AutoML in The Wild: Obstacles, Workarounds, and Expectations

Yuan Sun
yws5055@psu.edu
Pennsylvania State University
University Park, USA

Qiurong Song
qzs5098@psu.edu
Pennsylvania State University
University Park, USA

Xininig Gui
xinninggui@psu.edu
Pennsylvania State University
University Park, USA

Fenglong Ma
fenglong@psu.edu
Pennsylvania State University
University Park, USA

Ting Wang
inbox.ting@gmail.com
Pennsylvania State University
University Park, USA

ABSTRACT

Automated machine learning (AutoML) is envisioned to make ML techniques accessible to ordinary users. Recent work has investigated the role of humans in enhancing AutoML functionality throughout a standard ML workflow. However, it is also critical to understand how users adopt existing AutoML solutions in complex, real-world settings from a holistic perspective. To fill this gap, this study conducted semi-structured interviews of AutoML users (N = 19) focusing on understanding (1) the limitations of AutoML encountered by users in their real-world practices, (2) the strategies users adopt to cope with such limitations, and (3) how the limitations and workarounds impact their use of AutoML. Our findings reveal that users actively exercise user agency to overcome three major challenges arising from customizability, transparency, and privacy. Furthermore, users make cautious decisions about whether and how to apply AutoML on a case-by-case basis. Finally, we derive design implications for developing future AutoML solutions.

CCS CONCEPTS

• **Human-centered computing** → **User studies**.

KEYWORDS

Automated Machine Learning, Privacy, Transparency, Customizability, User Agency

ACM Reference Format:
Yuan Sun, Qiurong Song, Xininig Gui, Fenglong Ma, and Ting Wang. 2023. AutoML in The Wild: Obstacles, Workarounds, and Expectations. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23), April 23–28, 2023, Hamburg, Germany*. ACM, New York, NY, USA, 15 pages. https://doi.org/10.1145/3544548.3581082

CHI '23, April 23–28, 2023, Hamburg, Germany

ACM ISBN 978-1-4503-9421-5/23/04…$15.00
https://doi.org/10.1145/3544548.3581082

1 INTRODUCTION

While machine learning (ML) has been successfully applied to solve many challenging tasks across various domains, building performant ML solutions still requires substantial resources and extensive human expertise [34]. Automated machine learning (AutoML), a novel concept for automating the whole ML pipeline without (or as little as possible) human intervention [39], has emerged as a way to significantly reduce expensive development costs [75]. As illustrated in Fig. 1, envisioned to enable domain experts without considerable ML backgrounds (e.g., marketing and business analysts) to build ML solutions more easily, AutoML holds the promise of making ML techniques accessible to more people. Meanwhile, by liberating users from repetitive ML tasks (e.g., data preprocessing, parameter tuning, and feature selection), domain experts can spend more time on essential tasks, while data scientists can build more ML models in less time, improve model quality and accuracy, and experiment with more new algorithms.

Despite its tremendous potential, the current discourse around AutoML is a mixture of hope and frustration. On one hand, AutoML is believed to be the driving force of "democratizing data science" [62]. On the other hand, the realization of automating ML workflows faces severe challenges [87], resulting in limited adoption of AutoML by practitioners [20]. To facilitate the design of AutoML and support user experience, recent HCI studies have started investigating how users perceive and use AutoML in practice. One line of research focused on user trust in AutoML [24, 76] and studied how to incorporate visualization tools to improve AutoML's transparency [56, 81]. Another line of research gathered qualitative data about how users apply AutoML within the standard ML workflow to understand AutoML's benefits and deficiencies from the users' perspective [19, 55, 87]. These studies examined the roles of users as data scientists throughout a standard computational pipeline and emphasized the importance of bringing "human-in-the-loop" to strike a balance between human control and machine automation [3, 49]. The primary goal was to compensate for automation-induced functionality deficiencies through human intervention to improve AutoML performance [58, 87]. However, in complex, real-world settings, the concrete data science tasks are defined and executed by users with varying roles, interests, and backgrounds [58]; users may face a variety of issues beyond AutoML's functionality, and they often define their goals according to complex situations and use different resources to accomplish them [69]. Therefore, using the standard ML workflow as a scaffold with the goal of achieving AutoML effectiveness may limit our

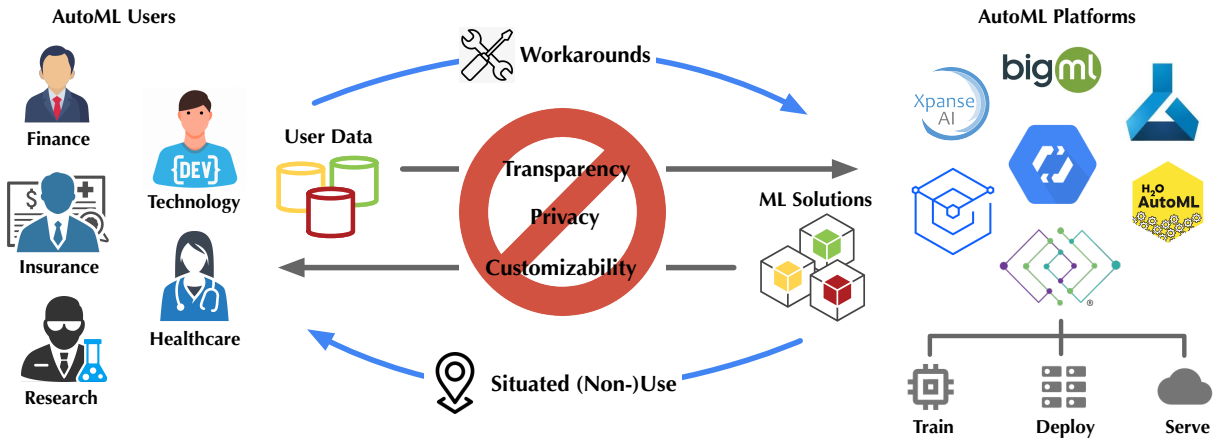

Figure 1: Obstacles and Workarounds of Using AutoML in Complex, Real-world Settings.

understanding of how users perceive and use AutoML in practical settings, while relying primarily on a technical perspective of data science may fail to account for the "social nuances, affective relationships, or ethical, value-driven concerns" of AutoML users [5]. This calls for a critical shift from examining the role of users in the standard ML workflow to understanding how users leverage AutoML as resources in their problem-solving processes.

The present study aims to fill this research gap by focusing on how users evaluate the role of AutoML in different situations and how they exercise user agency in leveraging AutoML in practice by developing various workaround strategies when facing challenges. As users' understanding of AutoML may be shaped by many situated factors, it is critical to investigate how users adopt and use AutoML in a variety of heterogeneous, specific situations. We conducted semi-structured interviews with 19 real-world AutoML users from different domains, varying in industries, job roles, and ML expertise levels. Specifically, our study discovered three major challenges, namely *lack of customizability*, *lack of transparency*, and *privacy concerns*, which impede users from effectively applying AutoML in complex, practical situations, as illustrated in Fig. 1. We analyzed the tactics devised by users to tackle such challenges and fit AutoML to diverse personal and organizational objectives, including various workaround strategies as well as selective and situated (non-)use of AutoML. Theoretical and design implications are provided.

2 RELATED WORK

2.1 HCI Research on AutoML

2.1.1 Users' Perceptions of AutoML. Recent studies have examined users' overall perceptions of AutoML. Wang et al. [80] focused on data scientists' perceptions of AutoML's utility and potential impacts on data science practices and occupations. Through a qualitative interview, they found that most data scientists consider AutoML as a complementary tool and perceive that AutoML is not able to replace human expertise. In another study, Wang et al. [78] developed an experimental AutoML system that only requires data scientists to upload their datasets and generates ML models automatically, and conducted a user study on data scientists' experience with such systems. They found that although the system provides high-quality models in less time, data scientists show less trust in the AutoML-generated models, largely due to a lack of transparency.

2.1.2 AutoML's Transparency. Many studies have focused on enhancing AutoML's transparency to foster user trust. Drozdal et al. [24] found that including transparency features such as model performance metrics and visualization is critical for increasing user trust. In a similar vein, Wang et al. [81] designed and implemented a visualization tool to support user understanding of the generated ML models and search space, and demonstrated that such visualization tools enhance user trust and enable users to better apply AutoML. Weidele et al. [83] developed an experimental system to visualize AutoML's model generation process and found that opening this black box improves user trust. Narkar et al. [56] designed a visualization tool to support user decision-making by analyzing AutoML's outcomes and comparing candidate ML models based on multiple performance metrics. In addition to this strand of design and experimental effort, Crisan and Fiore-Gartland [19] interviewed data scientists in enterprise settings, focusing on how they use visualization as a means to integrate humans into the automation loop. However, they found that participants' use of visualization is fairly limited due to the lack of usefulness and benefits [19].

2.1.3 Human-in-The-Loop AutoML. Although AutoML's goal is to "reduce the role of humans in the loop and fill the gap for non-expert users by playing the role of the domain expert" [27], there is a growing line of research that emphasizes the importance of "human-in-the-loop". Intuitively, AutoML with humans in the loop enables users to monitor and control AutoML's different stages yet without manually taking over the whole process [19], which also allows users to incorporate intuition and domain knowledge into AutoML's workflow to enhance its performance [30, 47]. To this end, Behnaz et al. [7] designed a new AutoML framework with interpretable feedback that allows users to leverage their domain knowledge. Gil et al. [30] proposed a hybrid framework that provides an intelligent interface enabling users to specify problem settings and explore different models, while AutoML functions under users' guidance. Besides, there have been studies focusing on understanding the roles of humans and automation in the standard ML workflow from a "human-in-the-loop" perspective. Wang et

al. [79] found that different data scientists desire for varying levels of automation depending on their experiences and the workflow stages. Crisan and Fiore-Gartland [19] further detailed the need for human involvement in different stages such as data preparation, analysis, deployment, and communication [54, 80]. In a similar vein, Xin et al. [87] examined user-reported AutoML's benefits and deficiencies and highlighted the importance of including humans in AutoML's process to combat AutoML's deficiencies such as system failures, lack of customizability, and lack of transparency.

