Visualization and Automation in Data Science: Exploring the Paradox of Humans-in-the-Loop

Jen Rogers, Marie Anastacio, Jürgen Bernard, Mehdi Chakhchoukh, Rebecca Faust, Andreas Kerren, Steffen Koch, Lars Kotthoff, Cagatay Turkay, Emily Wall

Abstract—We explore the interplay between automation and human involvement in data science. Emerging from in-depth discussions at a Dagstuhl seminar, we synthesize perspectives from Automated Data Science (AutoDS) and Interactive Data Visualization (VIS) – two fields that traditionally represent opposing ends of the human-machine spectrum. While AutoDS seeks to enhance efficiency through increasing automation, VIS underscores the critical value of human involvement in providing nuanced understanding, creativity, innovation, and contextual relevance. We explore these dichotomies through an online survey and advocate for a balanced approach that harmonizes the speed and consistency of effective automation with the indispensable insights of human expertise and thought. Ultimately, we confront the essential question: what aspects of data science *should* we automate?

Index Terms—Human-Machine Interaction, Automated Data Science, Human-Centered AutoDS

1 Introduction

In 1960, sociologist Ida Russakoff Hoos published "When the Computer Takes Over the Office", capturing in a line, the collective anxiety about automation's impact on the workforce [30]. Today, the extent of automation integrated into our daily tasks would seem like science fiction to Hoos' contemporaries. Despite significant advancements in data science aimed at automating human labor to enhance efficiency and accuracy, the concerns over automation persist, as does the debate over the appropriate boundaries of automation in our lives.

In data science, the tension between human involvement and automation presents a thought-provoking paradox; while automation can enhance efficiency, it also risks diminishing human agency, reducing system transparency, and eroding trust [34]. Swept up in the momentum to automate the entire data science (DS) pipeline, we risk alienating the human users we aim to support [23].

This position paper investigates the complex interplay between automation and human involvement within the field of data science. Originating from a week-long seminar at Dagstuhl, this work unites researchers from two domains that represent opposing ends of the human-machine axis: Automated Data Science (AutoDS), aimed at reducing the need for human intervention in the data science process, and Visualization (VIS), an intrinsically human-centered discipline focused on enhancing human understanding and interaction with data. After drafting our research agenda during the seminar, we further refined our perspectives through a survey distributed to both the AutoDS and VIS communities. By integrating our insights with feedback from the broader community, we explore the dynamic between the operational efficiency driven by AutoDS and the human-centric ethos of VIS.

In this work, we explore the changing landscape of automation within data science, beginning with foundational work from both automated and human-centered perspectives. We detail our approach to analyzing automation and human participation, drawing from the broader community and reflecting on the benefits and challenges of each. This exploration highlights the potential for systems that effectively integrate human contextual insights with automated processes. Ultimately, this position paper underscores that the most pressing question for the future of AutoDS technology is not just what we *can* automate, but rather, what we *should* automate.

2 Data Science and Opposing Ends of the Human-Machine Axis

This section outlines the foundational concepts of AutoDS and Human-Centered Data Science, establishing a groundwork that supports our subsequent discussions, survey, and reflections.

2.1 The Data Science pipeline

For our work, we define the DS pipeline (Figure 1) with eight stages: *Data Acquisition* – obtaining the data from the process we wish to model; *Data Preparation* – any cleaning, normalization, etc.; *Data Exploration* – initial analysis of what the data looks like; *Feature Engineering* – preparing the data to be suitable for building models; *Model Selection and Training* – deciding on a modeling approach; *Result Analysis and Validation* – is the model appropriate and what does it tell us; *Result Communication* – preparation of materials for the domain expert; and *Post-Deployment Monitoring* – does the data distribution shift, etc. While there is no single way of defining a DS pipeline, we use this pipeline as a "canonical" model, covering most aspects considered by researchers and practitioners [1, 8, 16, 31, 44], to explore the use of automation in data science.

2.2 Automated Data Science

The DS pipeline often requires significant time from data scientists. While part of this time involves making meaningful decisions, most are spent on the tedious fine-tuning of individual pipeline steps. The paradigm of automated data science stems from the idea that automation can allow data scientists to focus on meaningful decisions, providing a turnkey solution for non-experts.

While hyperparameter optimization has long been the most automated part of the DS pipeline (*e.g.,* [3]), the emerging field of Automated Machine Learning (AutoML) [17] now covers most steps of the pipeline such as feature engineering, neural architecture search or data labeling (*e.g.,* [27, 33]).

