The Next Decade of Data Science Practice

Justin D. Weisz
jweisz@us.ibm.com
IBM Research AI

Michael Muller
michael_muller@us.ibm.com
IBM Research AI

ABSTRACT

We present a vision of how the practice of data science will evolve over the next decade due to advancements in automated machine learning (known as AutoML or AutoAI). Akin to how certain other industries experienced a *shift in*, rather than an *elimination of*, jobs when automation technologies were introduced, data science work will experience a similar shift. Based on our experiences in building and studying user interfaces for automated machine learning tools, we propose seven new specializations of the future data scientist: translators, label-wranglers, interpreters, referees, operators, hackers, and auditors. These roles will offer new opportunities to data scientists, their organizations, and people entering the field.

CCS CONCEPTS

• **Social and professional topics** → **Automation**; • **Computing methodologies** → *Machine learning*; • **Human-centered computing** → **HCI theory, concepts and models**.

KEYWORDS

Data science, human-centered data science, AutoAI, AutoML

ACM Reference Format:
Justin D. Weisz and Michael Muller. 2020. The Next Decade of Data Science Practice. In *Proceedings of DaSH@KDD*. ACM, New York, NY, USA, 3 pages.

1 INTRODUCTION

New technologies in data science, known as AutoML [8] and AutoAI [25], automatically perform work such as data cleaning, feature creation, model building, and pipeline evaluation. These technologies threaten to deskill data scientists by eliminating the need to manually perform these activities [26]. We ask: *will these technologies spell the end for the practice of data science as we know it?*

When we examine how automation has historically affected jobs in other industries, we find some stories that involve a paradoxical combination of job-content improvement and overall job losses [16, 28], and other stories that are more about complementarity and job growth rather than displacement [3]. We believe that AutoAI and AutoML are in the latter category.

One popular example of automation is in the automotive industry. In the 1970s, many factories began adopting robotic technologies to automate dangerous manual tasks such as heavy lifting and welding that were previously conducted by human workers [24]. Automating these tasks resulted in huge improvements to worker safety [24] and shifted the kinds of tasks performed by human workers away from low-skill tasks toward higher-skill (and higher-paid) tasks, including the maintenance of the robots themselves and dealing with anomalies in their performance. While the robots saved humans from injury by doing the more hazardous work, they also displaced workers from those jobs.

By contrast, a second example occurs in the aviation industry, in which automation and "fly-by-wire" technologies made similar threats to the skills of pilots, who no longer needed to spend the majority of their time flying a plane under normal circumstances [10]. In this case, however, the work of pilots shifted toward intensely training for emergency scenarios, and commercial flights today still require human pilots in the cockpit. A similar effect occurred in banking, in which the introduction of automated teller machines led to a net *increase* in the number of jobs for human bank tellers [3].

For human work in data science [14], we anticipate similar shifts away from low-level, mechanistic ways of feature engineering, model building, and hyperparameter tuning (all of which are currently well-automated), toward higher-level reasoning about business goals and performance requirements of the resulting models.

2 FUTURE DATA SCIENCE SPECIALIZATIONS

We envision the future of data science work through a series of roles that highlight the different areas in which future data scientists will specialize. These roles are an outgrowth of our work examining current data science practices [17, 29] and adoption and use of automated machine learning technologies [6, 25–27]. We note that these roles imply a broadened view of the skills that make up data science practice, drawing from other disciplines such as XAI, systems engineering, security, DevOps, and behavioral science.

2.1 Translators

Much of data science work involves translating between the language of *business* and the language of *data*. Data scientists often mediate between business leaders who desire to define a set of important metrics or KPIs, and the underlying data and models from which those measures are computed. For example, a data scientist may translate a business requirement to predict the demand for a product in the next month into a predictive model trained on historical sales data. We anticipate a continued, if not increased need, for data scientists to act in this mediating role. Improvements in automation may provide efficiency gains in conducting the actual work of *translation*, but data scientists will always need to be involved in turning qualitative business desires and goals into a quantitative set of data to be collected and metrics to be computed.

