express non-consent of their outputs for products and services to which they do not want to contribute. We

invite CSCW scholars to reflect upon the lack of consent in extracting online volunteer work, discuss recent research proposals that aim to address this issue (e.g. behavioral licensing of AI models and data [4]), and brainstorm new opportunities.

- **Lack of volunteer involvement in decision making:** Online volunteers are not always involved in the decision making process in volunteer-powered organizations. For example, online volunteers who perform content moderation work for Reddit have long voiced their dissatisfaction with the company ignoring their suggestions and needs [7]. Our workshop aims to promote discussion on the potential steps towards democratic governance of online volunteer work.
- **Undue external influence:** While non-profit systems like Wikipedia and OpenStreetmap are mainly powered and shaped by online volunteers, they are still subject to corporate influence. In particular, recently, OpenStreetMap is faced with an influx of corporate editors from companies that use OpenStreetMap's data and services. Research shows that corporate editors' focuses and contribution patterns are distinct from those of volunteer editors, raising questions about the influence of corporate interests on the peer production platform [2, 11]. We invite conversations about what other online volunteer environments may be subject to corporate influence and the downstream implications of such influence.

In summary, we invite researchers to reflect upon the industry's extraction practices around online volunteer work and construct a research agenda to investigate tensions and potentially invisible harms stemming from these practices.

2.2 Constructing Ethical Norms

In online settings, the outputs of volunteer work are easily accessible to the public, opening up opportunities for companies to benefit from online volunteer work for free in a potentially exploitative manner. This workshop aims to promote discussion around what ethical norms should be followed when re-appropriating online volunteer work for use cases that are beyond what online volunteers intended. Specifically, we ask 1) under what circumstances does the extraction of online volunteer work become unethical and exploitative? 2) how and when should volunteers' and volunteer organizations' consent be sought when the extraction occurs? 3) should the beneficiaries of online volunteer work reciprocate with resources such as education and donations to support the online volunteer community? We will provide case studies for participants to reflect upon, speculate, and critique. We list two possible case studies below and if we find suitable ones from submissions, we will include them as well.

2.2.1 Case study 1: open source software. Prominent open source software projects, such as Linux, are possible because of numerous volunteer developers. Research has detailed the various complex activities that volunteers perform in order to keep their projects active, including engineering tasks and invisible managerial work [5]. The fruit of their labor has been integrated into many aspects of technology design and development and is foundational to many commercial products and services [9], such as Amazon Web Services

and android devices. Moreover, code written by volunteer developers may be used to train advanced models, as seen in GitHub's new product, Copilot. In light of the various downstream uses of open source software, should companies give back or compensate online volunteer communities? Should companies seek consent from online volunteers or volunteer organizations when re-appropriating their work?

2.2.2 Case study 2: Wikipedia and its Enterprise API. Open and free knowledge production in Wikipedia both provides key educational and informational resources for the public and powers a variety of services and products by corporate entities: fact-checking for YouTube, Knowledge Panel in Google Search, training datasets for large language models and knowledge graphs. These services draw in immense volumes of income to companies. In 2021, the Wikimedia Foundation launched an opt-in program for organizations and firms to use data from Wikimedia projects, including Wikipedia: Wikimedia Enterprise API [1]. Wikimedia Enterprise API is designed to provide easy access to high volume data with a cost and currently has Google and Internet Archive as its key customers. By repackaging Wikipedia editors' volunteer work, it generates revenue to support the operation of the Wikimedia foundation. Does the monetization of Wikipedia editors' work introduce ethical tensions? Given that volunteers have stated being less motivated to contribute their work if it is used for monetization [15], what are the potential ways for the Foundation to gain financial support without hurting the vibrancy and productivity of online volunteer communities?

