
Emerging Perspectives in Human-Centered Machine Learning

Gonzalo Ramos
Jina Suh
Soroush Ghorashi
Microsoft Research
goramos@microsoft.com
jinsuh@microsoft.com
sorgh@microsoft.com

Rebecca Fiebrink
Goldsmiths, University of London
r.fiebrink@gold.ac.uk

Gagan Bansal
University of Washington
bansalg@cs.washington.edu

Christopher Meek
Richard Banks
Saleema Amershi
Microsoft Research
meek@microsoft.com
rbanks@microsoft.com
samershi@microsoft.com

Alison Smith-Renner
University of Maryland, College Park
amsmit@cs.umd.edu

ABSTRACT

Current Machine Learning (ML) models can make predictions that are as good as or better than those made by people. The rapid adoption of this technology puts it at the forefront of systems that impact the lives of many, yet the consequences of this adoption are not fully understood. Therefore, work at the intersection of people's needs and ML systems is more relevant than ever. This area of work, dubbed Human-Centered Machine Learning (HCML), re-thinks ML research and systems in terms of human goals. HCML gathers an interdisciplinary group of HCI and ML practitioners, each bringing their unique, yet related perspectives. This one-day workshop is a successor of [6] and focuses on

CHI'19 Extended Abstracts, May 4–9, 2019, Glasgow, Scotland Uk

© 2019 Copyright held by the owner/author(s).

This is the author's version of the work. It is posted here for your personal use. Not for redistribution. The definitive Version of Record was published in *CHI Conference on Human Factors in Computing Systems Extended Abstracts (CHI'19 Extended Abstracts), May 4–9, 2019, Glasgow, Scotland Uk*, <https://doi.org/10.1145/3290607.3299014>.

recent advancements and emerging areas in HCML. We aim to discuss different perspectives on these areas and articulate a coordinated research agenda for the XXI century.

CCS CONCEPTS

• **Human-centered computing** → **Human computer interaction (HCI)**; *Interaction design*; • **Computing methodologies** → **Machine learning**;

KEYWORDS

Machine Learning; Machine Teaching; Explainable Systems; Human-Centered Machine Learning; User Experience Design, Fairness Accountability and Transparency.

ACM Reference Format:

Gonzalo Ramos, Jina Suh, Soroush Ghorashi, Christopher Meek, Richard Banks, Saleema Amershi, Rebecca Fiebrink, Alison Smith-Renner, and Gagan Bansal. 2019. Emerging Perspectives in Human-Centered Machine Learning. In *CHI Conference on Human Factors in Computing Systems Extended Abstracts (CHI'19 Extended Abstracts)*, May 4–9, 2019, Glasgow, Scotland Uk. ACM, New York, NY, USA, 8 pages. <https://doi.org/10.1145/3290607.3299014>

INTRODUCTION AND BACKGROUND

ML is a powerful technology that is at the forefront of computer science research. It is also under the spotlight of popular culture, which often misrepresents it as a Deus Ex-Machina plot device. Informally, ML allows a programmer to implement a function, not by coding, but by providing to a learning algorithm a set of examples that describes the function's behaviour. This allows for the implementations of functions, or models, that would otherwise be difficult or complex to code manually. Today, ML finds application in challenging domains such as object detection, image recognition, natural language understanding, and chatbots.

The applications of ML models can affect people in significant ways. ML systems can, for example, influence a judge's decision about a parole based on a "re-incidence score"¹; they can help a vision-impaired individual hear a description of an image post in Facebook², or take a predominant role in the practice of health care³. These are a few out of many examples of how ML affects people.

People's involvement with ML can take many roles, including data curation, learning algorithm selection and tuning, and labeling. These roles typically require specialized knowledge in data science, machine learning, and knowledge of the application's domain. These skills are often costly, and an issue for those without easy access to them. However, we believe that much more important are the emerging roles pointing at a different agency between people and ML, in particular where ML is used as a tool to assist in a larger task, or as a material at the foundation of a larger system.

