Making AI Machines Work for Humans in FoW

Sihem Amer-Yahia (CNRS, Univ. Grenoble Alpes, France), Senjuti Basu Roy (NJIT, USA), Lei Chen (HKUST, Hong Kong), Atsuyuki Morishima (Univ. of Tsukuba, Japan), James Abello Monedero (Rutgers University, USA), Pierre Bourhis (CNRS, CRIStAL, France), François Charoy (Univ. of Lorraine, INRIA, CNRS, France), Marina Danilevsky (IBM Research - Almaden, USA), Gautam Das (Univ. of Texas at Arlington, USA), Gianluca Demartini (Univ. of Queensland, Australia), Abhishek Dubey (Vanderbilt Univ., USA), Shady Elbassuoni (American Univ. of Beirut, Lebanon), David Gross-Amblard (Rennes 1 University, France), Emilie Hoareau (University Grenoble Alpes, France), Munenari Inoguchi (Univ. of Toyama, Japan), Jared Kenworthy (Univ. of Texas at Arlington, USA), Itaru Kitahara (Univ. of Tsukuba, Japan), Dongwon Lee (Pennsylvania State Univ., USA), Yunyao Li (IBM Research - Almaden, USA), Ria Mae Borromeo (UP Open Univ., Philippines), Paolo Papotti (EURECOM, France), Raghav Rao (Univ. of Texas at San Antonio, USA), Sudeepa Roy (Duke Univ., USA), Pierre Senellart (ENS, PSL University, France), Keishi Tajima (Kyoto Univ., Japan), Saravanan Thirumuruganathan (QCRI Qatar), Marion Tommasi (INRIA, France), Kazutoshi Umemoto (The Univ. of Tokyo, Japan), Andrea Wiggins (Univ. of Nebraska Omaha, USA), Koichiro Yoshida (CrowdWorks Inc., Japan)

1. OUR VISION

The Future of Work (FoW) is witnessing an evolution where AI systems (broadly machines or businesses) are used to the benefit of humans. Work here refers to all forms of paid and unpaid labor in both physical and virtual workplaces and that is enabled by AI systems. This covers crowdsourcing platforms such as Amazon Mechanical Turk, online labor marketplaces such as TaskRabbit and Qapa, but also regular jobs in physical workplaces. Bringing humans back to the frontier of FoW will increase their trust in AI systems and shift their perception to use them as a source of self-improvement, ensure better work performance, and positively shape social and economic outcomes of a society and a nation. To enable that, physical and virtual workplaces will need to capture human traits, behavior, evolving needs, and provide jobs to all. Attitudes, values, opinions regarding the processes and policies will need to be assessed and considered in the design of FoW ecosystems.

AI machines will become more specialized, more closely integrated and interoperable, and will automate many otherwise trivial tasks, as well as taking over more sophisticated functions that are currently done by humans only (e.g., onboarding and socializing). As intelligent systems are increasingly powerful and pervasive in augmenting, supporting, and sometimes replacing human work, making AI machines empower humans is necessary. This will leave workers with more time on exercising and refining human-specific skills, such as creativity and intuition and increasing the amount of specialized, highly-skilled work that they are able to handle by streamlining many supporting processes. This requires to rethink the design of FoW platforms to assist workers in continuously acquiring and improving skills through onboarding, upskilling and work delegation. Workers will take a more supervisory role, both over their work as well as the performance of AI machines that support their work, with their feedback providing corrective input that is used to continuously improve worker satisfaction and process performance.

2. INTELLECTUAL CHALLENGES

IC1: Capturing Human Capabilities.

In FoW, everyone can be a worker or an employer. Workers' perceptions of the fairness of recruitment, selection, allocation, and compensation processes will be crucial. Such perceptions must be measured to optimize not only the computational aspects of work, but also the human elements. This is a case where the measurement of key variables can be informed by social scientists and relevant theories, and put into practice by the computational communities.

New challenges at the crossroads of psychology, social science, organization studies and computational solutions will arise. These include questions such as the degree to which the variables capturing perceived fairness and transparency affect the satisfaction of workers and employers across different types of work and different platforms? Which cultural backgrounds best predict individual work metrics, and which combinations of human traits are predictors of collaborative work [4, 17]?

Addressing these questions will require adapting organizational commitment frameworks to different work contexts [1]. In particular, a major research question concerns the validation of theories from traditional workplaces in virtual marketplaces. From a modeling and computational perspectives, we need to rethink storage structures to easily update human factors, job assignment algorithms by making them adaptive, and querying capabilities to extract human capabilities over time. Additionally, as the number of human factors that are relevant to optimization are latent, subjective factors such as motivation, collaborativeness are not easy to acquire and learn. Current models of consent to tracking are all-or-nothing and there may not exist a one-size-fits-all solution. Additionally, FoW design needs to account for legal and social expectations.