However, automation comes at a cost: retraining a machine learning (ML) model many times during an optimization process is energy-demanding and goes against recent calls to look at the sustainability of AI (*e.g.,* [37]). Additionally, users of AutoDS systems often feel a lack

- *Jen Rogers, Tufts University, jen@cs.tufts.edu*
- *Marie Anastacio, RWTH Aachen University, anastacio@aim.rwth-aachen.de*
- *Jürgen Bernard, University of Zurich, bernard@ifi.uzh.ch*
- *Mehdi Chakhchoukh, University of Paris-Saclay, mehdi.chakhchoukh@universite-paris-saclay.fr*
- *Rebecca Faust, Tulane University, rfaust1@tulane.edu*
- *Andreas Kerren, Linköping University, andreas.kerren@liu.se*
- *Steffen Koch, University of Stuttgart, Steffen.Koch@vis.uni-stuttgart.de*
- *Lars Kotthoff, University of Wyoming, larsko@uwyo.edu*
- *Cagatay Turkay, University of Warwick, cagatay.turkay@warwick.ac.uk*
- *Emily Wall, Emory University, emily.wall@emory.edu*

Manuscript received xx xxx. 201x; accepted xx xxx. 201x. Date of Publication xx xxx. 201x; date of current version xx xxx. 201x. For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org. Digital Object Identifier: xx.xxxx/TVCG.201x.xxxxxxx

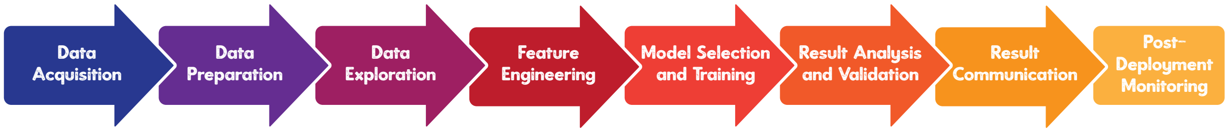

Fig. 1: Data science pipeline with eight main steps, some more data-oriented, some more model-oriented, but essentially all steps involve people. In the scope of this work, all steps are relevant for AutoDS and VIS practices alike. It represents a synthesis of process-oriented workflows and DS pipelines.

of control over the system's actions and outputs and are unsure how to assess result quality. This has led to calls for reintroducing human involvement in the automated data science loop [13, 42]. Automation makes this aim challenging, as AutoML systems often use ensemble models, making results hard to interpret. The adaptive nature of AutoDS, i.e., adjusting strategies to specifics of datasets and tasks, further complicates interpretation and explainability.

2.3 Human Factors in the Data Science Pipeline

Visualization often serves as the interface for human involvement in the DS pipeline, translating complex data into accessible visual formats that foster understanding and engagement (*e.g.,* [6, 32]). However, this is one aspect of a broader human factors approach in data science, which seeks to enhance human involvement and ensure that technology complements human capabilities. Research on human factors in the DS pipeline concentrates on ML pipelines. In light of this, much of the work we reference focuses on human factors within ML. Research on human involvement in data science generally adopts two approaches: direct participation through human-in-the-loop (HITL) and indirect involvement through understanding automated processes. Specifically, HITL ML involves active human participation across various pipeline stages, working collaboratively with automated tools and algorithms. Human-centered machine learning (HCML) intersects artificial intelligence (AI)/ML, Visualization, and Human-Computer Interaction (HCI) research, emphasizing human engagement in the pipeline to leverage contextual insights [36].

Regardless of the approach to human involvement, much of this work advocates for tailoring interactions and explanations within data science systems to the diverse roles of users —whether technical, business, regulatory, or end-user— ensuring the relevance and effectiveness of human involvement [10, 14, 21]. For example, technical users such as data scientists and ML engineers require detailed insights into performance metrics and model comparisons, whereas end-users benefit from clear explanations of final outcomes [21]. Such differentiation enriches HITL strategies and refines the scope and content of explanations throughout the DS process.

While HITL ML directly integrates human input throughout the data science process, Explainable AI (XAI) helps users understand automated decisions through explanations, visualizations, and interaction [7]. XAI emerged from the need to clarify how specific behaviors of automated systems lead to particular outcomes [10, 14], fostering user trust which is vital for the adoption and effectiveness of AI systems [19, 22, 26]. Effective explanation is both a human-centric and technical challenge [12], influenced by the user's expertise [4, 5].

2.4 Human-Machine Collaboration in AutoDS

Despite the emphasis on developing AutoDS tools, understanding their interaction with human operators remains underexplored. Wang *et al.* advocate that real-world applications of AutoDS not replace but rather enhance the collaborative partnership between humans and machines [38]. This sentiment is supported by others who highlight the significant human effort required in real-world scenarios [8], especially in decision-making [21] and contextual understanding [26]. Xin *et al.* take this sentiment further, advocating not only for a symbiotic relationship between humans and automation but discarding the term AutoDS as "complete automation is infeasible; instead, these tools can be better thought of as offering mixed-initiative ML solutions" [43].