These are the places where many HCI practitioners increasingly focus their attention; work that occurs at the intersection of HCI and ML, and aims to make ML (as a tool and a material) accessible

¹<https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>

²<https://www.theverge.com/2016/4/5/11364914/facebook-automatic-alt-tags-blind-visually-impaired>

³<https://www.mlforhc.org/>

and understandable to non-ML experts. In particular, the HCI community approaches research in this area by re-framing the work on and related to ML in terms of human goals [6]. This area is known as Human-Centered Machine Learning and highlights the need to take a human-centered look at how ML solutions are impacting people, how we should think about existing challenges, and how we should change the way we approach problems so that we don't face these challenges in the future. Despite seeing HCML as a topic of central importance, it has not reached a critical mass, for instance, to become a subcommittee at CHI. We aim to build on previous efforts [6] so as to widen the discussion and affirmation of HCML. HCML is a field that one workshop cannot cover. Efforts like [6] and the 2017 PAIR's symposium⁴ are examples of the work aimed at having an inclusive discussion on the topic. Our workshop differentiates from [6] in that it focuses in a subset of evolving and emerging HCML themes, so as to discuss them in more depth. In particular, we look at:

⁴<https://sites.google.com/view/pair-symposium2017/home>

Lowering the barrier for building ML models for non-experts.

End-User ML solutions lower the barrier of entry for building ML models for non-ML experts. The Crayon system, [5] introduced *Interactive Machine Learning* (IML) making model building accessible for many. In IML, there is a *human-in-the-loop* that passes knowledge to a learning system. As the system learns, the human chooses what other knowledge to pass to the system next. This knowledge can take different forms. *Sampling* is about picking a useful example of something a learning system has not seen before, and thus can learn from. *Labeling* is about providing the learning system with sources of truth. *Featuring* is about identifying or selecting the properties that improve the learning system's internal representation of the concept one wants to model. IML solutions fall outside the scope of cases where there are already a large number of labeled examples, as those can be better served with traditional batch ML pipelines.

Among ideas in the IML space, Machine Teaching (MT) [13] is an example that stands out as an emerging perspective. Here we use the term to refer to an IML process that builds on **the notion that humans are more than “a source of labels”** [2]. This process focuses on facilitating the interactive extraction of knowledge from a human teacher, towards the building of a model, while not knowing details about the learning algorithm under the hood. This process can also allow for the decomposition of a concept into smaller sub-concepts as another form of human knowledge teachers can pass to a learning system. It also encourages the definition, refinement, and reuse of semantic features. This makes models created in this way semantic by design; i.e., the decisions made by these models are explainable in human terms. These models are also easier to debug and maintain, as it is tractable why a prediction is incorrect: either the model is blind to a feature the teacher did not express, or the model needs to train with a new label. Finally, MT can be seen under the lens of programming: a teacher uses a teaching language (of labels, features, examples, decomposition) to code a particular concept. This view allows researchers to consider existing lessons and principles from the software

⁵<https://en.wikipedia.org/wiki/SOLID>

⁶<http://www.luis.ai>

engineering discipline, such as SOLID.⁵ Systems using the MT teaching process are already in use in the wild. With the LUIS service⁶, teachers provide examples, features, and labels interactively and progressively, without any predefined training or test sets. The models they create are always being improved, and the features teachers provide are not meant to fit the already labeled data, but instead given with the goal of the model being accurate on yet unseen data.

Lowering the barrier of entry to ML use remains an active research topic. Defining the languages that allow end-users, or teachers, to express knowledge about different types of data such as text, or images is important. These languages should not only enable someone to express labels and features; it must also allow someone to decompose a concept, as well as query the set of unlabeled examples to find the best ones at a given time. This language should let an end-user articulate why an example does or doesn't have a particular label. A core aspect of any IML loop is that of a teacher knowing about the effects their actions have on a model. While metrics to track how a model is performing, such as accuracy, precision or recall, are straightforward, they are only as useful and effective as the access one has to test and validation sets. For problems where these sets are not available, measuring the progress on an IML flow is not as straightforward. This challenge is amplified by the fact that traditional metrics say nothing about how reusable a concept or feature is, or the amount of new domain knowledge a teacher acquired in the process. We need a framework to think about how to measure IML in different dimensions. Last but not least, it is also worthwhile to discuss what are the foreseeable positive and negative consequences of democratizing ML.