IC2: Stakeholder Requirements.

FoW platforms must allow the declarative specification of job-related and workforce-related requirements. For instance, employers can only partially specify which workers to hire for their jobs (in AMT, they can specify a threshold for acceptance ratio but no other conditions, in Qapa, they can specify skills, location and qualifications but they are limited to what the proposed form lets them specify). Workers cannot specify which employers they want to ban (a recurring discussion point on TurkerNation).

Employers need to specify (i) jobs, (ii) execution requirements, such as skills, knowledge, and experiences, and (iii) delivery requirements, such as deadlines. They should also be able to express complex jobs requiring coordination among workers [14]. They need tools to estimate the available workforce on platforms and to predict how commitment and quality level they can expect from potential workers for a given job. The diversity of jobs constitutes a challenge in those estimations and predictions. Moreover, it is sometimes difficult for employers to translate their needs into concrete job specifications. It may also happen that employers obtain unexpected outputs because of some ambiguity in job descriptions [25], in which case, automatic verification using previous practices and communication channels between employers and workers, must be leveraged.

Workers should be able to specify jobs they want [20] and express expected rewards, deadlines, and required skills. They may also rely on AI machines to request which knowledge and skills they can acquire through jobs, and what sequence of jobs they could complete to further their career.

Platforms should be able to specify how to match workers and jobs and manage immoral jobs [8]. Sometimes, such jobs are decomposed into smaller ones, so that each piece does not look inappropriate, and AI algorithms for analyzing relations between jobs posted on multiple platforms are needed.

IC3: Social Processes.

Digital labor platforms change the dynamic of employer-worker and worker-worker relationships by creating an anonymous mediation between them. This weakens traditional workplace relationships.

Worker-worker and worker-employer communication constitute the social life at work. Workers exchange information and discuss job opportunities. They discuss with employers for clarification, feedback and training. Improving their ability to communicate in the workplace is essential for the success of FoW. Given that different workplaces, be they physical or virtual, have different credential systems, managing the skills portfolio of workers is a key challenge. FoW platforms should help workers not only share a CV of their work (like [19]) but also transfer their portfolio in a trustworthy manner. Additionally, onboarding for newcomers could be fully automated through AI machines or enabled via the ability to ask questions to more experienced workers. Upskilling is at least as important as onboarding. This process could be realized by AI machines that determine tutorials suitable for precision learning but also arranging job allocations in sequences of increasing difficulty. Prior work in CSCW related to onboarding has shown that, for example, retention of new Wikipedia editors is impacted by welcome messages from humans but not from bots [15].

Workers and jobs. Current platforms provide sophisticated services for task assignment but very little for other dimensions of task management like task delegation, task abandonment or team formation to complete complex tasks requiring different skills. In FoW, workers should be able to delegate part of their jobs to other workers or to AI machines join or leave teams of workers as they see fit. Incentives for interoperability is a policy issue that we do not address. This could be done through the market (as employers demand it) or through government (when major economy like EU creates regulations).

IC4: Platform Ecosystems.

Online labor markets are pervading every domain ranging from mobility (e.g. Uber), rental (e.g. Airbnb), food delivery (e.g. doordash), and freelance services (e.g. Fiverr, TaskRabbit). It is not possible for a single platform to support all these domains. Instead, due to specialized requirements there are different platforms for each domain. Within each domain and across domains, these platforms have to interoperate. That will enable different worker recruitment channels to reach diverse workers [11],[2]. At any time during job completion, AI machines should help workers if they wish to switch between tasks.

The technical challenges of interoperability include agreeing on schemas and APIs [10]. They should determine a class of interchangeable queries to exchange information on workers, employers and jobs. Such ecosystems would include (i) platforms where the actual work is performed, (ii) platforms similar to LinkedIn where workers can display their completed jobs along with credentials for skills to demonstrate their expertise, (iii) platforms for matching workers to jobs scattered across other platforms, and (iv) platforms that serve as an online watercooler where workers negotiate for employment benefits.

IC5: Computation Capabilities.

The first computational challenge is to support the design of adaptive utility functions and evaluation mechanisms, for both workers and employers, that support a variety of work types: human services, human supervision, data analysis, content creation, etc. These utility functions capture the benefit of getting involved in a platform for workers, employers and platforms by modeling preferences and constraints. AI machines must help in refining them from worker activity and feedback by leveraging methods from game theory and active learning.

