# Cultivating Human Expertise Through AI-Assisted Data Science

Josh Andres, Christine Wolf, Michael Muller, Justin Weisz, Narendra Nath Joshi, Aabhas Sharma, Krissy Brimijoin, Michael Desmond, Zahra Ashktorab, Qian Pan, Evelyn Duesterwald & Casey Dugan*
josh.andres@au1.ibm.com
IBM Research
Melbourne, Australia

## ABSTRACT

Advances in AI-systems are giving way to an era of AI-assisted human in the loop data science, where the data worker and the AI-system, work together to create better models faster. The interactions of AI-system and data worker offer a design space for cultivating data worker expertise. We propose opportunities for human in the loop data science focused on two data science worker roles: data science engineering and data labelling.

## KEYWORDS

Human-AI collaboration, Data Science, Expertise

**ACM Reference Format:**
Josh Andres, Christine Wolf, Michael Muller, Justin Weisz, Narendra Nath Joshi, Aabhas Sharma, Krissy Brimijoin, Michael Desmond, Zahra Ashktorab, Qian Pan, Evelyn Duesterwald & Casey Dugan. 2020. Cultivating Human Expertise Through AI-Assisted Data Science. In *DaSH@KDD'20, August 24, 2020, Virtual Conference* . ACM, New York, NY, USA, 2 pages.

## 1 INTRODUCTION

Advances in artificial intelligence (AI) and machine learning (ML) in recent years have once again sparked widespread interest in developing systems that harness these powerful technologies to transform everyday work practices in a number of domains through automation. Efforts to leverage AI/ML in data science work practices can be found under several monikers, such as "human-centered data science," [3, 10, 14], "human-in-the-loop machine learning" [5, 11], or "interactive machine learning" [1]. Examples of systems that enhance human efforts in the data science pipeline include label suggestions, enabling data labelers to perform their job faster [17], and algorithmic recommendations offered to data science engineers as they explore alternatives and build models for production [19].

At the same time, a growing body of empirical studies into data science work point to the situated and nuanced knowledge practices that constitute data science [3, 10, 15, 20, 22]. Data science work requires care [21]. Today, many interactive and automated human-in-the-loop paradigms focus on automating parts of the data science work to gain time efficiencies to make it less tedious to a broader range of actors varying in expertise levels [18].

While automation can provide efficiencies, our field needs safeguards to avoid data science workers over-relying on automated suggestions and result in losing critical engagement for expertise development [7]. In this extended abstract, we investigate the automated, human-in-the-loop future of AI-assisted data science. In particular, we propose opportunities to use automation not only to gain efficiencies but to also consider expertise development as part of the data science worker and automation system interactions. As yet, the consequences of data science automation remain open questions. *How might we design for mixed-initiative data science tasks, where the human and AI-system maximize their complementary abilities while supporting the human in developing expertise?*

We explore this question by drawing from our experiences in studying and designing various human-in-the-loop data science projects at a large, global technology organization. Our goal in reflecting on this opportunity is to provoke discussion at the KDD'20 DaSH Workshop on how we might cultivate human expertise through the human-in-the-loop systems we design.

## 2 HUMAN-IN-THE-LOOP DATA SCIENCE

We now describe two roles in the data science pipeline - data science engineering and data labeling - and discuss changes and opportunities in each on the topic of designing for human expertise.

### 2.1 AI-Assisted Data Science Engineering

Data science teams work together to extract insights from data through the orchestration of ML models [9, 15]. HCI research supporting data science work has focused on creating user interfaces to allow data scientists to data wrangle [6] and structurally study the anatomy and troubleshoot ML models to make sense of the model's performance [16]. Recent efforts have focused on automating pre-processing steps, such as algorithm selection, feature engineering, and hyper-parameter optimization, in multiple projects and services known collectively as "AutoML" [18], to reduce the time that data science teams spend on these pre-processing steps.