2.2 Label Wranglers

Discriminative models require a set of features by which to classify an instance, and ground-truth labels for those instances. In many cases, domain experts or SMEs are required to provide input on

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).
DaSH@KDD, August 24, 2020, Virtual Conference
© 2020 Copyright held by the owner/author(s).

what those features should be, and labels for a set of instances in order to train a model. Several techniques exist that ease the burden on these labeling and feature engineering tasks. Zhao et al. [30] describe a technique for propagating labels from a small subset of instances to other instances in the larger data set. Khurana et al. [13] developed a way to automatically collect domain knowledge from web searches as part of feature engineering. Active learning [21] selects instances to label that lie along the current decision boundary, forcing the classifier to converge faster. Despite these improvements in the efficiency of the labeling process, human effort and oversight will still be required in reviewing automatically engineered features for their relevance to the business problem, and in providing ground-truth labels to bootstrap supervised (or semi-supervised) discriminative algorithms.

2.3 Interpreters

The long-term goal of Explainable AI (XAI) is to provide fluent and understandable explanations of how AI algorithms make decisions. Many algorithms exist for producing explanations for different kinds of AI models [2], and we expect these off-the-shelf approaches to be useful for low-stakes scenarios, such as product recommenders. However, for higher-stakes applications[1], societies will require explanations not only for how AI models *operate* and *make decisions*, but also for how they were *created* and *evaluated* [1, 12, 22]. Thus, the model interpreter will need to apply data science forensic skills to help creators, users, and other stakeholders answer important questions about their production and operation:

- How the model was **produced**: what data was used to train the model, how was that data de-biased?
- How the model was **evaluated**: how accurate is the model on test data, how accurate is the model expected to be in a production environment, does the model produce fair outcomes?
- How the model **operates**: how does the model make decisions in general (e.g. global explanations), how did the model make a decision in a particular instance (e.g. local explanations), what are the performance characteristics of the model?
- Why the model is **trustworthy**: what governance processes were used to protect the integrity of the model during its development, how robust is the model to adversarial attack?

Techniques from XAI [2] and adversarial robustness testing [20] will be helpful in answering some of these questions, but more work needs to be done to capture and document the development of an AI model throughout its lifecycle [1, 12] in order for a model interpreter to address questions of trustworthiness.

2.4 Referees

Recently, many researchers, policy makers, and social justice advocates have focused their attention on the issue of *fairness* in AI systems. Many metrics and techniques have been developed to quantify and evaluate biases present in data sets, models, and outcomes, and to mitigate those biases [4]. Frustratingly, there are currently 21 different mathematical formulations of fairness [19, 23], some of which are wholly incompatible with each other. How is a data scientist to know how to use these mitigation techniques and metrics to evaluate their system in a way that translates to higher-level business goals or ethical principles?

We envision a role of a referee who helps intermediate between culturally-situated *social* definitions of fairness (e.g., [7]) and different *computational* formulations of fairness. This role would involve negotiating and operationalizing fairness criteria across the complex set of social aspirations and legal requirements for fairness that exist across society, as well as ensure that the resultant AI system conforms to those requirements.

2.5 Operators

Once a machine learning model has been trained and tested, it will be up to an operator to consider many issues when deploying that model in a production environment: what kinds of guardrails should be established to ensure the model continues to perform accurately while preserving desired levels of fairness, and who is responsible to act if the guardrails should fail? When a model's performance drifts, how should new training data be acquired to improve the model's performance? How can adversarial attacks against an AI service be detected and mitigated? Although we anticipate the actual monitoring of a model's real-time performance to be automated, business decisions will ultimately determine how to bound a model's performance drift and weigh the cost of acquiring additional training data against expected business value.