2.2 User Agency and AutoML

2.2.1 Human Agency and Automation. The aforementioned studies, from a broader view, reflect the desiderata to balance human agency and machine automation. While machine automation outperforms humans in terms of efficiency and cost reduction, over-reliance on automation may sacrifice humans' critical engagement and domain expertise [35]. The tension between human agency and automation creates critical challenges for system design [35]. HCI research has long advocated for more human-centered approaches that balance humans and automation [4, 35, 64, 65]. For instance, the debate over "direct manipulation versus interface agents" between HCI and AI researchers back in the 1990s arrived at the conclusion of "increased automation that amplifies the productivity of users and gives them increased capabilities in carrying out their tasks while preserving their sense of control and their responsibility" [65]. Along this line, Horvitz [37] proposed a set of principles for designing systems to support automation-human collaboration efficiently, such as "considering uncertainty about a user's goals" and "providing mechanisms for efficient agent-user collaboration to refine results". Recently, by synthesizing AI design from various sources, Amershi et al. [4] proposed a set of guidelines to support the interaction between humans and AI.

2.2.2 Supporting User Agency in ML. In the context of ML, existing work has investigated how to support user agency through interface mechanisms such as allowing user feedback [36], improving customizability for ML tools [85], and enabling users to interactively modify AutoML's search space [81]. Through self-generated actions, these technological affordances have been shown to effectively enhance user agency [72]. However, it is found that users often exert their own agency to modify or adapt ML solutions. For instance, Cai et al [13] designed refinement tools to assist medical decision-making and found that users often invent new strategies such as disambiguating ML and human errors to better understand the underlying algorithms. In a similar vein, Xin et al. [87] showed that users switch back to manually developed ML models when they perceive high risks of using AutoML due to its lack of transparency. These "workarounds" are conscious and creative acts to ensure ML works in practice [2]. Recent research [31] has pointed out that there is a lack of understanding of the nature of applying ML as a co-adaptive process, in which users adapt to using ML more effectively and adjust their goals accordingly.

2.3 Summary

While recognizing the importance of the role of humans in AutoML, existing HCI studies mainly focused on how humans may compensate for AutoML's deficiencies in each step of the standard ML workflow, such as developing interface affordances of customization. However, in practice, because users operate in complex, real-world settings, have diverse expertise and backgrounds, and develop varying experiences engaging with AutoML, they may adjust their use of AutoML in ways unimaginable by designers [21]; user actions are not limited by AutoML's deficiencies, and users may be resourceful enough to work around such limitations, fitting AutoML to their targets and needs [10]. Built upon the strand of existing research, our work takes on a more "user-centered" perspective, with a special focus on how real-world users leverage AutoML as one of the resources in their problem-solving processes and the social-technical implications of AutoML in their practices, which contributes to the understanding of how applying AutoML is also a co-adaptive process.

3 METHOD

We conducted semi-structured interviews to understand users' perceptions of AutoML and their exercises of human agency in working with and around AutoML. We further identified user expectations for the future design of AutoML.

3.1 Recruitment and Interviews

We focused on users who have hands-on experience with AutoML in different domains. We recruited participants by spreading recruitment messages through words of month (n = 10), mailing lists within enterprises (n = 4), and social media (n = 5). Participants were invited to complete a screening questionnaire about whether they had used AutoML before, which AutoML platforms they had used, and how long their experience was. We recruited 19 participants who reported having experience with AutoML in their workplaces.

The interviews were conducted remotely from May 2022 to July 2022 after receiving the institutional review board (IRB) approval. Each interview was scheduled for 60 minutes on video conferencing platforms and was audio-recorded for transcription purposes. The average interview duration was 45 minutes, with individual ones varying from 35 to 60 minutes. We recruited participants who span a diverse range of domains across healthcare (n = 3), finance (n = 1), human resources (n = 1), technologies (n = 9), and academic research (n = 3). In addition, participants have varying job roles, from the marketing manager at a social media agency to the head of technology at a healthcare company. The majority of our participants are based in the United States, with one based in China (P13), and one in Kenya (P6). Each participant received a $50 gift card upon completion of the interview. A summary of participants' ML expertise levels, industries, job roles, and the type(s) of AutoML platforms they used is presented in Table 1. We omitted the specific names of AutoML platforms to preserve the anonymity of participants.

3.2 Data Analysis

The dataset for analysis included all 19 interview transcripts. Two researchers (including the first author) manually transcribed the interviews. To ensure transcription accuracy, we carefully examined the data by repeatedly checking back against the original

Table 1: Characteristics of Participants in Our Study

Participant	Gender	ML Exp	ML Job Role	Industry	Organization Size	AutoML Platform
P1	Female	2 yrs	Marketing Manager	Marketing	10 - 100	Commercial
P2	Female	5 yrs	Software Engineer	Finance	1,000 - 10,000	Commercial
P3	Male	8 yrs	ML Engineer	Social Network	1,000 - 10,000	Internal & Commercial
P4	Male	3 yrs	ML Researcher	Healthcare	100 -1,000	Internal
P5	Male	10 yrs	NLP Researcher	Human Resource	10 - 100	Commercial
P6	Male	12 yrs	Head of Technology	Healthcare	100 - 1,000	Commercial
P7	Female	3 yrs	HCI Researcher	University	1,000 - 10,000	Commercial
P8	Male	2 yrs	ML Researcher	Social Network	1,000 - 10,000	Internal
P9	Male	15 yrs	Software Engineer	Information System	50,000+	Internal
P10	Female	6 yrs	Researcher	University	1,000 - 10,000	Internal
P11	Male	3 yrs	Software Engineer	Mobility Service	1,000 - 10,000	Internal
P12	Male	4 yrs	Software Engineer	Technology	50,000+	Internal
P13	Male	16 yrs	Research Manager	Technology	50,000+	Internal
P14	Male	2 yrs	Data Scientist	Travel Technology	1,000 - 10,000	Internal
P15	Male	5 yrs	ML Researcher	Retail	1,000 - 10,000	Commercial
P16	Male	2 yrs	Data Scientist	Healthcare	1,000 - 10,000	Commercial
P17	Male	3 yrs	Researcher	University	1,000 - 10,000	Internal
P18	Male	5 yrs	Software Engineer	Social Network	1,000 - 10,000	Commercial
P19	Male	2 yrs	ML Researcher	Music Streaming Service	1,000 - 10,000	Internal

audio recordings. To provide contextual information, each interview began with open-ended questions: (i) can you tell us about the company/industry you are working in? (ii) can you tell us your current job responsibilities? (iii) how long have you been working in ML-related work? The contextual information helped the transcribers to interpret recordings if they were not the researchers who collected the data.

To discover the main themes of the interviews, we followed an inductive approach [73] to perform thematic analysis [11, 12]. Four trained researchers (two HCI and two ML researchers) were involved in the data analysis. In the first stage, by reading the transcripts independently and repeatedly, we actively searched for the meanings and patterns of the content and wrote down analytic memos. Through this iterative process, we became familiar with every aspect of the data without being selective or skipping over the data following the instructions of thematic analysis [11].

After obtaining the initial understanding of how users apply AutoML in workplaces, we conducted multiple rounds of discussions about our understanding based on the analytic memos. Following a constant comparative approach [32], the process involved moving back and forth between the similarities and differences of emerging categories with reference to the collected data. After that, we individually returned to the data and began assigning basic codes to each idea. In this stage, each researcher coded the data by highlighting and noting the texts to indicate potential patterns. We followed the guidance of coding "as many potential themes/patterns as possible" [11] and generated a list of 209 basic codes. For example, the data extract "we mainly look at how the three learning curves of training, validation, and testing change during the training process and the testing process" was first identified as one of "workarounds for lacking transparency", and further coded as a sub-category of "tracking AutoML's process" of this category. We held regular meetings to discuss and compare the respective codes to note similarities and discrepancies. We held six meetings, each lasting an hour, to address any disagreements related to coding. We used the "Open Discussion" method [18] to resolve these disagreements. During these meetings, we created a table that summarized the codes used by each of the four coders for each quote. The main goal was to discuss and resolve any discrepancies. The coders addressed each disagreement in the order that it appeared on the summary table. Before making a final decision, the coders considered the codes used by other coders for a particular quote and took into account each coder's rationale for using a specific code. Note that we employed codes to facilitate the theory-development process, and avoided relying on inter-coder reliability to ensure all instances of variations can be captured and to prevent potential marginalization of viewpoints [52].

Upon generating the initial coding, we reconvened to compare and discuss the codes and explain how each basic code can be used to represent a potential theme. We then analyzed the codes and decided how different codes can be combined to form a higher-level theme through multiple rounds of discussions. After that, we re-examined the candidate themes and refined the themes to ensure internal homogeneity and external heterogeneity [59]. Lastly, we defined and named the themes and conducted multiple rounds of refinements before generating the final reports. The final satisfactory thematic map includes two primary themes: "acknowledging and working around AutoML's limitations" and "applying AutoML selectively and situationally".