2.3 Roadmapping Directions towards Ethical Extraction of Online Volunteer Work

Building upon participants' reflection on the ethical tensions and potential norms in this space, we plan to invite discussion and ideation to create a roadmap for promoting ethical use of online volunteer work. Specifically, we will facilitate brainstorming on the following questions: how could the research community encourage different roles (e.g. third-party practitioners and researchers) to follow participants' proposed norms? Given the long history of the CSCW research community's focus on labor, value-sensitive design, and empowerment, we are particularly interested in exploring what lessons we can learn from the body of CSCW scholarship and what future research directions can help promote ethical practices in leveraging the outputs of online volunteer work. Below are a few possible guiding questions:

- **How to design for the transparency of volunteer labor?** How can we recognize online volunteers' role in upholding the tech ecosystem through design? How can this be done at the time of labor production (e.g. measuring the value of volunteer labor [6])? And how will this differ in systems that have performed additional processing or repackaging of volunteer work, e.g. large language models and knowledge graph?
- **How to distribute the burden and cost of online volunteer work?** How can we ensure that companies that use online volunteer work share the burden/cost? What tools and measures need to be developed to make this happen?

- **Do online volunteers' goals and values always align with those of companies that re-appropriate online volunteer work?** Should volunteers have control over how their work is re-appropriated downstream so their values are respected? Do volunteers want control over their work? If so, in what ways?
- **Should volunteers receive some form of compensation for their labor?** In-person volunteering often comes with non-monetary compensation, such as tax credits or mutual aid when they fall on hard times. Are there similar mechanisms we can design within online ecosystems?

3 WORKSHOP LOGISTICS

We plan to host a one-day workshop including a keynote, a moderated panel discussion, lightning talks, small group discussions, and debriefing. To accommodate time zone differences, we plan to divide our workshop into two half-day sessions, with one session geared towards participants using the North American time zones and the other for East Asian time zones. Each session will be three hours long. The first session will include 1) the keynote address, 2) part of lightning talks, and 3) a small group discussion. The second session will include 1) the moderated panel, 2) the remaining part of lightning talks, 3) a small group discussion, and 4) debriefing. Because the two sessions will not overlap in content, we welcome and encourage participants to join both sessions. To facilitate interaction between the two sessions, we will create a digital space for participants to network, provide feedback, and discuss ideas asynchronously. Organizers will post polls and other structured activities in this space to bridge the two sessions together and create opportunities for cross-session interaction. We will also record our keynote address and panel discussion.

3.1 Description of Activities

A keynote and a moderated panel. We will recruit a speaker and four to six panelists to explicate the amount of efforts and expertise involved in online volunteer work, which provides important context for our workshop's themes. We will recruit at least one experienced online volunteer to participate in the panel discussion to bring in the perspective and experiences of front-line workers. As stated above, the keynote will take place in the first session and the panel will take place in the second session.

Lightning talks. We will have two blocks of lightning talks and each participant will have five minutes to present their extended abstracts. From past experience, we find that having lightning talks help participants better understand each other's background and stimulate discussion around the topics of the workshop.

Small group discussion. We will have two small group discussion sessions. The first small group discussion (in Session One) will focus on identifying ethical tensions in extraction of online volunteer work (Theme 1), and the second small group discussion (in Session Two) will focus on constructing ethical norms based on case studies and brainstorming potential ways to promote these norms (Theme 2 and 3). Both sessions will be mediated by organizers: each group will have at least one organizer facilitating discussion. Discussions will

also be scaffolded by shared note-taking and visual brainstorming on Miro.

Debriefing. We will have a short debriefing session at the end of the workshop summarizing key inputs from participants and mapping out main conclusions from activities. We will also provide the space for participants to brainstorm how to continue this conversation and establish collaborations with different disciplines and groups moving forward to promote ethical extraction of online volunteer work.

3.2 Schedule

As stated above, we plan to have two half-day sessions. The detailed schedule can be found below.

First half-day session (in North America Central time):

- 9:00am-10:00am: Keynote and Q&A
- 10:10am-11:00am: Lightning talks and Q&A
- 11:10am-12:00pm: Small group discussion

Second half-day session (in Taipei Standard time):

- 7:00am-8:00am: Moderated panel
- 8:10am-9:00am: Lightning talks and Q&A
- 9:10am-10:00am: Small group discussion

3.3 Participant Recruitment

We expect a maximum of 50 participants for this workshop. We will use social media (e.g. Twitter) and mailing lists to invite academic researchers to submit extended abstracts. We will also invite experienced online volunteers on prominent platforms, e.g. Wikipedia, Reddit, OpenStreetMap, to participate in our panel and dialogue with researchers. We are in the process of securing funding to compensate potential volunteers for their participation in our workshop.