One needs to aim for a global optimization in the long term, taking into account the utility functions of workers, employers, and the platform itself (these utility functions potentially evolve over time). Such optimization will be concerned with monitoring and evaluating the long-term health of the ecosystem, and especially in detecting and addressing bad actors. Employers may harm the platform by contributing fake or malicious tasks; workers may make malicious contributions, intentionally or unintentionally; and even the platform itself may be guilty of bias in work assignment or validation. Identifying such potentially harmful actions will require advancements in signal identification, outlier detection, pattern mining, and techniques for natural language understanding. As new regulations come into being to address such bad actors, they will require the availability of detailed provenance information regarding job assignment, performance evaluation, and complaints, among others.

A central focus must be placed on efficient and incremental management of the creation, storage, access, and protection of the necessary data to enable platform computation while respecting all stakeholders' privacy and well-being. This requires to monitor and mine streaming data about workers continuously and provide provenance tracking to faithfully record who produced which data, what decisions were made. This data will be leveraged in adapting AI machines to evolving human traits and needs as well as for auditing and fairness purposes.

IC6: Benchmark and Metrics.

Benchmarks and metrics for FoW will need to be developed to measure the effectiveness of humans and work interaction at various stages such as the discovery, matching, and interaction of jobs and workers. We envision benchmarks that take social and computational criteria into the metric design. The *social criteria* include social impact, capital advancement, criticality, accessibility, and robustness, while the *computational criteria* include effectiveness and efficiency. In addition, these benchmarks should be able to assess the effectiveness of human-human, human-machine, and human-job interactions.

One of the challenges is to measure human factors, such as cognitive overhead reflecting how interested workers are in their jobs, or retention which indicates whether the jobs lead to boredom and fatigue. Designed metrics may cover one or more criteria. For example, precision and recall measure effectiveness, equity measures easy and universal access to employment for a wide range of users (including users with disabilities or without access to mobile phones), and criticality evaluates whether a job is time critical.

Developing benchmarks requires understanding the context of various job marketplaces by conducting extensive surveys, and generation of synthetic datasets that correctly reflect real-world applications. Additionally, benchmarks must cover diverse applications. They need to capture subjective human factors allowing deviations and reproducibility, supporting interactions of humans with the available AI machines, and creation of *adversarial benchmarks* to evaluate the robustness of the platforms. Worker satisfaction must be assessed for continued participation of humans in the ecosystem.

IC7: Ethics.

Empowering workers and protecting their rights and privacy should be at the heart of FoW. This is a critical challenge since while platforms have a global reach, policies and regulations remain local for the most part. Advances in cybersecurity can be used to address privacy and access control mechanisms to guarantee that the right actors have visibility of the right data. Platforms should provide different privacy settings and be transparent about what worker data is exposed and to whom. Employers should be transparent about what the work is for, and how the work outcome will be utilized. They should also be able to protect their confidential information when needed. Fair compensation for workers, including base payments, bonuses, benefits and insurance should be ensured and regulated by law. Workers should have the freedom to choose the compensation type they deem acceptable. Finally, job allocation should be transparent, fair and explainable by design. Worker's sensitive attributes that might bias the job allocation process should be protected. Auditing mechanisms to ensure compliance with fair, transparent and explainable job allocation and compensation need to be developed and adopted.

In terms of fairness, an interdisciplinary approach will be required to develop novel methods to assess and quantify algorithmic fairness in job allocation practices. For example, looking at bias trade-offs between fully-algorithmic vs human-in-the-loop job allocation approaches where algorithmic bias could be different from implicit bias in humans. This will also result in higher levels of algorithmic transparency for job allocation where decisions should be easy to explain independently of whether they have been made by humans or by AI machines. Processes should be in place to specify how to best address unfair cases, e.g., by means of additional rewards for workers or novel/better job opportunities. We also envision novel methods to make job allocation distribution (i.e., the *long tail effect* where few workers complete most of the available jobs) and time spent on jobs more transparent to workers and external actors like compliancy agents. For example, visual analytics dashboards that communicate to workers how much time they spent and how much money they have earned on a platform with warnings on risks for addiction or unfair payments.

3. RELATED WORK

Kittur et al. discussed various challenges that prevent crowdsourcing from being a viable career [13]. This has inspired many follow-ups and there has been major upheaval in online labor marketplaces after [13] was published. The gig economy has become a major source of employment in various domains. Furthermore, [13] specifically focused on online paid crowdsourcing such as AMT. In contrast, our work casts a wider net. Our proposal affects not just a worker in AMT, a fully virtual marketplace, but also workers in virtual/physical labor markets such as Uber drivers and plumbers hired via TaskRabbit and Qapa.