4 FINDINGS

Overall, our study found that participants are well aware of AutoML's inadequacies, including incompatibility with specific task contexts, lack of transparency, and potential privacy issues. However, rather than being impeded by its limitations, they set clear objectives for what AutoML is able to achieve and thus adapt their

Table 2: Summary of Our Findings

Challenge	Workaround/Strategy	Participant
Customizability (§ 4.1.1)	Contextualizing input data	P6, P7
	Incorporating domain knowledge	P3, P9, P13, P15
	Building internal AutoML tools	P6, P11, P17
Transparency (§ 4.1.2)	Validating AutoML's outcomes manually	P2, P10, P16
	Tracking AutoML's process	P3, P8
	Creating customized visualization	P1, P2, P3, P5, P13, P15
Privacy (§ 4.1.3)	Uprooting privacy leakage	P6, P10, P13, P15, P17, P18, P19
	Applying privacy-preserving techniques	P3, P9
	Delegating to legal regulation	P1, P13
	Choosing trustworthy platforms	P2, P4, P5, P8, P9, P17
Use vs. Non-Use (§ 4.2)	Performance-driven (non-)use	P1, P3, P5, P6, P7, P8
	Task-oriented (non-)use	P1, P2, P4, P7, P8, P10, P13, P17, P18
	Context-specific (non-)use	P3, P9, P14, P15, P17

use accordingly. In practice, they adopt many strategies to cope with AutoML's inadequacies to maximize its practical usability. Our findings are summarized in Table 2.

4.1 Acknowledging and Working around AutoML's Limitations

4.1.1 **Tackling Lack of Customizability.** Requesting for more customizability is a sentiment shared by participants. Both P7 and P9 pointed out that existing AutoML platforms are often encapsulated, making it difficult for users to intervene with the automation process or perform fine-grained tuning of the generated results.

> *"There is actually nothing, not so much you can do in, you know, adjusting the parameters, or whatever algorithm they use."* (P7)
>
> *"The cloud-based AutoML doesn't allow the user to export the model or download the model to deploy it on their own machines. There is no such feature now because they are using the most advanced models. Those models are corporate properties and should not be disclosed to anyone else."* (P9)

Moreover, we found that AutoML's lack of customizability is multi-fold and participants need to derive various strategies to tackle this challenge in different scenarios.

Workaround 1: Contextualizing Input Data. As most current AutoML platforms are generic and do not provide the flexibility to configure their inner workings, participants often find lacking the capability to handle context-sensitive tasks. One common workaround is to contextualize the input data by adding "context hints", so that AutoML is able to utilize such additional information to generate context-specific ML solutions (P6, P7). For instance, P7, who is an HCI researcher, conducted a user study to understand the user experience with a voice-based self-tracking application. Due to her limited coding experience, she chose a commercial AutoML platform to provide the natural language processing (NLP) functionality. However, she reflected that the platform was not adaptive to the self-tracking context, and she needed to add additional contextual information to the input data for AutoML to work precisely:

> *"I feel that the AutoML services are not smart enough if you don't give them enough contextual information, they cannot accurately recognize users' voice input. I don't know how to improve that, so I tell my participants to give the system a little bit more contextual information. For example, "7 to 9" is often mistakenly translated into "729", and I asked my participants to say "7 to 9 AM" or "7 to 9 in the morning" so that it can help these systems improve their performance."* (P7)

Workaround 2: Incorporating Domain Knowledge. Another limitation related to AutoML's customizability perceived by participants is that it does not naturally fit the needs of different industries. Participants' workaround is to gather and incorporate domain knowledge into AutoML's optimization objectives (P3, P9, P13, P15). For example, P13, who works in a technology company focusing on providing AI and ML solutions to traditional industries, reported that AutoML was too generic and regarded it as *"a product made by obtaining the greatest common divisor among the needs of all users."* To fit AutoML to industry-specific tasks, he communicated with industry experts and transformed the experts' domain knowledge into AutoML's optimization objectives:

> *"Based on our cooperation with enterprises in traditional industries, the most difficult but valuable thing is how to convert domain knowledge of different industries into your model design. It's actually the most valuable part, but this is definitely something I can't do with AutoML. For example, in the optimization of the supply chain, a relatively reasonable level of inventory should be maintained, if you have no one to tell you about this kind of domain knowledge, you can not make AutoML fit into this specific task. Since either ML or data scientists are not particularly familiar with such issues, it actually requires us to have more communication with industry experts and transform this kind of domain knowledge into the objective in my model, so the bridging work is actually very important."* (P13)

Workaround 3: Building Internal AutoML Tools. AutoML's lack of customizability is also manifested in its limited support for uncommon data types. Correspondingly, participants often opt to build their own AutoML tools (P6, P11, P17). For instance, P11 explained that his company has developed AutoML tools to support tabular data, which is missing on mainstream platforms:

> *"Our company has developed our own AutoML platform. The AutoML platforms provided by companies like Google and Amazon are very mature. However, the functions of their AutoML platforms support generic data such as images and text but do not support tabular data, which our company deals with."* (P11)

Similarly, P17 described that in his company the data comes in different formats and with different features requiring refining the search space, while current AutoML platforms do not provide such configurability:

> *"Because the data in our field can be in many forms and has discrete features, it needs a better representation of the overall data. The process of its correction also needs to be searched. Our (internal) AutoML is designed to be more refined and can handle different kinds of input data."* (P17)

As another example, P6 works at a non-governmental organization (NGO) in Kenya. As the company provides healthcare information and helps patients connect with local medical resources, the ML solution needs to support the local language of Swahili. Therefore, the company is building an internal AutoML platform and is going to switch from the commercial AutoML service to its own platform, which can better support Swahili without sacrificing accuracy due to translation, as well as "*significantly saves money for the company.*"

In summary, building internal AutoML tools is the strategy to deal with special data types or data with unique features, or to provide better-localized solutions.

4.1.2 **Tackling Lack of Transparency.** The lack of transparency is another major concern frequently mentioned by participants. For example, P13 and P17 emphasized that while ML is already a black-box, automating ML adds another layer of "black-boxness"; thus, they perceived AutoML as a "double black-box". The main transparency issues perceived by participants include two aspects: (i) AutoML has limited support to evaluate its *outcomes*; and (ii) it also falls short to provide sufficient information to assess its *process*. Thus, participants have devised various workarounds to assess and evaluate AutoML's outcomes and process.

Workaround 1: Validating AutoML's Outcomes Manually. Several participants (P2, P10, P16) shared their struggles with evaluating and validating AutoML's *outcomes*. For example, P16 pointed out the lack of indicative performance metrics on the AutoML platform he has used:

> *"There is one issue with AutoML, at least according to our experience when cooperating with the NGO. The evaluation metrics it (AutoML) gave were relatively limited. I remember that it only had one metric of 'precision' for classification, but other metrics such as 'F1 score' and 'accuracy' were missing."* (P16)

To cope with such transparency issues, one common workaround by participants is to manually validate AutoML's outcomes using self-selected metrics or checking their backward compatibility with existing ML solutions:

> *"If it's just for the classification, I would just use the provided evaluation metrics like accuracy and some kind of like F1 score precision and recall, some kind of provided metrics; but for tasks like regression, I usually manually check whether the results are reliable."* (P10)
>
> *"In our company, we compare AutoML's results with the previous results. For example, when we want to run a credit score, we first use a well-trained model like our previous model to run to get a batch of results. Then we use AutoML to run the score. How much we can trust AutoML results depends on how different its results are from our previous results. If there is a big difference, there may be problems. If AutoML's results are within our acceptable range, there should be no big problem, but we definitely do a lot of such testing."* (P2)

Workaround 2: Tracking AutoML's Process. Further, most current platforms only provide explanations for AutoML's outcomes (e.g., the importance of different features for the models suggested by AutoML), while its dynamic process (e.g., how the models are actually found) remains fairly vague. However, as several participants (P3, P8) indicated, they equally care about evaluating the dynamics of AutoML's *process* to assess whether it performs as expected.

The reasons for this lack of process transparency on existing AutoML platforms may be multi-fold. For instance, commercial platforms often view the underlying AutoML techniques as proprietary intellectual properties and are unwilling to disclose the internal information. Also, as it often requires sufficient expertise to apprehend AutoML's process, providing the process transparency may be deemed unnecessary for AutoML platforms facing ordinary users. To work around this limitation, participants resort to manually tracking AutoML's learning curves, which are significant indicators of its optimization trajectories:

> *"We mainly look at how the three learning curves of training, validation, and testing change during the training process and the testing process. It may be the parameter settings or hyperparameter settings chosen for each set of AutoML. We check whether these learning curves make sense. If it makes sense, we will probably trust these results."* (P3)

Another way to assess the dynamics of AutoML's process is to compare the difference among multiple runs of this process under varying settings, as P8 shared his practice of this approach:

> *"Basically, I will look at the statistical results, but I could maybe randomly sample several searches, and that's where I got the AutoML algorithms and performance metrics like accuracy or latency, and I try to measure the differences among different algorithms or different*

> *neurons. I think it is a very straightforward way to evaluate the performance of AutoML."* (P8)

Workaround 3: Creating Customized Visualization. Many participants (P1, P2, P3, P5, P13, P15) also recognized the importance of visualization not only for understanding AutoML's inner workings but also for communicating AutoML's outcomes to internal (e.g., team members and executives) and external (e.g., clients and stakeholders) parties:

> *"Our company's internal AutoML platform has a function that I particularly like, it can visualize its running process, especially when there are so many tasks, it can tell us the running conditions of each task, and also tell us the overall comparisons by showing us a table that lists the differences between different tasks. This is quite useful. AutoML is no longer a black box; it can give us some insight that helps us to reduce the unnecessary search space of hyperparameters for this kind of experiment."* (P15)
>
> *"For my current job, we don't have many new models, because companies like us are relatively stable. But if we have relatively new models or features, we will need to explain them to the clients."* (P2)
>
> *"Personally, I don't use visualization very much. I just observe some specific numbers directly. However, if we need to report to clients, it's best to visualize it."* (P3)

However, this functionality is often underdeveloped or even absent on many AutoML platforms. Even when the existing visualization is sufficient for internal communication with experts, it often falls short to convey consumable information to external, non-expert parties who lack relevant backgrounds. To work around this limitation, participants often need to manually visualize AutoML's outcomes based on a set of pre-identified requirements to facilitate communication with external parties.

