# A Multivocal Literature Review on the Benefits and Limitations of Industry-Leading AutoML Tools

Luigi Quaranta[a], Kelly Azevedo[b], Fabio Calefato[a], Marcos Kalinowski[b]

[a]*University of Bari, Bari, Italy*
[b]*PUC-Rio, Rio de Janeiro, Brazil*

---

**Abstract**

***Context*** . Rapid advancements in Artificial Intelligence (AI) and Machine Learning (ML) are revolutionizing software engineering in every application domain, driving unprecedented transformations and fostering innovation. However, despite these advances, several organizations are experiencing friction in the adoption of ML-based technologies, mainly due to the current shortage of ML professionals. In this context, Automated Machine Learning (AutoML) techniques have been presented as a promising solution to democratize ML adoption, even in the absence of specialized people.
***Objective*** . Our research aims to provide an overview of the evidence on the benefits and limitations of AutoML tools being adopted in industry.
***Method*** . We conducted a Multivocal Literature Review, which allowed us to identify 54 sources from the academic literature and 108 sources from the grey literature reporting on AutoML benefits and limitations. We extracted explicitly reported benefits and limitations from the papers and applied the thematic analysis method for synthesis.
***Results***. In general, we identified 18 reported benefits and 25 limitations. Concerning the benefits, we highlight that AutoML tools can help streamline the core steps of ML workflows, namely data preparation, feature engineering, model construction, and hyperparameter tuning—with concrete benefits on model performance, efficiency, and scalability. In addition, AutoML empowers both novice and experienced data scientists, promoting ML accessibility. However, we highlight several limitations that may represent obstacles to the widespread adoption of AutoML. For instance, AutoML tools may introduce barriers to transparency and interoperability, exhibit limited flexibility for complex scenarios, and offer inconsistent coverage of the ML workflow.
***Conclusions***. The effectiveness of AutoML in facilitating the adoption of

Free author version, cite as:
Quaranta, L., Azevedo, K., Calefato, F., Kalinowski, M., A multivocal literature review on the benefits and limitations of industry-leading AutoML tools. *Information and Software Technology*, vol. 178, 2025. DOI: https://doi.org/10.1016/j.infsof.2024.107608

machine learning by users may vary depending on the specific tool and the context in which it is used. Today, AutoML tools are used to increase human expertise rather than replace it and, as such, require skilled users.

*Keywords:* multivocal literature review, automl, autoai, benefits, limitations

## 1. Introduction

Over the last decade, the technological landscape has undergone a significant transformation driven by advances in Artificial Intelligence (AI) and Machine Learning (ML). These innovations have revolutionized a wide range of industries, including automotive, business, and healthcare [1, 2]. The widespread adoption of machine learning has profoundly transformed data analysis and decision-making processes in various sectors. Furthermore, it has significantly impacted software engineering, giving rise to a new discipline called AI Engineering. This emerging field lies at the intersection of Software Engineering and Artificial Intelligence and is focused on the design, implementation, management, and maintenance of AI-enabled systems.

Despite these advances, numerous organizations still struggle to implement vital machine learning initiatives, primarily due to the dependency on highly specialized expertise [3, 4, 5]. Indeed, the demand for machine learning experts and AI engineers has dramatically increased [6] but there are not enough skilled professionals to cover these roles. This phenomenon limits industrial progress and slows down innovation.

AutoML represents a promising approach to reduce the dependence of industry on machine learning professionals, enabling organizations to efficiently adopt ML technologies [7]. Through automation, AutoML tools streamline the execution of several tasks in ML project development – including data cleaning, feature engineering, model selection, and hyperparameter tuning – typically offering faster (and sometimes better) outcomes than manually devised approaches. Hence, AutoML solutions can be crucial in driving machine learning progress by helping organizations integrate ML capabilities into their products. Moreover, AutoML has the potential to support AI engineers in ensuring crucial non-functional requirements of AI systems and their associated development processes, including reproducibility, maintainability, scalability, and reliability. For instance, integrating AutoML into production ML pipelines could enable real-time, autonomous retraining of

ML models, mitigating performance degradation phenomena and ensuring optimal system reliability.

Leading IT giants such as AWS[1], Google[2], IBM[3], and Microsoft[4] have acknowledged the growing business need for the AutoML technology and have created their own AutoML platforms. This increased focus highlights the importance of AutoML as a disruptive technology, which could drastically improve the development and deployment processes of AI-enabled systems.

In this work, our objective is to explore the evidence available on AutoML technologies, synthesizing their reported benefits and limitations. In particular, we aim to focus on the reported experiences with AutoML solutions currently deployed or under consideration in the software industry. Rather than reviewing AutoML research prototypes, our aim is to conduct a comprehensive analysis of the AutoML state-of-practice. The ultimate goal of this research is to equip practitioners with the knowledge necessary to make informed decisions regarding the integration of AutoML and AutoML-derived products into their development processes.