Automation like this does free up the attention of data science engineers; yet it may also impose built-in assumptions on the team's work, and may also mean that the knowledge work of reflecting on *why* one algorithm should be selected over another, is pulled away from the human and instead offloaded to the AI-system. This raises interesting design issues for HCI around human-AI agencies [7]. How might we design to support data science workers to challenge assumptions that are built-in to their tools? How might we design to

Permission to make digital or hard copies of all or part 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 components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
*DaSH@KDD'20, August 24, 2020, Virtual Conference*
© 2020 Association for Computing Machinery.

support exploration, learning, and in turn enrichment of their expertise? Such questions require us to consider issues of transparency [4] and explainability [2, 4, 8, 12], in AI-assisted data science. Flexibility is needed in human-in-the-loop systems to limit situations where data science teams develop "knowledge gaps" about what is being automated and how this step fits in the larger picture of work practices and work products. Not all data science workers have the same level of expertise, and future automation interactions should consider their diversity. We envision these questions to encourage reflection on the practices that our human-in-the-loop systems are enabling. In addition to metrics such as efficiency and accuracy, cultivating human expertise should also be a concern in designing enjoyable and enriching human-in-the-loop data science systems.

## 2.2 AI-Assisted Data Labelling

Data labelers play an important role in preparing ground-truth data for training ML models [17]. In enterprise settings, this work is often done by subject-matter experts (SMEs). Our current project focuses on SME labeling. In our interviews with 25 labelers and labeling-task creators, we found that there is limited automation functionality currently available to assist labeling work. One reason for this is likely the spectrum and complexity of domain labelling tasks, for example, from identifying obligations in legal contracts to pathologies in medical imaging [13]. This complexity is a pain point as it often results in the costly development of specialised UI's as there is not a generic user interface for cross-domain labelling.

The use of AI-assisted automation in the context of creating the labelling interface that supports labelling and the nuances of human-human and human-system protocols during data labelling could focus on knowledge building. For example, the labeling-task creator could ask the AI-system to bootstrap an interface for labelling retinal pathologies in images, and the system could follow up with questions informed by what is commonly labelled in retinal images, "labelling retinal images for Neovascular glaucoma can involve tracing veins and arteries, would you like tracing functionality?". In the labeller scenario, the AI-assisted automation could present batches of items that are likely to have the same label with an accompanying explanation of how these items relate to contributing to the labeller's understanding of the displayed items. This knowledge building framing can enrich the practice of data labelling to promote the development of expertise.

# 3 CONCLUSION

In this short paper, we investigate the automated, human-in-the-loop future of AI-assisted data science by reflecting on the changing roles of data science engineers the data labellers. From the perspective of using automation beyond the purpose of gaining efficiencies and instead consider data science worker and AI-system automation interaction as opportunities to support expertise development.

# REFERENCES


[1] 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.

[2] Josh Andres, Christine T Wolf, Sergio Cabrero Barros, Erick Oduor, Rahul Nair, Alexander Kjærum, Anders Bech Tharsgaard, and Bo Schwartz Madsen. 2020. Scenario-based XAI for Humanitarian Aid Forecasting. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems Extended Abstracts*. 1–8.

[3] Cecilia Aragon, Clayton Hutto, Andy Echenique, Brittany Fiore-Gartland, Yun Huang, Jinyoung Kim, Gina Neff, Wanli Xing, and Joseph Bayer. 2016. Developing a research agenda for human-centered data science. In *Proceedings of the 19th ACM Conference on Computer Supported Cooperative Work and Social Computing Companion*. 529–535.

[4] Jaimie Drozdal, Justin Weisz, Dakuo Wang, Gaurav Dass, Bingsheng Yao, Changruo Zhao, Michael Muller, Lin Ju, and Hui Su. 2020. Trust in AutoML: exploring information needs for establishing trust in automated machine learning systems. In *Proceedings of the 25th International Conference on Intelligent User Interfaces*. 297–307.

[5] Yolanda Gil, James Honaker, Shikhar Gupta, Yibo Ma, Vito D'Orazio, Daniel Garijo, Shruti Gadewar, Qifan Yang, and Neda Jahanshad. 2019. Towards human-guided machine learning. In *Proceedings of the 24th International Conference on Intelligent User Interfaces*. 614–624.

[6] Philip J Guo, Sean Kandel, Joseph M Hellerstein, and Jeffrey Heer. 2011. Proactive wrangling: Mixed-initiative end-user programming of data transformation scripts. In *Proceedings of the 24th annual ACM symposium on User interface software and technology*. 65–74.