Albeit the ongoing efforts to integrate human insights within the AutoDS framework [2, 18, 24, 35], progress has been limited and the challenges remain significant. AutoDS systems, which span multiple stages of the ML pipeline, necessitate increasingly complex explanations. Research varies from focusing on specific aspects, like hyperparameters [39], to broader system analyses such as PipelineProfiler [28], which scrutinizes the entire pipeline.

As previously mentioned, the adaptive nature of AutoDS complicates explainability. Addressing this involves transparent documentation, visualization tools, and user feedback integration to clarify the decision-making process and improve interpretability.

Provenance within the DS pipeline tracks the origin, history, and changes of data and digital artifacts, enhancing accountability and trust within the system [20]. It plays a key role in debugging and optimization [29], identifying and rectifying bottlenecks or errors, and addressing conceptual drift [15]. In AutoDS, provenance is multifaceted, covering data origins and transformations, model configurations and evolution, automation steps, and the rationale behind decision-making. This approach ensures transparency, reproducibility, and accountability in automated processes. Provenance becomes important in understanding and monitoring the pipeline's evolution, supporting effective human-machine collaboration. By integrating provenance and explainability with a strong emphasis on human understanding, AutoDS can make automated processes transparent and fosters a synergy that enhances uniquely human decision-making capabilities.

3 Methods

Our approach to developing this position paper was twofold. Initially, a week-long Dagstuhl seminar allowed us to lay the conceptual groundwork. Subsequently, we expanded and refined these ideas through an online survey to gather broader community insights. We then reflected on and improved our position for a year afterward.

3.1 Dagstuhl Seminar: Forming an Initial Perspective

The groundwork for this perspective paper was done during a Dagstuhl seminar titled "Human-Centered Approaches for Provenance in Automated Data Science," held in September 2023 [9]. This five-day event gathered researchers from both the VIS and AutoML communities, fostering a diverse exchange of ideas and collaborative efforts. During this seminar, our team—composed of experts from both fields—initiated discussions on the interplay of human factors within AutoDS.

These discussions consisted of brainstorming sessions where we mapped out the roles of humans, visualization, and automation within the AutoDS pipeline. From these sessions, we drafted an initial outline of broader themes of automation and human-in-the-loop/visualization, as outlined in Section 2 and our online survey.

3.2 Online Survey: The Role of VIS & AutoDS

We circulated a survey within the broader research community to expand upon our initial perspective and gain deeper insights into the balance between human involvement and automation within the DS pipeline. This survey explored perceptions of the current and desired levels of human and automated involvement across the pipeline stages.

The survey started with four Likert-scale questions to assess participants' views on current and desired levels of human and automation involvement, rated from 1 (low) to 5 (high). This format allowed us to capture a comparative perspective on each pipeline stage, as described in section 2. Additionally, we asked participants to indicate their preferred future level of human and automation involvement, with five options ranging from a "much smaller role" to a "much larger role". Additionally, four open-ended questions elicit detailed reflection on the advantages and challenges of automation, encouraging participants to offer examples. The survey concluded with an invitation for participants to join in follow-up interviews, to provide deeper insights into their responses.

We conducted our survey using Google Forms, distributing it via Twitter and Slack channels for both the VIS and AutoDS communities. To ensure privacy, participant responses remained anonymous, although individuals could provide their email for further inquiries. We requested participants specify their affiliation with either the VIS or AutoDS communities and define their role as either a researcher, developer, or end user. We received 11 responses, all identifying as researchers — five from the VIS and six from the AutoDS/AutoML communities. The survey took about ten minutes to complete.

We employed thematic analysis to identify emergent themes in the responses, beginning with a familiarization phase where we conducted a word frequency analysis to identify keywords and phrases. We then conducted a comparative analysis of responses from the VIS and AutoML communities to examine their views on human involvement and automation in data science. By extracting key points and phrases from responses, we identified thematic differences and compared the frequency of themes between the groups. This approach allowed us to delineate each group's distinct perspective.

4 The Good, The Bad, and the Human in Automated Data Science

We present the survey results, reflect on the survey results in dialogue with the discussions from the research seminar and interrogate the increasing role of automation in data science. Here, we examine the benefits and challenges, asking what *should* we automate versus what *can* we automate. The synthesis provides insights into the strengths and weaknesses of automation, identifying which aspects of data science should remain human-driven.