Social Computing Positioning. Initial work [2010-2020] focused on obtaining reliable results from unreliable workers or developing algorithms for involving crowd workers on diverse tasks. Recently, there has been increasing interest in making crowdsourcing platforms a better place for both workers and requestors. A key issue in making crowdsourcing as a viable career is low pay that is often less than minimum wage in many jurisdictions. The work in [23] enables a simple way that allows a requester to pay minimum wage in AMT. IC7 discusses the issue of fairness from a wider lens beyond pay. The work in [12] surveys 360 workers and identifies the various techniques such as the usage of scripts and tools that workers use to increase their pay. IC4 discusses a working environment in the near future where AI agents act as worker surrogates to improve their experience.

Other examples in [7], [6], [5] seek to build frameworks that enable the use of crowds to solve heterogeneous tasks and optimize simultaneously for cost-quality-time. However, these are often skewed towards one stakeholder such as an employer or worker. In IC2, we identify mechanisms to obtain the requirements of all stakeholders that help in the design of equitable platforms. [22] and [19] propose alternate mechanisms for worker reputation. In IC4, we discuss a generic approach of platform ecosystems that allow a worker to seamlessly move between platforms. There has been a lot of work (e.g., [9]) on understanding the various factors affecting quality of work. Recent efforts such as [3] explore ways to improve worker's skill development through coaching while [21] discusses efficient mechanisms to teach crowd workers new skills. IC1 proposes mechanisms to capture skills (among other human factors) efficiently while IC3 talks about the challenges of upskilling. We advocate for a ma-

Data Modeling	IC1, IC2
Declarative Languages	IC1, IC2, IC5
Indexing	IC1, IC5
Data Integration	IC4
Recommendations	IC3
Data Mining	IC1
Optimization	IC3, IC4, IC5
Benchmarks	IC6
Transactions	IC4

Table 1: ICs and Data Management Challenges

jor change in how platforms are designed to enable these.

Data Management Positioning. Since FOW is more than just crowdsourcing, and much of the lower-level work will be done by AI, data management problems related to AI are a major part of our challenges [16]. Similarly, how to enable human-in-the-loop machine learning at scale and fully integrate it into business processes poses many data management challenges [24].

Table 1 summarizes core data management challenges and their relationship to our ICs. Prior works such as Deco or Qurk focused on cost based optimization for homogenous microtasks. While recent work such as Cioppino [7] generalize them to multiple heterogeneous tasks that run in parallel. SmartCrowd [18] takes human factors into account for optimization. One of the central challenges is building a FoW platform that is modular, extensible and efficient. It must be able to leverage data management techniques such as query optimization, indexing for speeding up the algorithms. Incorporating a diverse set of human and AI crowd workers requires a fundamental rethink of task assignment algorithms. Finally, it is important to develop quantitative benchmarks for each of the ICs so that the progress could be tracked.

The computational platforms necessary to support FoW will require distributed processing of transparent, and immutable time stamped records of transactional data. Blockchain technologies could enable the necessary computational artifacts to support monitor supply chains, payments processing, money transfers, reward mechanisms, digital IDs, data sharing and backup, copyrights and royalty protection, digital voting, regulations and compliance, workers rights, equity trading, management of accessible devices, secure access to belongings, etc.

As platforms become more specialized (example of CrowdFlower, a general-purpose platform that became Figure Eight, solely dedicated to data generation for AI), the trend of claiming to support one of the intellectual challenges we describe is going to increase.

4. REFERENCES

[1] P. J. Bateman, P. H. Gray, and B. S. Butler. Research note—the impact of community commitment on participation in online communities. *Information systems research*, 22(4):841–854, 2011.

[2] M. Brambilla, S. Ceri, A. Mauri, and R. Volonterio. Adaptive and interoperable crowdsourcing. *IEEE Internet Computing*, 19(5):36–44, 2015.

[3] C.-W. Chiang, A. Kasunic, and S. Savage. Crowd coach: Peer coaching for crowd workers' skill growth. *Proceedings of the ACM on Human-Computer Interaction*, 2(CSCW):1–17, 2018.

[4] L. Coursey, B. Williams, J. Kenworthy, P. Paulus, and S. Doboli. Diversity and group creativity in an online, asynchronous environment. *J. Creat. Behav*, 2017.

[5] D. Difallah, A. Checco, G. Demartini, and P. Cudré-Mauroux. Deadline-aware fair scheduling for multi-tenant crowd-powered systems. *ACM Transactions on Social Computing*, 2(1):1–29, 2019.