Intelligible systems, explainable decisions, and algorithmic fairness

Systems that produce predictions that can have significant consequences to people are becoming ubiquitous. [10] sounds the alarm on the dangers of having such technologies deeply embedded into the fabric of everyday life. It becomes paramount then the need to add transparency, from the point of view of a human observer, to either the decisions (being explainable, the *why*) or the processes that generate them (being intelligible, the *how*) [7] [4] [8]. The same need also causes the research community to take a hard look at the “ethical, moral, social, and policy implications of social-technical systems”.⁷ These areas of research are highly interdisciplinary bringing to its fold practitioners not only from systems, HCI and ML, but also from topics such as law, social science, philosophy, and psychology. [1] highlights the many sources informing research on these issues.

⁷<https://fatconference.org/>

In particular, the implementation of the European Union's General Data Protection Regulation (GDPR) and its provision to contest any automated decision-making that was made on an algorithmic basis, brings the discussion about explainable and intelligible ML systems to the forefront of HCML research. Either through the creation and study of models that are explainable and/or intelligible [3, 9], or by the development of approaches that explain “black box” systems [11, 12], this perspective is only starting to shed light into the important questions we need to answer as a community to

make ML fair and trustworthy. It is also crucial to think and investigate how these issues affect the consumption of ML systems by end-users. For example, should a system reveal when bias exists in the data or the algorithmic decision and present end-users with action items, if any? Or what are the thresholds of unfairness, under different definitions, that end-users find acceptable? Does fairness, bias, accountability need to be part of the ML workflow? If so, how?

Workshop Goals

To summarize, Machine Learning has developed as a powerful technology that is embedded into the fabric of the systems and processes that affect millions of people. Human-Centered Machine Learning emerges in response to the above, to provide the understandings and means to put ML at the service of people in a way that is accessible, useful, and trustworthy to all. It is within this context that we set our workshop's goals to continue past efforts in the area and articulate an updated agenda for doing human-centered research that focuses on the above emerging perspectives. In particular, we hope to bring together a community of diverse practitioners and ideas connected by a common set of goals.

ORGANIZERS

Saleema Amershi is a Researcher in the Adaptive Systems and Interaction Group at Microsoft Research working on technologies for making people better at building and using ML systems. Throughout her work, she distills guiding principles applicable in a broader context to help provide a foundation for future human-driven machine learning systems.

Richard Banks is Principal Design Manager at the Human Experience and Design Group at Microsoft Research, Cambridge, UK. His work explores the boundaries of technology, creativity, and society through a human-centered lens. Much of Richard's current focus is on new issues and opportunities raised by artificial intelligence for both people, as well as the design profession.

Gagan Bansal is a graduate student at the Paul G. Allen School of Computer Science at the University of Washington, Seattle. He conducts interdisciplinary research on AI and HCI. He currently works on developing UI to explain ML models to users via a user-system dialog, and interfaces to evaluate the effect of explanations on human-AI team performance. He is advised by Dan Weld.

Rebecca Fiebrink is a Senior Lecturer at Goldsmiths, University of London. Much of her research involves combining techniques from HCI and ML to support new interactive technologies for music, art, and accessibility. She is also involved in research on ML education. She is the author of the Wekinator software for interactive ML.

Soroush Ghorashi is a Research Engineer at Microsoft Research working in the Machine Teaching Group, where he is trying to streamline the process of knowledge transfer from the human teachers to the learning machines. He received his M.Sc and Ph.D. from the Oregon State University specializing in

HCI focused on improving collaboration quality within software teams by building novel development environment and collaboration techniques.

Christopher Meek is a Principal Researcher at Microsoft Research in the Machine Teaching Group, where he is enabling people to effectively use machine learning to solve problems. He received his Ph.D. from Carnegie Mellon University, and worked on topics related to machine learning, many of which have led to product innovations and product features. He has a long-standing interest in learning causal, and graphical models. He is an affiliate professor at the University of Washington.