To this end, we conducted a Multivocal Literature Review (MLR), i.e., a form of systematic literature review that includes the analysis of non-peer-reviewed articles (the so-called 'grey literature') along with the academic literature (also known as 'white literature'). In particular, this study aims to contribute by identifying:

- The benefits revealed in scientific investigations on industry-relevant AutoML solutions and the perceived benefits reported by practitioners and companies that have embraced AutoML tools according to grey literature sources.

- The limitations revealed in scientific investigations on industry-relevant AutoML solutions and the perceived limitations reported by practitioners and companies when using AutoML tools according to grey literature sources.

To identify informative sources on industry-leading AutoML solutions, we narrowed our search to white and grey literature that mentions at least

[1]https://aws.amazon.com/machine-learning/automl/

[2]https://cloud.google.com/automl

[3]https://www.ibm.com/topics/automl

[4]https://learn.microsoft.com/azure/machine-learning

one of the AutoML leaders identified in the Gartner Magic Quadrant for Cloud AI Developer Services [8]. These leaders include the aforementioned Microsoft, Google, IBM, and Amazon Web Services (AWS).

By conducting a thorough thematic analysis of the retrieved sources, we have identified 18 benefits and 25 limitations of currently adopted AutoML solutions. The findings highlight that AutoML tools streamline machine learning workflows, simplify various development tasks, and can improve the performance of ML models. They enhance efficiency and scalability and accelerate prototyping, yet face challenges such as limited coverage of the typical ML workflow, insufficient replacement of human expertise, issues with transparency, and handling diverse data. While AutoML can bolster human capabilities, it does not replace the need for skilled ML professionals. A comprehensive understanding of AutoML tools is crucial for their effective application in industry.

The literature review presented in this paper represents a significant advancement in the field of AutoML, being the first secondary study to apply the MLR methodology to this emerging technology. Our systematic approach thoroughly examines the advantages and disadvantages of AutoML, synthesizing a wide range of sources from academic and grey literature published between 2017 and 2022. By complementing peer-reviewed scientific evidence with insights from non-peer-reviewed articles such as technical blogs, we offer a comprehensive and nuanced analysis of the AutoML field as a whole. As a result, this MLR addresses a significant knowledge gap in the literature by providing a multifaceted and in-depth exploration of the challenges and advancements in AutoML. Our analysis covers a wide range of perspectives, presenting a thorough examination of the reported benefits and limitations of AutoML. This comprehensive review aims to serve as a valuable resource for researchers, practitioners, and industries, fostering a deeper understanding of the current state and potential of AutoML. In doing so, we contribute to the foundation for future research and development in this rapidly evolving field.

The remainder of this paper is structured as follows. In Section 2, we briefly summarize the existing landscape of ML and AutoML, and in Section 3 we discuss related work. In Section 4, we describe the protocol followed to carry out our MLR of articles on AutoML. In Section 5 we present the study results, which are then discussed in Section 6. Finally, in Sections 7 and 8, we present the limitations of our research and draw conclusions.

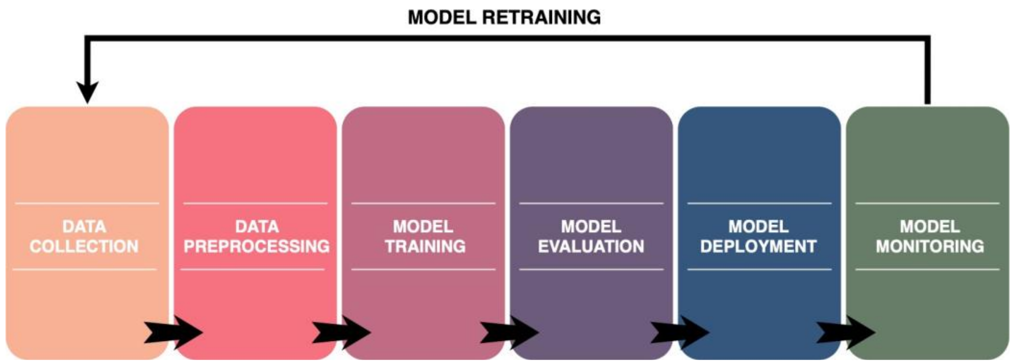



Figure 1: Representation of a simplified Machine Learning workflow (adapted from – "The nine stages of the machine learning workflow" by Amershi *et al.* [9]).

## 2. Background

Industries worldwide are seeing a significant shift, driven by the rapidly growing science of machine learning [1, 2]. This technology has enabled the possibility of automation, optimization, and data-driven insights, enabling computers to acquire knowledge and make predictions or judgments without explicit programming. This section will delve into the concepts of machine learning workflow and automated machine learning and summarize the pertinent existing research in this field that is related to this study.