> *"We have to visualize the explanations manually based on the results we got from AutoML, as the visualization auto-generated by AutoML is ugly, not informative, and not easy to understand . . . We need to make it easy to understand and look professional. There are certain requirements, such as avoiding text-heavy explanations and using more pictures. But internally, for example, within our group, we generally do not need manually created visualization, we can just use whatever features the platforms provide."* (P2)
>
> *"I manually visualize feature selection, which features are more important, the learning process, and learning curve changes in the performance of each model I trained, as the existing AutoML visualization function is simply not helpful."* (P3)

Participants also recognized that one major challenge to creating such explanations is to visualize information in an essential but not overwhelming manner, especially when communicating with external parties with limited expertise:

> *"Convincing people is not that easy from a technical perspective, but if you tell clients too many technical details like why a particular feature has a value of 0.7, people are going to be confused even more, so it's important to strike a balance to provide just enough information in the right way, not too much, not too technical, or too detailed. This can also protect us by preventing other people from copying our idea."* (P5)

4.1.3 **Mitigating Potential Privacy Risks.** In addition to AutoML's functionality limitations, many participants (P1, P3, P4, P5, P6, P9, P10, P11, P13, P15, P17, P18, P19) also expressed serious concerns about potential data privacy issues in using AutoML platforms.

One major concern is whether using AutoML platforms may entail the privacy leakage of training data, which is especially consequential for critical domains (e.g., healthcare and insurance) involving sensitive information such as health history, credit history, and demographic information:

> *"The problem of data privacy is quite serious. Some projects I have done before were related to medical information, which involved patient data . . . The initial data may first be provided by the hospital itself. However, if an institution provides you with a very small amount of data, while we need to do this experiment on a large scale, we must involve patient data provided by different institutions or different hospitals. Then there are privacy risks: First of all, the patient's information cannot be leaked. Second, each hospital may not want its data to be somehow leaked to other hospitals. So privacy is definitely a very important part to consider when it comes to whether to use AutoML or which one to use."* (P18)

Another major concern is whether AutoML-generated ML solutions are subject to potential inference attacks if disclosed to and used by unauthorized parties. For instance, the models generated by AutoML based on confidential medical data carry a significant amount of sensitive information from the original data, while malicious parties, if given access, may infer such sensitive information by reverse-engineering the models:

> *"If I get a parameter after training on the data of a bank or a hospital, I want to use it in the parameter space of other hospitals or other banks. Sometimes the parameter itself can be used to infer what the previously trained dataset looks like, which may cause data leakage."* (P18)

To alleviate the privacy concerns above, participants resort to various workarounds as detailed below.

Workaround 1: Uprooting Privacy Leakage. One straightforward workaround by participants (P6, P10, P13, P15, P17, P18, P19) is to limit privacy leakage at its root. This strategy can be adopted at either the user or organization level. Specifically, at the user level, they purposely collect less sensitive data during the data collection stage before using such data on AutoML platforms:

> *"We are only collecting minimal identifiable data right now such as users' phone numbers because that's how we engage with the users over SMS (Short Message Service). We also have information about health facilities or hospitals when users sign up. [This information] is enough for us to help the users connect to health services,*

> *but we are not collecting other information such as user names, addresses, or ages."* (P6)

In addition, participants (P13, P15, P17) also mentioned that their organizations may have already performed certain data anonymization to protect data privacy before handing over the data:

> *"If our company asked us to compress a model, we won't have too many images due to user privacy, or we may have a lot of data, but we do not have sensitive information such as gender since such information has been masked by the company."* (P15)
>
> *"Basically, the data I get may already be processed in advance. The information that can be stored in the general database is basically not related to any personal information."* (P17)
>
> *"Clients in the healthcare industry, for example, a pharmaceutical company we have collaborated with before, have strong compliance requirements, so they will also do a lot of processing on their side."* (P13)

In general, only uploading non-sensitive data to AutoML platforms greatly reduces the risks of privacy breaches during AutoML's process. On the downside, this strategy may significantly affect data authenticity and negatively impact AutoML's performance, as noted by P11:

> *"Google definitely doesn't want users to worry that their data will be leaked, so they (Google) may mask some data and may use other means, such as protecting the user's ranking layer to protect data privacy, but such techniques actually damage the authenticity of the original data and affect performance."* (P11)

Workaround 2: Applying Privacy-Preserving Techniques. Another workaround mentioned by participants (P3, P9) is to proactively apply privacy-preserving techniques during AutoML's process. Examples include "black-box optimization" [8] that avoids direct access to data, and "federated learning" [88] that constructs ML models using data spread across multiple parties yet without sharing data between different parties:

> *"There are some black-box optimization methods that AutoML does not touch [your data in optimization]. In this case, it can be done at least a little better to guarantee privacy. Another way is through federated learning, which is equivalent to giving data to local users without uploading [the data to the server]. It relies on the local side to do some [AutoML] searches. What AutoML receives is some high-level [data] or metadata instead of data from users' own devices."* (P3)

However, this workaround is not for every AutoML user, as many may lack the necessary technical expertise to apply advanced privacy-preserving techniques.

Workaround 3: Delegating to Legal Regulation. In addition, several participants (P1, P13) who use commercial AutoML platforms referred to data privacy as a legal issue that should be clearly specified in the privacy agreements:

> *"I think it is a legal issue between the platform and the company. Before the company decides to use these platforms, it must clearly state the privacy issue in the confidential agreement. If the AutoML platform violates the regulations, it will be a legal issue."* (P1)
>
> *"Before we collaborate with the companies that provide the AutoML services, we must first make it clear about what data can be shared, and to what extent the data can be shared. For such issues, these (AutoML) companies actually have their own standards and their own legal team will handle such issues."* (P13)

Delegating to legal regulations helps AutoML users clarify their responsibilities and secure their data from a legal perspective.

Workaround 4: Choosing Trustworthy Platforms. Participants reflected contrastive views towards the trustworthiness of cloud-based AutoML platforms in terms of privacy protection. While some participants (P2, P4, P8, P9, P17) raised concerns about the privacy risks of using cloud-based AutoML platforms, other participants (P5) trusted the AutoML services of renowned companies. For example, P4 from a healthcare company explained his strategy of avoiding using cloud-based AutoML platforms when it involves private data:

> *"If using the public dataset to try out the AutoML service, I'm not worried about data privacy; if I'm going to use some private datasets, I will probably not upload the data but run it locally. If I'm using cloud AutoML services, I will not choose to upload all the private data to the cloud server."* (P4)

P9 also echoed P4's concerns:

> *"The people who really have this concern won't even need to ask, they have very strict rules to prevent them from uploading any data to the cloud so this option was rolled out at first glance, so they wouldn't need to ask and this definitely a concern for many companies they don't want to disclose their data, because they have strict rules to upload data to any other servers besides their internal servers."* (P9)

On the contrary, other participants, especially ones from startup companies (P5), prefer reputed cloud-based AutoML platforms (e.g., Amazon AWS) over their own in terms of data privacy protection and believe these large companies are better positioned to protect data privacy given their plenty of infrastructure and personnel resources:

> *"If I'm hosting service on my own server, I'll be very concerned about getting attacks, because as a start-up, we cannot afford to have an onsite, fully dedicated security team. But those bigger companies have teams of experts and engineers that can take care of this."* (P5)

Apparently, this view contradicts that of participants who choose to use internal platforms for risk control, implying the complex landscape of how users choose among different AutoML platforms, which are affected by perceived privacy risks, operational costs, and platform trustworthiness.

4.1.4 Summary. Acknowledging AutoML's limitations, participants exercise user agency to adjust the use of AutoML in their work

including contextualizing input data, incorporating domain knowledge, and building internal AutoML tools; they apply a variety of strategies to assess and evaluate AutoML's outcomes and process to increase its transparency; meanwhile, they also create new ways to combat privacy concerns, ranging from data anonymization to switching between different AutoML platforms.

We further probed participants about where they often seek help to tackle AutoML's deficiencies. Multiple complementary resources are mentioned, including official documentation of AutoML platforms (P1, P2, P4, P5, P7, P10, P18, P19), online forums where ML practitioners gather (P1, P2, P4, P5, P12, P17, P18, P19), online searches (P2), internal training (P11), and personal networks such as friends and colleagues who are experienced in AutoML (P1).

4.2 Applying AutoML Selectively and Situationally

As our study showed, participants are fully aware of AutoML's deficiencies and devise various strategies to work around or mitigate such limitations. Yet, they also understand that it is impractical to completely address all the limitations. Thus, they perform a careful cost-benefit analysis, set clear expectations, and strategically decide whether and when to apply AutoML on a case-by-case basis. Although this is not a direct way to tackle the limitations per se, it is a fundamental strategy to leverage AutoML in complex situations.

4.2.1 Performance-Driven (Non-)Use of AutoML. We found that participants set clear performance expectations and decide how to use AutoML accordingly. For example, P8 explained that he evaluates the performance of AWS AutoML before making the decision of whether to use its service:

> *"I only use AutoML as a mechanism that automatically runs the new inputs through the whole training process in the background and then presents me the results. If the results pass a certain threshold, I can automatically deploy a new model. Amazon AutoML service does a very good job in training the whole model and in all implementations on a certain API (application programming interface)."* (P8)

In certain cases, participants (P1, P3, P5, P6, P7) acknowledged that AutoML may not produce the optimal results; however, as long as the results meet their performance expectations, they are willing to adopt AutoML for its convenience:

> *"It doesn't necessarily mean that I have to find a perfect model when I have a standard for performance, I will continue using it to improve something based on it, and I don't expect anything else."* (P3)
> *"In our company, we only apply AutoML to get the NLP (natural language processing) for our chatbots, and we are satisfied with 80% accuracy or so."* (P6)

This finding corroborated previous studies [19] that data scientists are often interested in using AutoML to produce "good enough" results.