4.1 High-Level Overview

The survey results, illustrated in Figure 2, provide an overview of both the current and desired roles of human involvement and automation across all stages of the DS process. Respondents identified several stages where human engagement is pronounced, including *Data Preparation*, *Data Exploration*, *Result Communication*, and *Post-Deployment Monitoring*. Of these, *Data Exploration* and *Result Communication* are particularly highlighted as areas where human insight and oversight are crucial for ensuring the accuracy and relevance of findings.

Data Preparation and *Post-Deployment Monitoring* currently involve significant human effort (Fig. 2.A) and most respondents agree that these stages should be increasingly automated in the future (Fig. 2.D). On the other hand, whilst *Model Selection and Training* are currently the stages with the least human involvement (Fig. 2.A-B), respondents indicated a desire for even less human involvement (Fig. 2.C-D). This suggests confidence in the ability of automated systems to handle these complex tasks more effectively as technologies evolve.

Overall, the survey indicates a desire to reduce human involvement in the DS pipeline, with notable exceptions for stages that benefit from human judgment, such as *Data Exploration* and the *Communication of Results*. Respondents favor a future where automation is balanced with strategic human oversight for stages that require deep understanding and critical thinking. This reflects the evolving landscape in data science. As automation increases, practitioners continuously adapt their vision of how much automation to include at each step. The integration of advanced automation tools aims to optimize efficiency and accuracy, while human expertise is applied where it adds the most value, harnessing the strengths of both.

4.2 Benefits of Automation

The case for automation and its benefits seem self-evident. In data science, automation is widely embraced because it accelerates analysis and decision-making, offering computational speed and memory capacity while addressing human limitations in the pipeline. By exploring the strategic advantages of automation in data science, this section synthesizes insights from our survey and seminar discussions, focusing on automation's role in improving efficiency, reproducibility, collaboration, and reducing human errors and biases throughout the entire pipeline.

Automation for Efficiency: Notably, survey participants emphasized the human tendency to *"[slow] down the process in literally every step"* underscoring automation's potential to streamline and expedite aspects of the pipeline. Efficiency is a key factor identified in automation, accelerating various stages of the data science process, particularly in repetitive or time-consuming tasks such as data preparation, exploration, and post-deployment monitoring. As one survey respondent emphasized, *"If post-deployment monitoring can be automated, it would immensely reduce cost and effort from systems and projects becoming stale after publication."* This increase in efficiency is especially pertinent as datasets become larger and more complex. *"Automated methods are urgently needed due to increasing size of data."*

In addition to time efficiency, automating routine work within the pipeline decreases cognitive load for the human user, *"so that they may focus more on understanding the problem, the data, and the models."*. By automating the lower-level tasks, we can allow more space for higher-level tasks and decision-making.

Automation for reproducibility and collaboration: Automation in model selection and result analysis enhances the clarity of protocols, improving reproducibility in data science processes. This aspect, as highlighted in our survey, makes it easier for others to understand the methodology, contributing to better transparency and collaboration.

Automation allows for parallel processing and provides provenance information for parallel solutions. This capability facilitates collaboration and concurrent work, addressing the need for robust mechanisms to track and understand the evolution of solutions in collaborative data science projects.

Automating out human error and bias: Automation mitigates the risk of human error and subjective bias, particularly in data preparation, feature selection, result analysis, and validation. One respondent cited that human *"intervention may be subjective/biased in data preparation/feature selection."* While another stated; *"during result analysis and validation, humans might be prone to injecting their subjective bias."* Survey participants recognized the importance of reducing reliance on human intervention in these critical aspects, emphasizing the potential for more objective and unbiased outcomes in data science tasks.

While automation can add objectivity and efficiency to the DS pipeline, it is not immune to bias. The extent to which bias is introduced depends on the type of automation. For example, models can introduce biases based on how they were modeled and trained. Additionally, automation relies on the assumption that the initial data is unbiased, which is often not the case. Although automation does not introduce bias, it cannot detect it, lacking a representation of what an unbiased model should be.

4.3 Challenges in Automation

Despite the benefits, automation still has challenges, specifically in the lack of context, transparency, interpretability, and validation.

No model to rule them all: A universal or one-size-fits-all model that can outperform others across all data types and tasks does not exist. Referred to as the "no free lunch" theorem [41], this poses a significant challenge for automation without human discretion. As diverse datasets and tasks are common in data science, this theorem implies that no single algorithm or optimization approach can consistently deliver optimal results for every scenario and domain, emphasizing the importance of a flexible and adaptive approach that considers the nuances of specific data and problem domains. AutoDS systems must navigate the trade-offs and variations in model performance, making it essential for developers and practitioners to carefully choose or customize algorithms based on the specific tasks and context.