[6] K. Goel, S. Rajpal, and M. Mausam. Octopus: A framework for cost-quality-time optimization in crowdsourcing. In *Fifth AAAI Conference on Human Computation and Crowdsourcing*, 2017.

[7] D. Haas and M. J. Franklin. Cioppino: Multi-tenant crowd management. In *Fifth AAAI Conference on Human Computation and Crowdsourcing*, 2017.

[8] C. G. Harris. Dirty deeds done dirt cheap: a darker side to crowdsourcing. In *2011 IEEE Third International Conference on Privacy, Security, Risk and Trust and 2011 IEEE Third International Conference on Social Computing*, pages 1314–1317. IEEE, 2011.

[9] K. Hata, R. Krishna, L. Fei-Fei, and M. S. Bernstein. A glimpse far into the future: Understanding long-term crowd worker quality. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*, pages 889–901, 2017.

[10] P. G. Ipeirotis and J. J. Horton. The need for standardization in crowdsourcing. In

Proceedings of the workshop on crowdsourcing and human computation at CHI, 2011.

[11] J. Jarrett and M. B. Blake. Interoperability and scalability for worker-job matching across crowdsourcing platforms. In *2017 IEEE 26th International Conference on Enabling Technologies: Infrastructure for Collaborative Enterprises (WETICE)*, pages 3–8. IEEE, 2017.

[12] T. Kaplan, S. Saito, K. Hara, and J. P. Bigham. Striving to earn more: a survey of work strategies and tool use among crowd workers. In *Sixth AAAI Conference on Human Computation and Crowdsourcing*, 2018.

[13] A. Kittur, J. V. Nickerson, M. Bernstein, E. Gerber, A. Shaw, J. Zimmerman, M. Lease, and J. Horton. The future of crowd work. In *Proceedings of the 2013 conference on Computer supported cooperative work*, pages 1301–1318, 2013.

[14] A. Kulkarni, M. Can, and B. Hartmann. Collaboratively crowdsourcing workflows with turkomatic. In *Proceedings of the acm 2012 conference on computer supported cooperative work*, pages 1003–1012, 2012.

[15] J. T. Morgan and A. Halfaker. Evaluating the impact of the wikipedia teahouse on newcomer socialization and retention. In *Proceedings of the 14th International Symposium on Open Collaboration*, pages 1–7, 2018.

[16] N. Polyzotis, S. Roy, S. E. Whang, and M. Zinkevich. Data management challenges in production machine learning. In *Proceedings of the 2017 ACM International Conference on Management of Data*, pages 1723–1726, 2017.

[17] Y. Ren, J. Chen, and J. Riedl. The impact and evolution of group diversity in online open collaboration. *Management Science*, 62(6):1668–1686, 2016.

[18] S. B. Roy, I. Lykourentzou, S. Thirumuruganathan, S. Amer-Yahia, and G. Das. Task assignment optimization in knowledge-intensive crowdsourcing. *The VLDB Journal*, 24(4):467–491, 2015.

[19] C. Sarasua and M. Thimm. Crowd work cv: Recognition for micro work. In *International Conference on Social Informatics*, pages 429–437. Springer, 2014.

[20] T. Schulze, S. Krug, and M. Schader. Workers' task choice in crowdsourcing and human computation markets. 2012.

[21] N.-C. Wang, D. Hicks, and K. Luther. Exploring trade-offs between learning and productivity in crowdsourced history. *Proceedings of the ACM on Human-Computer Interaction*, 2(CSCW):1–24, 2018.

[22] M. E. Whiting, D. Gamage, S. N. S. Gaikwad, A. Gilbee, S. Goyal, A. Ballav, D. Majeti, N. Chhibber, A. Richmond-Fuller, F. Vargus, et al. Crowd guilds: Worker-led reputation and feedback on crowdsourcing platforms. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*, pages 1902–1913, 2017.

[23] M. E. Whiting, G. Hugh, and M. S. Bernstein. Fair work: Crowd work minimum wage with one line of code. In *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*, volume 7, pages 197–206, 2019.

[24] D. Xin, L. Ma, J. Liu, S. Macke, S. Song, and A. Parameswaran. Accelerating human-in-the-loop machine learning: challenges and opportunities. In *Proceedings of the Second Workshop on Data Management for End-To-End Machine Learning*, pages 1–4, 2018.

[25] Y. Yamakata, K. Tajima, and S. Mori. A case study on start-up of dataset construction: In case of recipe named entity corpus. In *2018 IEEE International Conference on Big Data (Big Data)*, pages 3564–3567. IEEE, 2018.