4.2.2 Task-Oriented (Non-)Use of AutoML. According to participants, they also strategically allocate different parts of their tasks to AutoML depending on the specific contexts. For example, for P2, who works in an insurance company, one daily task is to explore the consumer complaint database and classify clients based on their comments on the financial products. Her strategy is to first run the sub-task of text classification internally using her familiar techniques and then assign the other sub-tasks to the AutoML platform:

> *"We need to classify customers based on their complaint records. Part of the task is to perform a classification based on the clients' complaints. We already have a fixed model to use, which we are familiar with, and have more confidence in this process than using AutoML. We use AutoML for the rest of the task."* (P2)

In the same vein, other participants (P1, P4, P7, P8, P18) set clear needs in mind and only use AutoML for sub-tasks that strictly fit AutoML's intended use (e.g., architecture search, model training, and hyper-parameter tuning):

> *"We mainly use AutoML to find specific architectures and also fit parameters."* (P4)
> *"All of those AutoML tools are that automatically help me handle those issues of speech-to-text [translation] and NLP issues, so it saves me time to deliver the app that I want to design for my users in my project."* (P7)
> *"The first thing I need to do is to define the search space for AutoML before I train models. Then I also define the search algorithm. After these, … just use AutoML to build the models for us."* (P17)
> *"I only need AutoML to do the hyper-parameter search."* (P18)

As another concrete example, P13 works for a company in China that provides whole-package solutions to help enterprise users, especially those from traditional industries, leverage AI and ML in their production. AutoML only accounts for a small portion of the whole solution.

> *"For AutoML, it can speed up the process of searching for optimal hyper-parameter settings, but it cannot replace our early stage. The understanding of the problem, the definition of the problem, and the possibility that we keep re-designing our model according to the model architecture. Actually, there are very limited now. If the model is already trained and I want to refine it, then we may be able to use AutoML to search for such an optimal parameter of a hyperparameter. For example, when we cooperate with financial companies, they have the task of automatically searching for and selling investments, which mainly rely on AutoML. In fact, it only accounts for a part of our actual delivery."* (P13)

While the aforementioned participants apply AutoML as part of the solutions in their tasks, other participants (P10, P17) mainly use AutoML for research purposes. For instance, P10 uses AutoML as a baseline check for their own models, while P17 considers AutoML as a promising research topic and focuses on improving its practical impacts:

> *"I use AutoML for doing experiments like comparing the AutoML models we received from Microsoft to the one we trained; we just run some experiments on it."* (P10)

> *"I would consider AutoML as a research topic but would not use AutoML tools at work. The current applications of AutoML are very limited, and I do not use it for my work. It is because the application of AutoML in practice is not very successful, and there is still room for AutoML to improve, that's why it requires more scientific research in this area."* (P17)

Further, we also found that as the functionality offered by different AutoML platforms varies, participants often decide to use a specific platform that best fits their target tasks after performing a careful pro-con analysis:

> *"My question is which one is more convenient for me. For example, Azure does punctuation recognition automatically, while Google AutoML provides some punctuation, but you need to do some work manually. I would just use Azure instead of Google AutoML."* (P7)

4.2.3 Context-Specific (Non-)Use of AutoML. Despite acknowledging its benefits, many participants (P3, P9, P14, P15, P17) also expressed concerns about using AutoML in high-stake contexts (e.g., healthcare). In fact, participants who work in the healthcare industry indicated that they tend to avoid using AutoML but still rely on human-designed models and features for multiple reasons. First, as health data is often highly noisy and complicated, its processing requires domain knowledge and past experience. Second, the cost can be prohibitive if they switch from traditional methods to AutoML. For example, P17 described the current practices of the healthcare company he works for:

> *"In our company, many features are still manually designed and processed with very traditional ML models. In fact, our company does not use the deep learning method, let alone AutoML, because the traditional methods may be more stable. In addition, traditional methods have been used for a long time; they are easier for people to use and may achieve better results. The feature engineering process in health data can be very complicated, and many experts who are already very experienced in this field can better identify this type of data."* (P17)

Furthermore, these non-use cases are not derived from AutoML's performance issues only, but rather reflect an even larger topic of trust issues in AI or automated systems in general, known as "algorithm aversion" [22]. For example, both P15 and P3 expressed concerns about the reliability and trustworthiness of AutoML in critical domains (e.g., medical analysis) and emphasized the importance of human expertise:

> *"People can still not be replaced by machines in some very critical scenarios, such as doing some analysis of medical treatment. For example, read a CT scan or some X-rays to determine whether this person has lung cancer. I think such a thing is difficult to replace by AutoML because it requires human knowledge and the cost will be too high if AI makes a mistake."* (P15)
>
> *"In some medical situations, in which some data itself is more sensitive, many people do not trust ML models designed by people, let alone the ones designed by AutoML. For example, you have 99% human inspections. If it is true that there is a problem with this model in 1% of practical applications, you have no way to hold it accountable."* (P3)

We further probed the reasons behind the distrust in AutoML and found that it is partially due to the challenge of holding AutoML accountable. According to participants (P9, P14), when AutoML is involved in the decision-making process, it is difficult to attribute responsibility when an error occurs or when AutoML's performance falls short of expectations. As P14 explained:

> *"I think things that are simple and highly repetitive can be handed over to AutoML, but the more critical parts, such as decision-making or analysis, still need people because these tools cannot take responsibility. A very key problem is to what extent the achievements of these AI tools represent humans' achievements. This needs to be carefully defined because what AI does is not necessarily what humans will do."* (P14)

4.2.4 Summary. As our study showed, besides working around AutoML's inadequacies, participants also strategically decide whether and when to use AutoML based on their performance-driven, task-oriented, and context-oriented motivations, and pragmatically adjust their use of AutoML to fit their needs and background knowledge.

5 DISCUSSION

Our study reveals the highly heterogeneous nature of how users incorporate AutoML as available resources to help them accomplish their tasks and goals in practice. Users often develop different understandings and expectations of AutoML's capabilities, strategically adjust their use of AutoML technologies and develop pragmatic workarounds when facing challenges. Previous work has mentioned similar challenges (e.g., customizability [87] and transparency [19, 81]) and suggested that resolving these issues requires human intervention. Our study extends previous work by elucidating users' efforts to overcome specific challenges associated with AutoML and balance AutoML with other resources in practice. In addition, our study discovers privacy concerns as a novel issue associated with AutoML use, which has yet to be discussed in the current discourse of AutoML in HCI. Our study unpacks users, as situated actors, who develop workarounds in overcoming the challenges and examines the underlying decision-making process.

5.1 User Agency and Workarounds

According to the workarounds theory [2], "workarounds are fundamentally about human agency, the ability of people to make choices related to acting in the world." In other words, human agency manifests in the development and execution of workarounds to meet individual needs, goals, and expectations [40] and users often define their goals according to complex situations and use different resources to accomplish them [69]. In our study, users' workarounds reflect their needs to take control over AutoML and to restore user agency when facing constraints, challenges, and unmet expectations derived from technological and situational factors. Below, we

detail the active role of user agency in this decision-making process, specifically, how to develop workarounds and how to decide (non-)use of AutoML.

5.1.1 User Agency in Developing Workarounds. The activity theory [84] proposes that users "appropriate" tools to empower them to achieve goals and they often need to combine multiple tools to do so [40]. In our context, how users exercise their agency is reflected in their developing, selecting, and executing workarounds. Further, we observed that the level of user agency hinges upon external factors such as resources available as well as internal factors such as users' ML expertise. For example, while "power users", or users with extensive ML experience, often apply more advanced techniques to verify AutoML's performance and leverage various tools to customize the use of AutoML, users with less ML expertise rely more on the available features of AutoML platforms and are less likely to enact workarounds. This finding demonstrates how users "use their available knowledge to create and execute an alternate path to achieve the goal" [2], and corroborates previous HCI research on that experienced users are superior at recognizing design defaults [23], comprehending functional relationships [15], and developing problem-solving strategies [44]. Therefore, for users with less ML knowledge, we expect an even narrower range of enactment as AutoML solutions become more tightly integrated into the data science workflow. As previous studies reflected that lay users tend to overly rely on ML [16, 17], we should be cautious. Although AutoML is seen as a solution to accelerate "democratizing data science" [62], limited knowledge, experience, or expertise in ML could result in over-trust and over-reliance, which may threaten user agency as users develop cognitive heuristics or mental shortcuts to perceive and use AutoML [71].

5.1.2 User Agency in (Non-)Use of AutoML. Based on the activity theory [40], users' goals, interests, and intentions are the starting points for analyzing situations in problem-solving. In our study, participants perceive AutoML services as part of the tools to accomplish the tasks at hand [42]. Further, according to the expectation confirmation theory [9], the continuous use of technology is largely determined by perceived expectation confirmation. Our findings show that participants' use of AutoML depends on their evaluations of AutoML's capabilities and performance expectations. As long as AutoML's performance and functionality meet such expectations, they will use AutoML in practice. This view is different from previous studies that examined the reactive role of humans from the technical perspective of data science workflows [19, 87]. From our perspective, users play a more proactive role when considering the relationship between AutoML and data science tasks.