Lack of semantic understanding and contextual relevance: The lack of a universal model requires additional semantic understanding and context to find the right fit in light of the trade-offs. As a participant of our survey had voiced; *"[a]utomated methods cannot consider domain knowledge (e.g., importance/impact of detected features) yet,"* emphasizing the critical role human insights have in guiding the au-

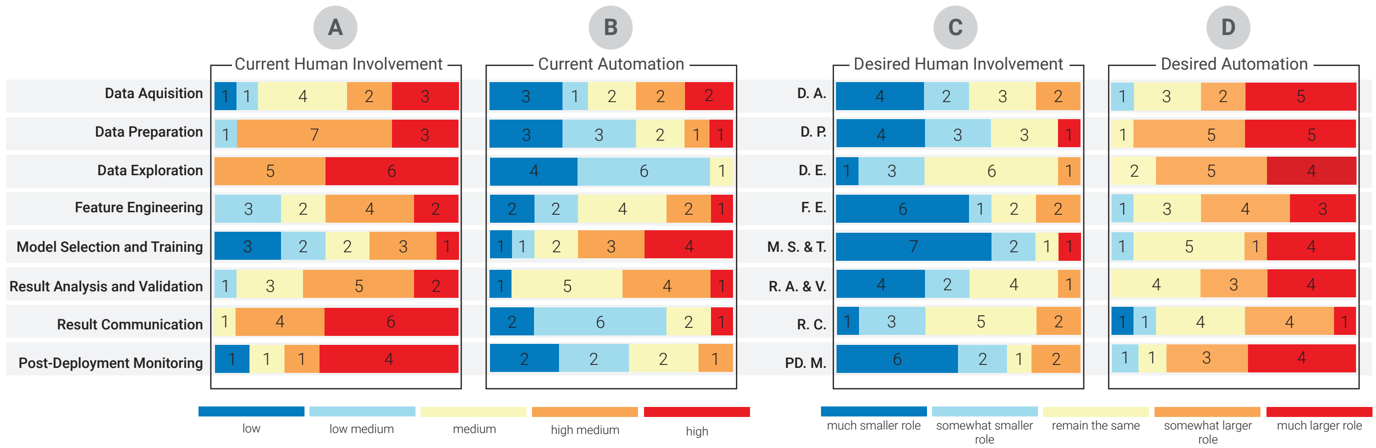

Fig. 2: Summary of survey results for current (A-B) and desired (C-D) levels of human involvement and automation. Stages of the pipeline are represented as rows in each chart. These rows are color-coded by the heuristic specified in the color legend. Each segment's length is proportional to the percentage of respondents who selected that category and the number of respondents is indicated by the number in each colored segment. Charts A and B reveal that 'Data Exploration' involves the most human labor, while 'Model Selection and Training' involves the most automation. The right charts show an overall preference for increased automation in the pipeline, especially in the 'Data Preparation' stage. Though participants prefer more automation, 'Data Exploration' and 'Result Communication' where most respondents also want human involvement to remain the same or play a somewhat larger role while also wanting more automation. This indicates areas for more integrated human-machine collaboration, highlighting a shift toward balancing automation with human expertise in key areas.

tomated modeling process. Context is critical for mapping solutions to real-world tasks. Without context, *"the found solution might not contribute to solving the actual task."*

Additionally, context is an imperative consideration in defining fairness. What does fairness mean within an automated system? Fairness is often context-dependent, and the same model may exhibit different levels of fairness in different situations. Understanding and addressing contextual factors that influence fairness, such as the specific task or domain, is crucial for developing fair automated systems.

Lack of interpretability, transparency, and meta-problems: As automated systems still rely on human discretion and oversight, a major challenge in current automated systems is interpretability and transparency. In particular, the challenge of model transparency is compounded by the creation of ensembles with numerous parameters and the intricacies introduced by deep learning techniques. An added layer of complexity arises from the occasional failure of automated processes, often without clear indications for human users regarding the cause and nature of the failure. From a human-centric perspective, the Algorithmic Imprint concept, as discussed by Ehsan *et al.* [11], highlights the lasting effects of algorithmic failures and emphasizes the need to understand the consequences of such failures. The black-box nature of automation exacerbates these issues, obscuring the mechanisms behind the solutions proposed by AutoDS systems. Key questions surrounding the trustworthiness of results become pivotal—can the output of an automated data science system be relied upon, and to what extent can the human user place confidence in the proposed solutions? Addressing these challenges requires a concerted effort to enhance transparency, refine the interpretability of automated processes, and allow users to discern the reliability of their results.

4.4 The Trade-off of Human-Involvement

As discussed in the previous section, challenges remain in current AutoDS technology. Automation enhances efficiency in cost and resources, streamlines lower-level processes, facilitates decision-making, and accelerates analytical processes in data science. While technology will continue to improve, solving some of these pain points, we reflect on what is uniquely human about this process, and which aspects, if any, should remain under human supervision.