2.6 Hackers

As AI systems perform increasingly important work for society, they will become attractive sites for hacking. In software engineering, the defensive practice of *ethical hacking* has become so important that one can earn certification in that field [9]. We anticipate a similar role in which certified, ethical AI hackers use a multitude of techniques to identify weaknesses in an AI model. For example, work by Li et al. [15] discusses a blackbox method for fooling an image classifier with a 100% success rate. Tools that help ethical AI hackers ensure the adversarial robustness of their models are beginning to emerge [20], and we believe this role will be critically important for the future of data science.

2.7 Auditors

To establish trust in AI systems, we envision a need for independent auditors who would use their technical and reputational authority to certify and strengthen crucial properties of an AI system, such as performance, fairness, robustness, and ethical application. The auditor would examine all aspects of a system's development, validating not just the actual AI system, but also the people who designed and built the system. The auditor is a meta-role that would draw upon the skills of all the roles previously discussed in order to certify the trustworthiness of an AI system and provide a "seal of approval."

3 CONCLUSION

Based on our experience in building and studying automated machine learning technologies, we view these technologies as *evolving enablers* [3]: they will enable data scientists to focus on higher-value work by automating low-level tasks; they will enable organizations to expand the range of data science jobs; and they will enable aspiring data scientists to join the field by lowering barriers of entry.

[1]Examples of higher-stakes applications that incorporate AI models include prison sentence recommendation [11], credit decisions [18], and hiring [5].

REFERENCES

[1] Matthew Arnold, Rachel KE Bellamy, Michael Hind, Stephanie Houde, Sameep Mehta, A Mojsilović, Ravi Nair, K Natesan Ramamurthy, Alexandra Olteanu, David Piorkowski, et al. 2019. FactSheets: Increasing trust in AI services through supplier's declarations of conformity. *IBM Journal of Research and Development* 63, 4/5 (2019), 6–1.

[2] Vijay Arya, Rachel KE Bellamy, Pin-Yu Chen, Amit Dhurandhar, Michael Hind, Samuel C Hoffman, Stephanie Houde, Q Vera Liao, Ronny Luss, Aleksandra Mojsilović, et al. 2019. One explanation does not fit all: A toolkit and taxonomy of ai explainability techniques. *arXiv preprint arXiv:1909.03012* (2019).

[3] David H. Autor. [n.d.]. Why are there still so many jobs? The history and future of workplace automation. ([n. d.]).

[4] Rachel KE Bellamy, Kuntal Dey, Michael Hind, Samuel C Hoffman, Stephanie Houde, Kalapriya Kannan, Pranay Lohia, Jacquelyn Martino, Sameep Mehta, Aleksandra Mojsilovic, et al. 2018. AI Fairness 360: An extensible toolkit for detecting, understanding, and mitigating unwanted algorithmic bias. *arXiv preprint arXiv:1810.01943* (2018).

[5] Ben Dattner, Tomas Chamorro-Premuzic, Richard Buchband, and Lucinda Schettler. 2019. The Legal and Ethical Implications of Using AI in Hiring. *Harvard Business Review* (25 April 2019). Retrieved 2-June-2020 from https://hbr.org/2019/04/the-legal-and-ethical-implications-of-using-ai-in-hiring

[6] 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.

[7] Jean Ensminger and Joseph Henrich. 2014. *Experimenting with social norms: Fairness and punishment in cross-cultural perspective*. Russell Sage Foundation.

[8] Matthias Feurer, Katharina Eggensperger, Stefan Falkner, Marius Lindauer, and Frank Hutter. 2018. Practical automated machine learning for the automl challenge 2018. In *International Workshop on Automatic Machine Learning at ICML*. 1189–1232.

[9] Kimberly Graves. 2010. *CEH certified ethical hacker study guide*. John Wiley & Sons.

[10] Christopher J Hamblin, Candace Gilmore, and Alex Chaparro. 2006. Learning to fly glass cockpits requires a new cognitive model. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, Vol. 50. SAGE Publications Sage CA: Los Angeles, CA, 1977–1981.