[7] Jeffrey Heer. 2019. Agency plus automation: Designing artificial intelligence into interactive systems. *Proceedings of the National Academy of Sciences* 116, 6 (2019), 1844–1850.

[8] Robert R Hoffman, Shane T Mueller, Gary Klein, and Jordan Litman. 2018. Metrics for explainable ai: Challenges and prospects. *arXiv preprint arXiv:1812.04608* (2018).

[9] Miryung Kim, Thomas Zimmermann, Robert DeLine, and Andrew Begel. 2016. The emerging role of data scientists on software development teams. In *2016 IEEE/ACM 38th International Conference on Software Engineering (ICSE)*. IEEE, 96–107.

[10] Marina Kogan, Aaron Halfaker, Shion Guha, Cecilia Aragon, Michael Muller, and Stuart Geiger. 2020. Mapping Out Human-Centered Data Science: Methods, Approaches, and Best Practices. In *Companion of the 2020 ACM International Conference on Supporting Group Work*. 151–156.

[11] Doris Jung-Lin Lee, Stephen Macke, Doris Xin, Angela Lee, Silu Huang, and Aditya Parameswaran. 2019. A Human-in-the-loop Perspective on AutoML: Milestones and the Road Ahead. *Data Engineering* 58 (2019).

[12] Q. Vera Liao, Daniel Gruen, and Sarah Miller. 2020. Questioning the AI: Informing Design Practices for Explainable AI User Experiences. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. 1–15. https://doi.org/10.1145/3313831.3376590

[13] Ruibin Ma, Po-Hsuan Cameron Chen, Gang Li, Wei-Hung Weng, Angela Lin, Krishna Gadepalli, and Yuannan Cai. 2019. Human-centric Metric for Accelerating Pathology Reports Annotation. *arXiv preprint arXiv:1911.01226* (2019).

[14] Michael Muller, Melanie Feinberg, Timothy George, Steven J Jackson, Bonnie E John, Mary Beth Kery, and Samir Passi. 2019. Human-Centered Study of Data Science Work Practices. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–8.

[15] Michael Muller, Ingrid Lange, Dakuo Wang, David Piorkowski, Jason Tsay, Q Vera Liao, Casey Dugan, and Thomas Erickson. 2019. How Data Science Workers Work with Data: Discovery, Capture, Curation, Design, Creation. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–15.

[16] Jeffrey M Perkel. 2018. Why Jupyter is data scientists' computational notebook of choice. *Nature* 563, 7732 (2018), 145–147.

[17] Szilárd Vajda, Yves Rangoni, and Hubert Cecotti. 2015. Semi-automatic ground truth generation using unsupervised clustering and limited manual labeling: Application to handwritten character recognition. *Pattern recognition letters* 58 (2015), 23–28.

[18] Dakuo Wang, Parikshit Ram, Daniel Karl I Weidele, Sijia Liu, Michael Muller, Justin D Weisz, Abel Valente, Arunima Chaudhary, Dustin Torres, Horst Samulowitz, et al. 2020. AutoAI: Automating the End-to-End AI Lifecycle with Humans-in-the-Loop. In *Proceedings of the 25th International Conference on Intelligent User Interfaces Companion*. 77–78.

[19] Dakuo Wang, Justin D. Weisz, Michael Muller, Parikshit Ram, Werner Geyer, Casey Dugan, Yla Tausczik, Horst Samulowitz, and Alexander Gray. 2019. Human-AI Collaboration in Data Science. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (Nov 2019), 1–24. https://doi.org/10.1145/3359313

[20] Christine Wolf and Drew Paine. 2020. Sensemaking Practices in the Everyday Work of AI/ML Software Engineering. (2020).

[21] Christine T Wolf. 2019. Conceptualizing Care in the Everyday Work Practices of Machine Learning Developers. In *Companion Publication of the 2019 on Designing Interactive Systems Conference 2019 Companion*. 331–335.

[22] Christine T Wolf. 2020. AI Models and Their Worlds: Investigating Data-Driven, AI/ML Ecosystems Through a Work Practices Lens. In *International Conference on Information*. Springer, 651–664.