Domain knowledge and context: Humans bring a contextual understanding of the domain and link between data and real-world objectives. *"[They] can look at a problem with a broader perspective and define helpful objectives that ensure progress in a meaningful direction."* Though the survey results largely advocate for more automation in stages such as Data Preparation, some responses stress that humans remain critical in this stage, as well as in *"Feature Engineering, Result Analysis, and Validation,"* as their domain knowledge enables more informed and meaningful decisions. We view this divergence in opinion as indicative of the dynamic collaboration we advocate between humans and automation, where involvement is dependent on the specific situation. Humans can provide context and intervene when needed, especially if trust in the automated system is compromised.

Subjectivity and nuance: Humans excel at recognizing nuances and our survey highlights this in several examples. Respondents indicated that they do not want the result communication stage to be heavily automated (Fig. 2.C-D), as this stage *"can be very nuanced, it's important to have a human in the loop who can explain the subtleties and relevance, especially when communicating results with people from other fields/domains."*

This sentiment extends to other analysis aspects, such as anomaly detection. Though automation can improve the identification of outliers [25, 40], it may miss more subtle ones. *"Some anomalous data point, which has not been flagged as important by the model can still be important and having a "manual" look at the data often surfaces these nuances — specifically in qualitative analysis."*

Fairness and ethical considerations Humans play a key role in defining fairness, especially in automated processes. Contextual considerations underscore the ethical dimensions of data science tasks. *"Humans can consider fairness by looking at the context of the task, e.g., medical applications."* While this may seem contrary to the idea that automation reduces human biases, it also suggests a dynamic where humans and machines can each leverage their respective strengths.

5 Conclusion

Will we ever be entirely comfortable with full automation without human oversight? If so, what implications does this have for visual analytics and more broadly, data visualization? If not, how can we leverage the benefits of automation to reduce cognitive load and facilitate more effective decision-making? What measure can we implement to enhance transparency within the highly heterogeneous DS pipeline?

The future of data science is inherently collaborative, where automation complements rather than replaces human capabilities. Human judgment remains indispensable for imparting ethical considerations, domain-specific knowledge, and nuanced decision-making into automated systems. This work is a call to action for developing AutoDS systems that leverage the strengths of automation while embracing the irreplaceable unique value of human intellect and intuition.

Acknowledgments

This work was initiated by discussions at the Dagstuhl Seminar 23372 "Human-Centered Approaches for Provenance in Automated Data Science." We thank Dagstuhl and the organizers for making this happen.

References

[1] S. Amershi, A. Begel, C. Bird, R. DeLine, H. C. Gall, E. Kamar, N. Nagappan, B. Nushi, and T. Zimmermann. Software engineering for machine learning: a case study. In *Int. Conf. on Softw. Eng.: Software Engineering in Practice, ICSE (SEIP)*, pp. 291–300. IEEE / ACM, 2019. 1

[2] M. Anastacio and H. Hoos. Model-based algorithm configuration with default-guided probabilistic sampling. In T. Bäck, M. Preuss, A. Deutz, H. Wang, C. Doerr, M. Emmerich, and H. Trautmann, eds., *Parallel Problem Solving from Nature – PPSN XVI*, pp. 95–110. Springer Int. Publishing, Cham, 2020. 2

[3] J. S. Bergstra, R. Bardenet, Y. Bengio, and B. Kégl. Algorithms for Hyper-Parameter Optimization. In *Advances in Neural Information Processing Systems 24*, pp. 2546–2554. Curran Associates, Inc., 2011. 1

[4] A. Chatzimparmpas, K. Kucher, and A. Kerren. Visualization for trust in machine learning revisited: The state of the field in 2023. *IEEE Computer Graphics and Applications*, 44(3):99–113, 2024. 2

[5] A. Chatzimparmpas, R. M. Martins, I. Jusufi, K. Kucher, et al. The state of the art in enhancing trust in machine learning models with the use of visualizations. *Computer Graphics Forum*, 39(3):713–756, 2020. 2

[6] A. Chatzimparmpas, R. M. Martins, K. Kucher, and A. Kerren. StackGenVis: Alignment of data, algorithms, and models for stacking ensemble learning using performance metrics. *IEEE Transactions on Visualization and Computer Graphics*, 27(2):1547–1557, 2020. 2

[7] J. Choo and S. Liu. Visual analytics for explainable deep learning. *IEEE computer graphics and applications*, 38(4):84–92, 2018. doi: 10.1109/MCG.2018.042731661 2