In addition, some participants mentioned that they avoid using AutoML in high-stake contexts such as healthcare while having the tendency to trust humans over AutoML and ML in general. In HCI, this phenomenon of "non-use" is also a matter of users' own choice that resemblances the sense of agency [67]. Empirical research offers possible explanations for such decisions. One explanation is "algorithm aversion" that describes users' predisposition that favors humans over automated systems [22]. Another possible explanation is blame attribution, which is widely discussed in human-AI collaboration, especially when moral responsibility is involved [6]. Therefore, such pre-existing attitudes toward automation, in general, could drive the decision-making towards the non-use of AutoML.

5.2 Designing to Mitigate AutoML's Limitations

Our study corroborated existing HCI research on AutoML (e.g., [19, 81, 87]) by confirming that the lack of customizability and transparency are users' two major concerns. In addition, our study also uncovered privacy as an additional significant, non-functional concern, which has not been identified in previous research. Moreover, our study detailed the workaround strategies and users' rationale and practices of applying AutoML selectively and situationally. Below we discuss what insights our findings yield in terms of designing to mitigate the three perceived limitations of AutoML.

5.2.1 Supporting Domain-Specific Customizability. AutoML's lack of customizability is one major limitation perceived by participants. Previous studies reported a similar issue but emphasized the need for more user control over AutoML's process [19, 87]. However, this effort to improve AutoML customizability focuses mainly on giving the user control over each stage of the standard data science workflow due to the tension between full automation and human intervention as identified by the "human-in-the-loop" approach [77]. Our study pointed out that, in addition to providing workflow-focused customizability, it is also important to design for domain-specific customizability.

Our study found that the lack of customizability also derives from the tension between AutoML as the main service and the support for inclusive services for users with special domain needs. Participants frequently referred to AutoML's lack of customizability as its incapability to fit domain-specific situations. This finding calls for more attention to AutoML platforms that are specialized for concrete application domains and can be adapted to specific contexts [33]. HCI research may consider investigating how to design such customization in a user-friendly manner.

5.2.2 Providing Multifaceted Transparency. The second challenge we found for AutoML is its lack of transparency. Previous research has also pointed out this issue, but focused mainly on user evaluation of current visualization tools [19] or the development of new visualization techniques to improve user understanding of AutoML's outcomes [24, 56, 81, 83]. Our results reflected a more nuanced need for AutoML's transparency: users tend to favor different transparency features in evaluating the outcomes and/or process of AutoML. Some users rely on the performance metrics provided by AutoML platforms, while others need to comprehend AutoML's training and selection processes. Such findings corroborated explainable AI (XAI) research indicating that there is no one-size-fits-all solution for transparency and that a more personalized, interactive method of interpreting ML to users that supports user requirements is often necessary [25, 49, 70].

Additionally, current AutoML transparency tools are designed only to support internal validation and understanding of AutoML solutions, while user expectations go beyond this scope. Our study revealed that AutoML users require transparency features not just

for internal evaluation (e.g., model explanations) but also for external communication with clients, which raises the requirement for layman's comprehension through easy-to-understand visualization. This finding echoed previous work that showed that data scientists tend to have high expectations of what transparency tools can achieve that are beyond the tools' capabilities [43]. The gap between user expectations of what AI can do and the designed features in reality may undermine user trust and adoption of AI [45]. To enhance transparency, it is therefore imperative to consider different stakeholders involved in the process of understanding AutoML [60].

5.2.3 Enhancing Data Privacy. Moreover, a new issue of privacy surrounding AutoML has surfaced, which has not yet been discussed in the current HCI literature on AutoML. The striding advances in ML techniques, especially deep learning techniques, are built upon the consumption of massive amounts of data, which are often sensitive by nature, leading to unprecedented privacy concerns. Prior work has investigated possible privacy challenges and risks arising in the ML process [66, 68, 74], showing that both ML models and training data can be the targets of privacy attacks and leak sensitive information. For instance, Fredrikson et al. [28] demonstrated that even if ML service providers (e.g., Google) provide prediction capabilities as query-only services, the attacker is able to reconstruct the confidential ML models via querying such services in a black-box manner. Besides the privacy issues in common with general ML (e.g., the requirement for massive amounts of training data), AutoML also incurs unique privacy risks and challenges. For instance, Pang et al. [57] showed that AutoML-generated models tend to more easily leak sensitive information about training data. Additionally, the practice of AutoML brings in new stakeholders (e.g., AutoML service providers), entailing privacy risks unaccounted for in the literature.

Our study found that most participants concern about privacy while their coping strategies range from non-use of AutoML platforms to privacy-preserving techniques. They indicated that they use AutoML with privacy in mind but do not see it as a reason to avoid using AutoML as they feel that AutoML's perceived benefits (e.g., dependable and fast services) outweigh its potential privacy risks. Laufer et al. [46] referred to this behavior as privacy calculus, a cognitive process in which people estimate future outcomes of current decisions by calculating the costs and advantages of sacrificing some privacy for better outcomes. Previous HCI research has focused on technological or behavioral mechanisms as privacy coping techniques, which are defined as active problem solving based on deliberate cognitive assessments [46]. These coping strategies include evaluating privacy policies [41], adopting privacy-enhancing tools [89], and information withholding [29].

5.3 Supporting Collaborative Work behind AutoML

Previous work has shown that substantial human work is necessary to apply AutoML in enterprise contexts [19]. Humans are "valuable contributors, mentors, and supervisors to AutoML" by providing guidance throughout the data science workflow [87]. Our study found that the need for human support goes beyond this scope, necessitating collaborations among ML experts, data scientists, and domain experts for the successful adoption of AutoML.

Existing AutoML platforms are often designed for generic ML problems and lack capabilities to integrate domain knowledge. To bridge this gap, data scientists could play the role of helping translate domain knowledge into AutoML's process. Existing research has shown that data scientists increasingly work with domain experts to solve complex scientific problems [50]. Our study also corroborated this finding, as stated by participants (P13), that the "bridging role" of data scientists is critical for helping translate domain knowledge into AutoML.

In addition, our findings surfaced how teams within organizations collaborate. For example, to remedy AutoML's privacy issues, it often requires the company's legal team to assess the privacy compliance of AutoML services; also, technical teams need to develop intuitive visualization to explain the ML solutions generated by AutoML to other stakeholders. In theory, a human-in-the-loop paradigm for augmenting the data science workflow can be useful for understanding the types of engagement between humans and machines to ameliorate certain trust concerns. However, we found that human-in-the-loop is limiting since an AutoML correction and refinement loop not only exists within a wider scope of data science processes but also within organizational processes. While the nomenclature of human-in-the-loop is not exclusive to single individuals interacting with AutoML, we argue that the notion of "*humans*-in-the-loops" more accurately captures how AutoML technologies are used within enterprise settings [19].

5.4 Design Implications

First, our study highlighted the heterogeneous nature of real-world data science. Participants reflected that current AutoML platforms are not customizable for certain domains (P7) or have not addressed language localization (P6). One practical implication is to design domain-specific AutoML platforms (e.g., healthcare [82] and finance [1]). Another way to improve compatibility with specific contexts is to enable users to select relevant tasks or contexts [42]. Techniques, such as active learning, that help AutoML capture user preferences may be promising. Also, techniques such as backward compatible learning [38, 63] may help ensure the backward compatibility of AutoML-generated solutions. As participants indicated, domain knowledge is also vital for resolving the customizability issue; thus, it is crucial to develop features that incorporate domain knowledge to support context-specific applications. For example, it may allow users to specify domain knowledge as first-order relations or introduce connections into neural network models based on the logical constraints enforced by such domain relations [48]; it may also allow users to refer to domain-specific sources (e.g., knowledge graphs) to perform data augmentation before feeding the data to AutoML [51].

Meanwhile, our research discovered that the present transparency tools do not yet fulfill the needs of AutoML users. Although some participants are satisfied with the transparency of AutoML's outcomes, others address more concerns about understanding and monitoring AutoML's process. The design of future transparency tools could consider adding different features to provide both static (i.e., results) and dynamic (i.e., process) transparency. In addition, as

some participants reported, there is an increasing need to communicate AutoML results to other stakeholders, such as clients without relevant ML backgrounds, transparency designs could include not only unpacking algorithmic black-boxes, but also addressing how to communicate data and models among teams, products, and services.

Lastly, our research also highlighted the necessity of resolving privacy concerns when implementing AutoML in real-world contexts, which echoed the call for addressing privacy by design [86]. Specifically, our study found that participants with more extensive ML expertise tend to show a higher level of privacy awareness. On the other hand, participants with less expertise and those from enterprises with fewer resources rely on the legal compliance of AutoML platforms. Thus, designers of AutoML platforms could consider more proactive approaches, such as incorporating privacy notices [26] and nudges [14] to make users aware of privacy risks and engage in privacy-enhancing practices.

6 LIMITATIONS AND FUTURE WORK

The study has a few limitations that could inform future research. We discovered that our participants come from a wide variety of backgrounds and have varying understandings of AutoML. Despite the fact that our findings shed fresh light on how users actively adopt AutoML in real-world contexts, our participants are not sufficiently representative to permit comparisons. For example, we recruited our participants mainly through words of mouth, mailing lists within enterprises, and social media, with diversity potentially limited by the nature of such channels. Future research could examine the difference between AutoML users with more varying ML expertise levels or from more diverse domains to further elucidate how their competence and domain needs may impact their perceived challenges and workarounds. Future studies could also expand on the results of our study through quantitative research methods that lead to more generalizable implications. Furthermore, the voluntary participation in our study may result in self-selection bias [61], with participants who consented to participate in our interviews may be more active users compared to those who did not. This opens the questions for future research on comparing active users with non-active users of AutoML. In addition, although our study discovered three main challenges that users develop strategies to cope with (i.e., customizability, transparency, and privacy), future work is needed to systematically investigate other emerging concerns and challenges not covered in this study. For example, how AutoML may magnify the fairness and bias concerns in ML [53] and how users perceive and seek solutions for such difficulties are potential topics that could be explored in future research. Lastly, participants outside the United States revealed novel issues such as lacking local language support from AutoML. Future research could compare AutoML use cases across different countries and examine other factors that may impact the adoption of AutoML (e.g., languages and cultures).