Gonzalo Ramos is a Senior Researcher at Microsoft Research working in the Machine Teaching Group, where he works in lowering the barrier of entry for people to harness the power of ML. He received his M.Sc and Ph.D. from the University of Toronto's Computer Science Department, specializing in Scientific Visualization and HCI, respectively. Gonzalo also worked as a Senior Design Technologist and UX Scientist at Amazon, and as a Scientist at Microsoft.

Alison Smith-Renner leads the Machine Learning Visualization Lab for Decisive Analytics Corporation. Her focus is on enhancing users' understanding and analysis of data without requiring expertise in ML or data science. She is currently a Computer Science Ph.D. Candidate at the University of Maryland, College Park with a focus on human-centered design for interactive machine learning under the direction of Leah Findlater and Jordan Boyd-Graber.

Jina Suh is a graduate student at the Paul G. Allen School of Computer Science at the University of Washington advised by Dan Weld and James Fogarty. She is a Principal Research Software Developer at Microsoft Research in the Affective Computing Group. Jina researches how humans interact with ML systems focusing on model intelligibility and bias. She was part of the Machine Teaching Group at Microsoft Research and continues to contribute in this area.

WEBSITE

We will host a website (<https://aka.ms/hcmlperspectives>) to advertise and disseminate the workshop's information and call for proposals. We will advertise the workshop through email distribution lists at relevant research institutions, direct communication with colleagues in the field, and social media.

PRE-WORKSHOP PLANS

We will expect and encourage participation from practitioners working at the intersection of HCI and ML as well as individuals interested in getting started in HCML. Participants interested in giving a presentation at the workshop need to submit a short paper between 3 and 6 pages long. We encourage position papers that present diverse points of view on each of the aforementioned HCML perspectives. We are most interested in having presentations that generate a constructive discussion, or present a compelling blueprint or vision of the future. We will select papers through a review process by at least two workshop committee members.

WORKSHOP STRUCTURE

This one-day workshop will begin with a short opening session where the organizers will frame the overall themes and give a preview of the sessions ahead. We will then divide the workshop into 4 sessions lasting between 60 and 90 minutes. The first 3 sessions will focus on a theme aligning with lowering the barrier of entry to creating ML models, Machine Teaching, and Explainable and fair ML, respectively. Each session will consist of 3 to 4 presentations chaired by a member of the workshop's committee. Each presentation will take about 10 minutes. Each session will give participants and attendees enough time for a meaningful discussion, moderated by the session's chair. The fourth session will consist of a participatory design activity including the audience. Finally, we will close the workshop with a general discussion to summarize the thoughts presented by participants and proposing future action items.

POST-WORKSHOP PLAN

At the final session, we will discuss how to disseminate the materials presented during the workshop; e.g., research questions and positions, design principles or tenets. At a minimum, we will build a lasting website linking to archival versions of all position papers, as well as video recordings of all sessions. We hope that the discussions generated during this workshop will fuel an expanded position paper, or series of articles to be published on special issues of peer-reviewed journals such as TOCHI, similarly to how [6] lead to [14]. We will also discuss how to continue building the HCML community through social media and/or a topical SIG. We will also explore the possibility of follow-up events in HCI, design and ML venues in order to increase the size and awareness of the HCML community.

CALL FOR PARTICIPATION

Machine Learning is currently one of the most important technologies used across systems affecting research, commerce, healthcare, entertainment, security, etc. As its use grows woven into our society, and affecting the lives of people, it has become important to study and design the interactions between ML and its stakeholders from a Human-Centered point of view. Human-Centered Machine Learning (HCML) shares this view and re-frames work on and related to ML in term of human goals. This workshop continues prior workshops on HCML by inviting ideas focusing on, but not limited to the following themes from an HCML point of view:

- Lowering the barrier for building ML solutions,
- People engaging with ML in richer ways than just providing labels, and
- Intelligible systems, explainable decisions, and fair ML.