[11] Karen Hao. 2019. AI is sending people to jail – and getting it wrong. *MIT Technology Review* (21 January 2019). Retrieved 2-June-2020 from https://www.technologyreview.com/s/612775/algorithms-criminal-justice-ai/

[12] Michael Hind, Stephanie Houde, Jacquelyn Martino, Aleksandra Mojsilovic, David Piorkowski, John Richards, and Kush R Varshney. 2020. Experiences with Improving the Transparency of AI Models and Services. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–8.

[13] Udayan Khurana, Fatemeh Nargesian, Horst Samulowitz, Elias Khalil, and Deepak Turaga. 2016. Automating feature engineering. *Transformation* 10, 10 (2016), 10.

[14] 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.

[15] Huichen Li, Xiaojun Xu, Xiaolu Zhang, Shuang Yang, and Bo Li. 2020. QEBA: Query-Efficient Boundary-Based Blackbox Attack. *arXiv preprint arXiv:2005.14137* (2020).

[16] James Manyika, Michael Chui, Mehdi Miremadi, et al. 2017. A future that works: AI, automation, employment, and productivity. *McKinsey Global Institute Research, Tech. Rep* 60 (2017).

[17] 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.

[18] John Murawski. 2019. Mortgage Providers Look to AI to Process Home Loans Faster. *Wall Street Journal* (18 March 2019). Retrieved 2-June-2020 from https://www.wsj.com/articles/mortgage-providers-look-to-ai-to-process-home-loans-faster-11552899212

[19] Arvind Narayanan. 2018. Translation tutorial: 21 fairness definitions and their politics. In *Proc. Conf. Fairness Accountability Transp., New York, USA*.

[20] Maria-Irina Nicolae, Mathieu Sinn, Minh Ngoc Tran, Ambrish Rawat, Martin Wistuba, Valentina Zantedeschi, Nathalie Baracaldo, Bryant Chen, Heiko Ludwig, Ian M Molloy, et al. 2018. Adversarial Robustness Toolbox v0.4.0. *arXiv preprint arXiv:1807.01069* (2018).

[21] Burr Settles. 2009. *Active learning literature survey*. Technical Report. University of Wisconsin-Madison Department of Computer Sciences.

[22] Ben Shneiderman. 2020. Human-Centered Artificial Intelligence: Reliable, Safe & Trustworthy. *International Journal of Human–Computer Interaction* 36, 6 (2020), 495–504.

[23] Sahil Verma and Julia Rubin. 2018. Fairness definitions explained. In *2018 IEEE/ACM International Workshop on Software Fairness (FairWare)*. IEEE, 1–7.

[24] Jim Vinoski. 2018. Will A Robot Take My Manufacturing Job? Yes, No, And Maybe. Retrieved 19-May-2020 from https://www.forbes.com/sites/jimvinoski/2018/09/03/will-a-robot-take-my-manufacturing-job-yes-no-and-maybe

[25] 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.

[26] 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: Exploring Data Scientists' Perceptions of Automated AI. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (2019), 1–24.

[27] Daniel Karl I Weidele, Justin D Weisz, Erick Oduor, Michael Muller, Josh Andres, Alexander Gray, and Dakuo Wang. 2020. AutoAIViz: opening the blackbox of automated artificial intelligence with conditional parallel coordinates. In *Proceedings of the 25th International Conference on Intelligent User Interfaces*. 308–312.

[28] Richard Wike and Bruce Stokes. 2018. In Advanced and Emerging Economies Alike, Worries About Job Automation. *Pew Research Center, Global Attitudes & Trends* (2018).

[29] Amy X Zhang, Michael Muller, and Dakuo Wang. 2020. How do Data Science Workers Collaborate? Roles, Workflows, and Tools. *arXiv preprint arXiv:2001.06684* (2020).

[30] Liyue Zhao, Gita Sukthankar, and Rahul Sukthankar. 2011. Robust active learning using crowdsourced annotations for activity recognition. In *Workshops at the Twenty-Fifth AAAI Conference on Artificial Intelligence*.