7 CONCLUSION

In this research, we demonstrated how privacy concerns, lack of transparency, and lack of customizability in AutoML affect and complicate real-world data science activities. The study revealed a range of tactics used by AutoML users to control, resolve, and utilize various resources (such as internal AutoML resources, legal teams, and manual checking) to overcome those obstacles in imperfect but ultimately practical workarounds. Understanding the situated and discretionary adoption and use of AutoML opens up new possibilities for research and practices to promote more effective human-automation collaboration in applied data science.

ACKNOWLEDGMENTS

We thank our participants for sharing their thoughts and experiences. We also thank the anonymous reviewers for their valuable feedback. This work is supported by the National Science Foundation under Grant No. 2212323, 1951729, and 1953893.

REFERENCES

[1] Anna Agrapetidou, Paulos Charonyktakis, Periklis Gogas, Theophilos Papadimitriou, and Ioannis Tsamardinos. 2021. An AutoML application to forecasting bank failures. *Applied Economics Letters* 28, 1 (2021), 5–9.
[2] Steven Alter. 2014. Theory of Workarounds. *Commun. AIS* 34 (2014), 55.
[3] Saleema Amershi, Andrew Begel, Christian Bird, Robert DeLine, Harald Gall, Ece Kamar, Nachiappan Nagappan, Besmira Nushi, and Thomas Zimmermann. 2019. Software engineering for machine learning: A case study. In *Proceedings of the International Conference on Software Engineering: Software Engineering in Practice*.
[4] Saleema Amershi, Dan Weld, Mihaela Vorvoreanu, Adam Fourney, Besmira Nushi, Penny Collisson, Jina Suh, Shamsi Iqbal, Paul N. Bennett, Kori Inkpen, Jaime Teevan, Ruth Kikin-Gil, and Eric Horvitz. 2019. Guidelines for human-AI interaction. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*.
[5] 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 ACM Conference on Computer Supported Cooperative Work and Social Computing Companion*.
[6] Alexander Arntz, Sabrina C Eimler, Carolin Straßmann, and H Ulrich Hoppe. 2021. On the influence of autonomy and transparency on blame and credit in flawed human-robot collaboration. In *Companion of the ACM/IEEE International Conference on Human-Robot Interaction*.
[7] Behnaz Arzani, Kevin Hsieh, and Haoxian Chen. 2021. Interpretable feedback for AutoML and a proposal for domain-customized AutoML for networking. In *Proceedings of the ACM Workshop on Hot Topics in Networks*.
[8] Charles Audet and Warren Hare. 2018. *Derivative-Free and blackbox optimization*. Springer.
[9] Anol Bhattacherjee. 2001. Understanding information systems continuance: An expectation-confirmation model. *MIS Quarterly* 25, 3 (2001), 351–370.
[10] Marie-Claude Boudreau and Daniel Robey. 2005. Enacting integrated information technology: A human agency perspective. *Organization Science* 16, 1 (2005), 3–18.
[11] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative Research in Psychology* 3, 2 (2006), 77–101.
[12] Virginia Braun and Victoria Clarke. 2019. Reflecting on reflexive thematic analysis. *Qualitative Research in Sport, Exercise and Health* 11, 4 (2019), 589–597.
[13] Carrie J Cai, Emily Reif, Narayan Hegde, Jason Hipp, Been Kim, Daniel Smilkov, Martin Wattenberg, Fernanda Viegas, Greg S Corrado, Martin C Stumpe, et al. 2019. Human-centered tools for coping with imperfect algorithms during medical decision-making. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*.
[14] Daphne Chang, Erin L Krupka, Eytan Adar, and Alessandro Acquisti. 2016. Engineering information disclosure: Norm shaping designs. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*.
[15] Michelene TH Chi, Paul J Feltovich, and Robert Glaser. 1981. Categorization and representation of physics problems by experts and novices. *Cognitive Science* 5, 2 (1981), 121–152.
[16] Chun-Wei Chiang and Ming Yin. 2021. You'd better stop! Understanding human reliance on machine learning models under covariate shift. In *Proceedings of the ACM Web Science Conference*.
[17] Chun-Wei Chiang and Ming Yin. 2022. Exploring the effects of machine learning literacy interventions on laypeople's reliance on machine learning Models. In *Proceedings of the International Conference on Intelligent User Interfaces*.
[18] Bonnie Chinh, Himanshu Zade, Abbas Ganji, and Cecilia Aragon. 2019. Ways of qualitative coding: A case study of four strategies for resolving disagreements. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–6.

[19] Anamaria Crisan and Brittany Fiore-Gartland. 2021. Fits and starts: Enterprise use of automl and the role of humans in the loop. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*.

[20] Koen Van der Blom, Alex Serban, Holger Hoos, and Joost Visser. 2021. AutoML adoption in ML software. In *Proceedings of the ICML Workshop on Automated Machine Learning*.

[21] Gerardine DeSanctis and Marshall Scott Poole. 1994. Capturing the complexity in advanced technology use: Adaptive structuration theory. *Organization Science* 5, 2 (1994), 121–147.

[22] Berkeley J Dietvorst, Joseph P Simmons, and Cade Massey. 2015. Algorithm aversion: People erroneously avoid algorithms after seeing them err. *J. Exp. Psychol. Gen.* 144, 1 (2015), 114–126.

[23] Hubert L Dreyfus and Stuart E Dreyfus. 1987. Mind over machine: The power of human intuition and expertise in the era of the computer. *IEEE Expert* 2, 2 (1987), 110–111.

[24] 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 International Conference on Intelligent User Interfaces*.

[25] Julia Earp and Jessica Staddon. 2016. "I had no idea this was a thing": On the importance of understanding the user experience of personalized transparency tools. In *Proceedings of the Workshop on Socio-Technical Aspects in Security and Trust*.

[26] Serge Egelman, Janice Tsai, Lorrie Faith Cranor, and Alessandro Acquisti. 2009. Timing is everything? The effects of timing and placement of online privacy indicators. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*.

[27] Radwa Elshawi and Sherif Sakr. 2020. Automated machine learning: Techniques and frameworks. In *Big Data Management and Analytics*.

[28] Matt Fredrikson, Somesh Jha, and Thomas Ristenpart. 2015. Model inversion attacks that exploit confidence information and basic countermeasures. In *Proceedings of the ACM Conference on Computer and Communications Security*.

[29] Andrew Gambino, Jinyoung Kim, S Shyam Sundar, Jun Ge, and Mary Beth Rosson. 2016. User disbelief in privacy paradox: Heuristics that determine disclosure. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*.

[30] 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 International Conference on Intelligent User Interfaces*.

[31] Marco Gillies, Rebecca Fiebrink, Atau Tanaka, Jérémie Garcia, Frédéric Bevilacqua, Alexis Heloir, Fabrizio Nunnari, Wendy Mackay, Saleema Amershi, Bongshin Lee, et al. 2016. Human-centred machine learning. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*.

[32] Barney G Glaser and Anselm L Strauss. 2017. *The discovery of grounded theory: Strategies for qualitative research*. Routledge.

[33] Mojtaba Haghighatlari, Gaurav Vishwakarma, Doaa Altarawy, Ramachandran Subramanian, Bhargava U Kota, Aditya Sonpal, Srirangaraj Setlur, and Johannes Hachmann. 2020. ChemML: A machine learning and informatics program package for the analysis, mining, and modeling of chemical and materials data. *WIREs Computational Molecular Science* 10, 4 (2020), e1458.

[34] Xin He, Kaiyong Zhao, and Xiaowen Chu. 2021. AutoML: A survey of the state-of-the-art. *ArXiv e-prints* (2021).

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

[36] Andreas Holzinger, Markus Plass, Katharina Holzinger, Gloria Cerasela Crişan, Camelia-M Pintea, and Vasile Palade. 2016. Towards interactive Machine Learning (iML): Applying ant colony algorithms to solve the traveling salesman problem with the human-in-the-loop approach. In *Proceedings of the International Conference on Availability, Reliability, and Security*.

[37] Eric Horvitz. 1999. Principles of mixed-initiative user interfaces. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*.

[38] Weihua Hu, Rajas Bansal, Kaidi Cao, Nikhil Rao, Karthik Subbian, and Jure Leskovec. 2022. Learning backward compatible embeddings. In *Proceedings of the ACM Conference on Knowledge Discovery and Data Mining*.

[39] Frank Hutter, Lars Kotthoff, and Joaquin Vanschoren. 2019. *Automated machine learning: Methods, systems, challenges*. Springer.

[40] Victor Kaptelinin and Bonnie A Nardi. 2006. *Acting with technology: Activity theory and interaction design*. MIT Press.

[41] Clare-Marie Karat, John Karat, Carolyn Brodie, and Jinjuan Feng. 2006. Evaluating interfaces for privacy policy rule authoring. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*.

[42] Shubhra Kanti Karmaker, Md Mahadi Hassan, Micah J Smith, Lei Xu, Chengxiang Zhai, and Kalyan Veeramachaneni. 2021. Automl to date and beyond: Challenges and opportunities. *ACM Comput. Surv.* 54, 8 (2021), 1–36.

[43] Harmanpreet Kaur, Harsha Nori, Samuel Jenkins, Rich Caruana, Hanna Wallach, and Jennifer Wortman Vaughan. 2020. Interpreting interpretability: understanding data scientists' use of interpretability tools for machine learning. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*.

[44] Gary A Klein. 2017. *Sources of power: How people make decisions*. MIT Press.