Participant Submissions and Review Process. Individuals interested in participating will be asked to submit an extended abstract for a 5-minute lightning talk that they will give as part of the workshop activities. Submissions should be 2-4 pages long and should outline the participants' research interests and how they relate to the key themes of our workshop. Accepted abstracts will be used to create subgroups for activities and adjust any activities appropriately, especially depending on the skew of time zones of participants.

The organizing team will review submissions and seek external reviewers if necessary. We plan to have two to three reviewers for each submission. Reviewer comments will be shared with authors. We welcome early-stage analytical work, case studies, essays, and design fictions.

3.4 Workshop Modality, Accessibility, and Accommodation

We will host the workshop in a private Zoom meeting, with live captioning on. We will use Miro, a visual collaboration platform, for shared note-taking and brainstorming. Additionally, as noted earlier, we will create a separate Miro board for participants to have asynchronous networking and discussion. This board will have empty name cards for participants to fill out with their research interests, contact information, and lightning talk presentations.

3.5 Expected Outcomes

Through this workshop, we expect the following main outcomes:

- (1) At a high level, facilitating conversations between the research community and online volunteers on the extraction of online volunteer work.
- (2) A research agenda to guide future work to investigate ethical tensions, potential harms, and implications of the extraction of online volunteer work for profit.
- (3) A manifesto or proposal of ethical norms for the extraction of online volunteer work for broader circulation.
- (4) A roadmap outlining opportunities to mitigate the negative implications of the extraction of online volunteer work and promote ethical approaches.

3.6 Post-Workshop Plans

We will explicate our workshop outcomes in the format of a blog post or a submission to the Communications of ACM. Our hope is by distributing the outcomes of our workshop more broadly, we can engage with more members of the CSCW community in the discourse on ethical vs. exploitative extraction of online volunteer work. We will also create a mailing list for participants to facilitate post-workshop communications.

4 ORGANIZERS

Our organizing team consists of researchers coming from a diversity of disciplines (HCI, computational social science, information systems) across five institutions, who collectively contribute expansive research expertise on a variety of online volunteer work.

Hanlin Li is a PhD candidate in Technology and Social Behavior at Northwestern University. She studies the social and economic impact of online volunteer work such as content moderation and online ratings. Her recent research has examined the monetary value of Reddit moderators' volunteer labor.

Leah Ajmani is a PhD student in Computer Science at the University of Minnesota. She studies human-centered machine learning in mental health settings and online communities. Recently, she has been working on how different philosophy methods can assist ML practitioners in ethical reflections.

Moyan Zhou is a PhD student in Computer Science at the University of Minnesota. She is interested in online community formation, development, and regulation in the context of peer production platforms including Wikipedia. She is also interested in information usage and user effects in recommendation systems.

Nicholas Vincent is a PhD candidate in Technology and Social Behavior at Northwestern University. His work focuses on studying the dependence of modern computing technologies, including the broad set of systems called "AI", on human-generated data, with the goal of mitigating negative impacts of these technologies.

Sohyeon Hwang is a PhD student in the Media, Technology, and Society program at Northwestern University. She studies patterns in heterogeneous governing practices across diverse and overlapping online communities, to understand the downstream implications for social computing platforms such as Wikipedia and Reddit.

Tiziano Piccardi is a Ph.D. student at EPFL (Swiss Federal Institute of Technology) in the Data Science Lab and a formal collaborator of Wikimedia Research. His work is focused on understanding

and improving Wikipedia by designing recommender systems and modeling users' behavior on the platform.

Sneha Narayan is an Assistant Professor of Computer Science at Carleton College. Her research focuses on understanding how technical features and social processes in volunteer-run online communities are used to foster participation, support newcomers, and bolster self-governance.

Sherae Daniel is an Associate Professor of Operations, Business Analytics and Information Systems in the Carl H. Lindner College of Business at the University of Cincinnati. She earned her Ph.D. in Information Systems from the Robert H. Smith School of Business at the University of Maryland. Sherae's research seeks to reveal how to best manage collaboration challenges in nontraditional work environments. In particular, she seeks to uncover the keys that will unlock doors to future success for OSS collaborators. Sherae's research has been published in premier outlets such as *Information Systems Research*, *MIS Quarterly*, and the *Journal of Association for Information Systems*. She is a member of the Association for Information Systems.