[8] A. Crisan and B. Fiore-Gartland. Fits and starts: Enterprise use of automl and the role of humans in the loop. In *Proc. CHI*, pp. 1–15, 2021. 1, 2

[9] A. Crisan, L. Kotthoff, M. Streit, and K. Xu. Human-Centered Approaches for Provenance in Automated Data Science (Dagstuhl Seminar 23372). *Dagstuhl Reports*, 13(9):116–136, 2024. 2

[10] U. Ehsan and M. O. Riedl. Human-centered explainable AI: Towards a reflective sociotechnical approach. In *Proc. HCI Int.*, pp. 449–466. Springer, 2020. doi: 10.1007/978-3-030-60117-1_33 2

[11] U. Ehsan, R. Singh, J. Metcalf, and M. Riedl. The algorithmic imprint. In *Proc. ACM Conf. on Fairness, Accountability, and Transparency*, pp. 1305–1317, 2022. doi: 10.1145/3531146.3533186 4

[12] U. Ehsan, P. Wintersberger, Q. V. Liao, E. A. Watkins, C. Manger, H. Daumé III, A. Riener, and M. O. Riedl. Human-centered explainable ai (hcxai): beyond opening the black-box of ai. In *Proc. CHI*, pp. 1–7, 2022. doi: 10.1145/3491101.3503727 2

[13] J. Giovanelli and G. Pisano. Towards human-centric automl via logic and argumentation. In *EDBT/ICDT Workshops*, 2022. 2

[14] D. Gunning, M. Stefik, J. Choi, T. Miller, S. Stumpf, and G.-Z. Yang. Xai—explainable artificial intelligence. *Science robotics*, 4(37):eaay7120, 2019. doi: 10.1126/scirobotics.aay7120 2

[15] X. Han, T. Pasquier, A. Bates, J. Mickens, and M. Seltzer. Unicorn: Runtime provenance-based detector for advanced persistent threats. *arXiv preprint arXiv:2001.01525*, 2020. doi: arXiv:2001.01525 2

[16] F. Hohman. *Interactive Scalable Interfaces for Machine Learning Interpretability.* PhD thesis, Georgia Institute of Technology, Atlanta, GA, USA, 2021. 1

[17] F. Hutter, L. Kotthoff, and J. Vanschoren, eds. *Automated Machine Learning: Methods, Systems, Challenges*. The Springer Series on Challenges in Machine Learning. Springer, Cham, 1 ed., 2019. 1

[18] C. Hvarfner, D. Stoll, A. Souza, L. Nardi, M. Lindauer, and F. Hutter. πBO: Augmenting acquisition functions with user beliefs for bayesian optimization. In *Int. Conf. on Learning Representations*, 2022. 2

[19] S. Jia, Z. Li, N. Chen, and J. Zhang. Towards visual explainable active learning for zero-shot classification. *IEEE Transactions on Visualization and Computer Graphics*, 28(1):791–801, 2021. 2

[20] A. Kale, T. Nguyen, F. C. Harris Jr, C. Li, J. Zhang, and X. Ma. Provenance documentation to enable explainable and trustworthy ai: A literature review. *Data Intelligence*, 5(1):139–162, 2023. 2

[21] T. T. Khuat, D. J. Kedziora, and B. Gabrys. The roles and modes of human interactions with automated machine learning systems. *Foundations and Trends® in Human–Computer Interaction*, 17(3-4):195–387, 2023. 2

[22] B. La Rosa, G. Blasilli, R. Bourqui, D. Auber, G. Santucci, R. Capobianco, E. Bertini, R. Giot, and M. Angelini. State of the art of visual analytics for explainable deep learning. In *Computer Graphics Forum*, vol. 42, pp. 319–355. Wiley Online Library, 2023. doi: 10.1111/cgf.14733 2

[23] D. J. L. Lee, S. Macke, D. Xin, A. Lee, S. Huang, and A. G. Parameswaran. A human-in-the-loop perspective on automl: Milestones and the road ahead. *IEEE Data Eng. Bull.*, 42(2):59–70, 2019. 1

[24] D. J. L. Lee, S. Macke, D. Xin, A. Lee, S. Huang, and A. G. Parameswaran. A human-in-the-loop perspective on automl: Milestones and the road ahead. *IEEE Data Eng. Bull.*, 42:59–70, 2019. 2

[25] B. Montambault, C. D. Brumar, M. Behrisch, and R. Chang. Pixal: Anomaly reasoning with visual analytics. *arXiv preprint arXiv:2205.11004*, 2022. doi: 10.48550/arXiv.2205.11004 4

[26] E. Mosqueira-Rey, E. Hernández-Pereira, D. Alonso-Ríos, J. Bobes-Bascarán, and Á. Fernández-Leal. Human-in-the-loop machine learning: a state of the art. *Artif. Intell. Rev.*, 56(4):3005–3054, 2023. 2