[45] Rafal Kocielnik, Saleema Amershi, and Paul N Bennett. 2019. Will you accept an imperfect AI? Exploring designs for adjusting end-user expectations of ai systems. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*.

[46] Robert S Laufer and Maxine Wolfe. 1977. Privacy as a concept and a social issue: A multidimensional developmental theory. *Journal of Social Issues* 33, 3 (1977), 22–42.

[47] Doris Jung-Lin Lee and Stephen Macke. 2019. A Human-in-the-loop perspective on AutoML: Milestones and the road ahead. *IEEE Data Eng. Bull.* 42 (2019), 59–70.

[48] Tao Li and Vivek Srikumar. 2019. Augmenting neural networks with first-order logic. In *Proceedings of Annual Meeting of the Association for Computational Linguistics (ACL)*.

[49] Q Vera Liao, Daniel Gruen, and Sarah Miller. 2020. Questioning the AI: Informing design practices for explainable AI user experiences. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*.

[50] Yaoli Mao, Dakuo Wang, Michael Muller, Kush R Varshney, Ioana Baldini, Casey Dugan, and Aleksandra Mojsilović. 2019. How data scientists work together with domain experts in scientific collaborations: To find the right answer or to ask the right question? *Proc. ACM Hum.-Comput. Interact.* 3, GROUP (2019), 1–23.

[51] Kenneth Marino, Xinlei Chen, Devi Parikh, Abhinav Gupta, and Marcus Rohrbach. 2020. KRISP: Integrating Implicit and Symbolic Knowledge for Open-Domain Knowledge-Based VQA. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*.

[52] Nora McDonald, Sarita Schoenebeck, and Andrea Forte. 2019. Reliability and inter-rater reliability in qualitative research: Norms and guidelines for CSCW and HCI practice. *Proc. ACM Hum.-Comput. Interact.* 3, CSCW (2019), 1–23.

[53] Ninareh Mehrabi, Fred Morstatter, Nripsuta Saxena, Kristina Lerman, and Aram Galstyan. 2021. A survey on bias and fairness in machine learning. *ACM Comput. Surv.* 54, 6 (2021), 1–35.

[54] 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 CHI Conference on Human Factors in Computing Systems*.

[55] Manfred Musigmann, Burak Han Akkurt, Hermann Krähling, Nabila Gala Nacul, Luca Remonda, Thomas Sartoretti, Dylan Henssen, Benjamin Brokinkel, Walter Stummer, Walter Heindel, et al. 2022. Testing the applicability and performance of Auto ML for potential applications in diagnostic neuroradiology. *Scientific Reports* 12, 1 (2022), 1–11.

[56] Shweta Narkar, Yunfeng Zhang, Q Vera Liao, Dakuo Wang, and Justin D Weisz. 2021. Model LineUpper: Supporting interactive model comparison at multiple levels for AutoML. In *Proceedings of the International Conference on Intelligent User Interfaces*.

[57] Ren Pang, Zhaohan Xi, Shouling Ji, Xiapu Luo, and Ting Wang. 2022. On the security risks of AutoML. In *Proceedings of the USENIX Security Symposium*.

[58] Samir Passi and Steven J Jackson. 2018. Trust in data science: Collaboration, translation, and accountability in corporate data science projects. *Proc. ACM Hum.-Comput. Interact.* 2, CSCW (2018), 1–28.

[59] Michael Quinn Patton. 1990. *Qualitative evaluation and research methods*. SAGE.

[60] Alun Preece, Dan Harborne, Dave Braines, Richard Tomsett, and Supriyo Chakraborty. 2018. Stakeholders in explainable AI. *ArXiv e-prints* (2018).

[61] Oliver C Robinson. 2014. Sampling in interview-based qualitative research: A theoretical and practical guide. *Qualitative Research in Psychology* 11, 1 (2014), 25–41.

[62] Zeyuan Shang, Emanuel Zgraggen, Benedetto Buratti, Ferdinand Kossmann, Philipp Eichmann, Yeounoh Chung, Carsten Binnig, Eli Upfal, and Tim Kraska. 2019. Democratizing data science through interactive curation of ml pipelines. In *Proceedings of the ACM Conference on Management of Data*.

[63] Yantao Shen, Yuanjun Xiong, Wei Xia, and Stefano Soatto. 2020. Towards backward-compatible representation learning. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*.

[64] Ben Shneiderman. 1997. *Designing the user interface: Strategies for effective human-computer interaction*. Addison-Wesley Longman.

[65] Ben Shneiderman and Pattie Maes. 1997. Direct manipulation vs. interface agents. *Interactions* 4, 6 (1997), 42–61.

[66] Reza Shokri, Marco Stronati, Congzheng Song, and Vitaly Shmatikov. 2016. Membership inference attacks against machine learning models. In *Proceedings of the IEEE Symposium on Security and Privacy*.

[67] Patrick Skeba, Devansh Saxena, Shion Guha, and Eric P. S. Baumer. 2021. Who has a Choice?: Survey-based predictors of volitionality in Facebook use and non-use. *Proc. ACM Hum.-Comput. Interact.* 5, GROUP (2021).

[68] Congzheng Song, Thomas Ristenpart, and Vitaly Shmatikov. 2017. Machine learning models that remember too much. In *Proceedings of the ACM Conference on Computer and Communications Security*.

[69] Lucy A Suchman. 1987. *Plans and situated actions: The problem of human-machine communication.* Cambridge University Press.

[70] Yuan Sun and S Shyam Sundar. 2022. Exploring the effects of interactive dialogue in improving user control for explainable online symptom checkers. In *Proceedings of the CHI Conference on Human Factors in Computing Systems.*

[71] S Shyam Sundar. 2020. Rise of machine agency: A framework for studying the psychology of human–AI interaction (HAII). *Journal of Computer-Mediated Communication* 25, 1 (2020), 74–88.

[72] S Shyam Sundar and Sampada S Marathe. 2010. Personalization versus customization: The importance of agency, privacy, and power usage. *Human Communication Research* 36, 3 (2010), 298–322.

[73] David R Thomas. 2006. A general inductive approach for qualitative data analysis. *American Journal of Evaluation* 27, 2 (2006), 237–246.

[74] Florian Tramèr, Fan Zhang, Ari Juels, Michael K. Reiter, and Thomas Ristenpart. 2016. Stealing machine learning models via prediction APIs. In *Proceedings of the USENIX Security Symposium.*

[75] Lukas Tuggener, Mohammadreza Amirian, Katharina Rombach, Stefan Lörwald, Anastasia Varlet, Christian Westermann, and Thilo Stadelmann. 2019. Automated machine learning in practice: State of the art and recent results. *ArXiv e-prints* (2019).

[76] Oleksandra Vereschak, Gilles Bailly, and Baptiste Caramiaux. 2021. How to evaluate trust in AI-assisted decision making? A survey of empirical methodologies. *Proc. ACM Hum.-Comput. Interact.* 5, CSCW2 (2021), 1–39.

[77] Chi Wang, Qingyun Wu, Markus Weimer, and Erkang Zhu. 2021. FLAML: A fast and lightweight AutoML library. In *Proceedings of the Conference on Machine Learning and Systems.*

[78] Dakuo Wang, Josh Andres, Justin D Weisz, Erick Oduor, and Casey Dugan. 2021. AutoDS: Towards human-centered automation of data science. In *Proceedings of the CHI Conference on Human Factors in Computing Systems.*

[79] Dakuo Wang, Q Vera Liao, Yunfeng Zhang, Udayan Khurana, Horst Samulowitz, Soya Park, Michael Muller, and Lisa Amini. 2021. How much automation does a data scientist want? *ArXiv e-prints* (2021).

[80] 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. *Proc. ACM Hum.-Comput. Interact.* 3, CSCW (2019), 1–24.

[81] Qianwen Wang, Yao Ming, Zhihua Jin, Qiaomu Shen, Dongyu Liu, Micah J Smith, Kalyan Veeramachaneni, and Huamin Qu. 2019. Atmseer: Increasing transparency and controllability in automated machine learning. In *Proceedings of the CHI Conference on Human Factors in Computing Systems.*

[82] Jonathan Waring, Charlotta Lindvall, and Renato Umeton. 2020. Automated machine learning: Review of the state-of-the-art and opportunities for healthcare. *Artif. Intell. Med.* 104 (2020), 101822.

[83] 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 International Conference on Intelligent User Interfaces.*

[84] James V Wertsch. 1998. *Mind as action.* Oxford University Press.

[85] Marcus Winter and Phil Jackson. 2020. Flatpack ML: How to support designers in creating a new generation of customizable machine learning applications. In *Proceedings of the International Conference on Design, User Experience, and Usability.*

[86] Richmond Y Wong and Deirdre K Mulligan. 2019. Bringing design to the privacy table: Broadening "design" in "privacy by design" through the lens of HCI. In *Proceedings of the CHI Conference on Human Factors in Computing Systems.*

[87] Doris Xin, Eva Yiwei Wu, Doris Jung-Lin Lee, Niloufar Salehi, and Aditya Parameswaran. 2021. Whither AutoML? Understanding the role of automation in machine learning workflows. In *Proceedings of the CHI Conference on Human Factors in Computing Systems.*

[88] Qiang Yang, Yang Liu, Tianjian Chen, and Yongxin Tong. 2019. Federated machine learning: Concept and applications. *ACM Transactions on Intelligent Systems and Technology* 10, 2 (2019), 1–19.

[89] Jun Yu, Baopeng Zhang, Zhengzhong Kuang, Dan Lin, and Jianping Fan. 2016. iPrivacy: Image privacy protection by identifying sensitive objects via deep multi-task learning. *IEEE Transactions on Information Forensics and Security* 12, 5 (2016), 1005–1016.