By articulating these emerging perspectives, we look at how, as a community, we keep moving the HCML field forward. We invite the submission of positions papers between 3-6 pages long. Position

papers should follow the CHI Extended Abstract format and be submitted through the workshop's submission site available through <https://aka.ms/hcmlperspectives>. The workshop's organizing committee will review the submissions.

Accepted papers will be presented at the workshop. We ask that at least one of the authors of each accepted position paper attends the workshop. Presenting authors must register for the workshop and at least one full day of the conference. Each presentation will take place within a session focused around a particular theme. Sessions will consist of 3-4 presentations, each lasting approximately 10 minutes and will be followed by a group discussion.

REFERENCES

- [1] Ashraf Abdul, Jo Vermeulen, Danding Wang, Brian Y. Lim, and Mohan Kankanhalli. 2018. Trends and Trajectories for Explainable, Accountable and Intelligent Systems: An HCI Research Agenda. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. ACM, New York, NY, USA, Article 582, 18 pages.
- [2] Saleema Amershi, Maya Cakmak, William Bradley Knox, and Todd Kulesza. 2014. Power to the people: The Role of Humans in Interactive Machine Learning. *AI Magazine* 35, 4 (2014), 105–120.
- [3] R. Caruana, Y. Lou, J. Gehrke, P. Koch, M. Sturm, and N. Elhadad. 2015. Intelligent Models for Healthcare: Predicting Pneumonia Risk and Hospital 30-day Readmission.
- [4] Finale Doshi-Velez, Mason Kortz, Ryan Budish, Chris Bavitz, Sam Gershman, David O'Brien, Stuart Schieber, James Waldo, David Weinberger, and Alexandra Wood. 2017. Accountability of AI Under the Law: The Role of Explanation. *CoRR* abs/1711.01134 (2017). arXiv:1711.01134
- [5] Jerry Alan Fails and Dan R. Olsen, Jr. 2003. Interactive Machine Learning. In *Proceedings of the 8th International Conference on Intelligent User Interfaces (IUI '03)*. ACM, New York, NY, USA, 39–45.
- [6] Marco Gillies, Rebecca Fiebrink, Atsu Tanaka, Jérémie Garcia, Frédéric Bevilacqua, Alexis Heloir, Fabrizio Nunnari, Wendy Mackay, Saleema Amershi, Bongshin Lee, Nicolas d'Alessandro, Joëlle Tilmanne, Todd Kulesza, and Baptiste Caramiaux. 2016. Human-Centred Machine Learning. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems (CHI EA '16)*. ACM, New York, NY, USA, 3558–3565.
- [7] Bryce Goodman and Seth Flaxman. 2017. European Union Regulations on Algorithmic Decision-Making and a "Right to Explanation". *AI Magazine* 38, 3 (2017), 50–57.
- [8] Todd Kulesza, Margaret Burnett, Weng-Keen Wong, and Simone Stumpf. 2015. Principles of Explanatory Debugging to Personalize Interactive Machine Learning. In *Proceedings of the 20th international conference on intelligent user interfaces*. ACM, 126–137.
- [9] Himabindu Lakkaraju, Ece Kamar, Rich Caruana, and Jure Leskovec. 2017. Interpretable & Explorable Approximations of Black Box Models.
- [10] C. O'Neil. 2016. *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. Crown. <https://books.google.com/books?id=CxD-DAAAQBAJ>
- [11] M. Ribeiro, S. Singh, and C. Guestrin. 2016. Why Should I Trust You?: Explaining the Predictions of any Classifier.
- [12] M. Ribeiro, S. Singh, and C. Guestrin. 2018. Anchors: High-Precision Model-Agnostic Explanations.
- [13] Patrice Y Simard, Saleema Amershi, David M Chickering, Alicia Edelman Pelton, Soroush Ghorashi, Christopher Meek, Gonzalo Ramos, Jina Suh, Johan Verwey, Mo Wang, et al. 2017. Machine Teaching: A New Paradigm for Building Machine Learning Systems. *arXiv preprint arXiv:1707.06742* (2017).
- [14] Michelle Zhou (Ed.). 2018. *ACM Trans. Interact. Intell. Syst.* 8, 2 (2018).