Veniamin Veselovsky is a MSc Research Scholar at EPFL in the Data Science Lab. His work has focused on understanding editing behaviour and the social connotations of knowledge. Previously he worked in the Computational Social Science Lab at the University of Toronto.

REFERENCES

- [1] 2022. Wikimedia Enterprise announces Google and Internet Archive as its first customers; allows new customers to self sign-up for free trials. (2022). <https://wikimediafoundation.org/news/2022/06/21/wikimedia-enterprise-announces-google-and-internet-archive-first-customers>
- [2] Jennings Anderson, Dipto Sarkar, and Leysia Palen. 2019. Corporate editors in the evolving landscape of OpenStreetMap. *ISPRS International Journal of Geo-Information* 8, 5 (2019), 232.
- [3] Slaven Brumec and Neven Vrčec. 2013. Cost effectiveness of commercial computing clouds. *Information Systems* 38, 4 (2013), 495–508.
- [4] Danish Contractor, Daniel McDuff, Julia Katherine Haines, Jenny Lee, Christopher Hines, Brent Hecht, Nicholas Vincent, and Hanlin Li. 2022. Behavioral Use Licensing for Responsible AI. In *2022 ACM Conference on Fairness, Accountability, and Transparency* (Seoul, Republic of Korea) (FAccT '22). Association for Computing Machinery, New York, NY, USA, 778–788. <https://doi.org/10.1145/3531146.3533143>
- [5] R Stuart Geiger, Dorothy Howard, and Lilly Irani. 2021. The Labor of Maintaining and Scaling Free and Open-Source Software Projects. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW1 (2021), 1–28.
- [6] Hanlin Li, Brent Hecht, and Stevie Chancellor. 2022. Measuring the Monetary Value of Online Volunteer Work. In *Proceedings of the International AAAI Conference on Web and Social Media*, Vol. 16. 596–606.
- [7] J Nathan Matias. 2019. The civic labor of volunteer moderators online. *Social Media+ Society* 5, 2 (2019), 2056305119836778.
- [8] Connor McMahon, Isaac L. Johnson, and Brent Hecht. 2017. The Substantial Interdependence of Wikipedia and Google: A Case Study on the Relationship Between Peer Production Communities and Information Technologies.. In *ICWSM*. 142–151.
- [9] Pratyush Nidhi Sharma, Sherae L Daniel, Tingting Rachel Chung, and Varun Grover. 2022. A Motivation-Hygiene Model of Open Source Software Code Contribution and Growth. *Journal of the Association for Information Systems* 23, 1 (2022), 165–195.
- [10] Olivia Solon. 2019. Facial recognition's 'dirty little secret': Millions of online photos scraped without consent. *NBC News* (2019).
- [11] Veniamin Veselovsky, Dipto Sarkar, Jennings Anderson, and Robert Soden. 2022. An Automated Approach to Identifying Corporate Editing. In *Proceedings of the International AAAI Conference on Web and Social Media*, Vol. 16. 1052–1063.
- [12] Nicholas Vincent and Brent Hecht. 2021. A Deeper Investigation of the Importance of Wikipedia Links to Search Engine Results. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW1 (2021), 1–15.
- [13] Nicholas Vincent, Isaac Johnson, and Brent Hecht. 2018. Examining Wikipedia with a broader lens: Quantifying the value of Wikipedia's relationships with other large-scale online communities. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. 1–13.
- [14] Nicholas Vincent, Isaac Johnson, Patrick Sheehan, and Brent Hecht. 2019. Measuring the importance of user-generated content to search engines. In *Proceedings of the International AAAI Conference on Web and Social Media*, Vol. 13. 505–516.
- [15] Charles Chuankai Zhang, Mo Houtti, C Estelle Smith, Ruoyan Kong, and Loren Terveen. 2022. Working for the Invisible Machines or Pumping Information into an Empty Void? An Exploration of Wikidata Contributors' Motivations. *Proceedings of the ACM on Human-Computer Interaction* 6, CSCW1 (2022), 1–21.