[27] I. Namatēvs, K. Sudars, and I. Polaka. Automatic data labeling by neural networks for the counting of objects in videos. *Procedia Computer Science*, 2019. doi: 10.1016/j.procs.2019.01.118 1

[28] J. P. Ono, S. Castelo, R. Lopez, E. Bertini, J. Freire, and C. Silva. Pipelineprofiler: A visual analytics tool for the exploration of automl pipelines. *IEEE Transactions on Visualization and Computer Graphics*, 27(2):390–400, 2020. doi: 10.1109/TVCG.2020.3030361 2

[29] J. Rogers and A. Crisan. Tracing and visualizing human-ml/ai collaborative processes through artifacts of data work. In *Proc. CHI*, pp. 1–22, 2023. doi: 10.1145/3544548.3580819 2

[30] I. Russakoff Hoos. When the computer takes over the office. *Office Technology and People*, 2(1):57–67, 1983. doi: 10.1108/eb022621 1

[31] D. Sacha, M. Sedlmair, L. Zhang, J. A. Lee, J. Peltonen, D. Weiskopf, S. C. North, and D. A. Keim. What you see is what you can change: Human-centered machine learning by interactive visualization. *Neurocomputing*, 268:164–175, 2017. doi: 10.1016/J.NEUCOM.2017.01.105 1

[32] D. Sacha, M. Sedlmair, L. Zhang, J. A. Lee, J. Peltonen, D. Weiskopf, S. C. North, and D. A. Keim. What you see is what you can change: Human-centered machine learning by interactive visualization. *Neurocomputing*, 268:164–175, 2017. doi: 10.1016/j.neucom.2017.01.105 2

[33] S. Saxena and J. Verbeek. Convolutional neural fabrics. pp. 4053–4061, 2016. 1

[34] B. Shneiderman. Human-centered artificial intelligence: Reliable, safe & trustworthy. *Int. Journal of Human-Computer Interaction*, 36(6):495–504, 2020. doi: 10.1080/10447318.2020.1741118 1

[35] A. Souza, L. Nardi, L. B. Oliveira, K. Olukotun, M. Lindauer, and F. Hutter. Bayesian optimization with a prior for the optimum. In N. Oliver, F. Pérez-Cruz, S. Kramer, J. Read, and J. A. Lozano, eds., *Machine Learning and Knowledge Discovery in Databases. Research Track*, pp. 265–296. Springer International Publishing, Cham, 2021. 2

[36] F. Sperrle, M. El-Assady, G. Guo, R. Borgo, D. H. Chau, A. Endert, and D. Keim. A survey of human-centered evaluations in human-centered machine learning. In *Computer Graphics Forum*, vol. 40, pp. 543–568. Wiley Online Library, 2021. doi: 10.1111/cgf.14329 2

[37] A. van Wynsberghe. Sustainable ai: Ai for sustainability and the sustainability of ai. *AI and Ethics*, 1:213 – 218, 2021. 1

[38] D. Wang, J. Andres, J. D. Weisz, E. Oduor, and C. Dugan. Autods: Towards human-centered automation of data science. In *Proc. CHI*, pp. 1–12, 2021. doi: 10.1145/3411764.3445526 2

[39] Q. Wang, Y. Ming, Z. Jin, Q. Shen, D. Liu, M. J. Smith, K. Veeramachaneni, and H. Qu. Atmseer: Increasing transparency and controllability in automated machine learning. In *Proc. CHI*, pp. 1–12, 2019. 2

[40] L. Wilkinson. Visualizing big data outliers through distributed aggregation. *IEEE transactions on visualization and computer graphics*, 24(1):256–266, 2017. doi: 10.1109/TVCG.2017.2744685 4

[41] D. H. Wolpert and W. G. Macready. No free lunch theorems for optimization. *IEEE transactions on evolutionary computation*, 1(1):67–82, 1997. doi: 10.1109/4235.585893 3

[42] I. Xanthopoulos, I. Tsamardinos, V. Christophides, E. Simon, and A. Salinger. Putting the human back in the automl loop. In *EDBT/ICDT Workshops*, 2020. 2

[43] D. Xin, E. Y. Wu, D. J.-L. Lee, N. Salehi, and A. Parameswaran. Whither automl? understanding the role of automation in machine learning workflows. In *Proc. CHI*, pp. 1–16, 2021. 2

[44] M.-A. Zöller and M. F. Huber. Benchmark and survey of automated machine learning frameworks. *Journal of artificial intelligence research*, 70:409–472, 2021. doi: 10.1613/jair.1.11854 1