### *2.1. Machine Learning Workflow*

Machine Learning (ML) workflows are systematic processes that transform raw data into trained models capable of making informed predictions or decisions. These workflows are key to numerous applications, and a typical ML workflow consists of several key steps. In [9], Amershi *et al.* present a nine-stage machine learning workflow, including both data-centric (collecting, cleaning, labeling) and model-centric (requirements, feature engineering, training, evaluation, deployment, monitoring) stages. The study highlights the importance of having several feedback loops, suggesting that the assessment and surveillance of models might impact any previous step. Based on this nine-stage ML workflow and the CRISP-DM industry-independent process model phases [10], we abstracted six generic stages characterizing typical machine learning workflows (Figure 1).

Data collection and data pre-processing are the initial steps. Data is collected, cleaned, and prepared for analysis (*e.g.*, labeled, transformed) during these steps. The quality of data in this phase can significantly impact the final model's performance. The data pre-processing stage also includes feature engineering, which is a creative and knowledge-intensive task involving the selection, transformation, and even creation of features from the training dataset. This process enhances the model's ability to make accurate predictions by making the data more informative and relevant.

The next critical step is model training, which starts with the selection of the ML algorithm that is best suited to solve a particular problem. The decision takes into account the nature of the data and the specific task, whether it is classification, regression, clustering, or another application. The chosen algorithm establishes the foundation for the entire workflow.

With the algorithm selected, the actual training of a model begins. Here, the algorithm learns patterns and relationships within the data; it adjusts its internal parameters to better represent the underlying patterns in the data. After each training iteration, the resulting model is evaluated. Evaluating the model is essential to understand how well it is likely to perform on unseen data. Various metrics, depending on the problem (*e.g.*, accuracy, precision, recall, or F1 score could be used for classification problems while the mean squared error and R-squared could be used for regression problems), are used to assess the model's performance.

The model can be improved by adjusting the hyperparameters. The model does not learn hyperparameters, and they must be configured before training. This step, called hyperparameter tuning, is an iterative process to optimize the hyperparameters of the model. This fine-tuning step optimizes the model's performance, and techniques such as grid search or random search are commonly employed.

After successful training and validation, the model is ready for deployment in a real-world setting. Model deployment can take the form of APIs, embedded systems, or cloud-based services, enabling it to make predictions or automate decisions.

After deployment, a model requires continuous monitoring in production. Model performance can degrade over time due to phenomena like data drift, where the distribution of production data shifts relative to the training data. Timely detection of such performance degradation and triggering of a new iteration of the ML workflow is essential to maintain the reliability of production-grade ML systems.

ML workflows come with a set of challenges. Nahar *et al.* [11] collected and organized overall challenges related to building products with ML-components, some of which are directly related to the ML workflow itself. For instance, data quality is a critical issue, as poor quality data can significantly impact model performance, introducing biases, inaccuracies, and decreasing overall accuracy. Furthermore, effective feature engineering requires domain expertise and an in-depth understanding of the data, making it a complex and creative process. Model selection can also be challenging, as different algorithms perform differently based on the data and task, requiring a deep understanding of the field. Balancing model complexity and generalization is a common challenge, as complex models may overfit training data, while simple ones may underfit, both resulting in poor generalization. Additionally, privacy and security concerns may arise, particularly when handling sensitive information. Scalability may become an issue as datasets grow, requiring additional computational resources. Moreover, complex "black-box" models can be difficult for stakeholders or regulatory bodies to comprehend, which is crucial in sensitive applications like healthcare or finance.

In summary, ML workflows encompass well-defined phases, each associated with its particular set of challenges. Effectively addressing these challenges is essential to successfully deploy ML solutions in real-world applications and derive meaningful insights from the data.

### *2.2. Automated Machine Learning*

Automated machine learning (AutoML) solutions have emerged as a response to the growing demand for machine learning, aiming at simplifying and expediting the model generation process.

AutoML solutions have significantly streamlined the complex landscape of machine learning processes. These tools promise to empower users to build high-quality machine learning models with minimal manual effort, overseeing everything from data pre-processing to hyperparameter optimization [12, 13, 14, 15, 16, 17]. AutoML technologies aim to democratize the field by making machine learning more accessible, even to individuals with limited coding or data science experience [18, 19, 13, 20, 21, 15, 22, 23].

They perform these tasks using a blend of statistical methods and optimization algorithms, which not only save time but also diminish the need for manual intervention. Additionally, some AutoML tools extend their offerings to include features such as model deployment, scaling efficiency [13, 24], and improving efficiency and productivity [19, 21, 14, 15, 17].

The wide variety of AutoML tools and the range of user needs they address are clear indicators of their adaptability. Some tools focus on providing

APIs and libraries that work well with R and Python, while others provide graphical user interfaces to enable the creation of machine learning models without scripting. Open-source AutoML frameworks also give users a lot of freedom to adapt to their needs. Commercial AutoML solutions introduce advanced features, including enhanced scalability and automated model deployment.

By automating routine tasks and allowing the potential elevation of model performance, AutoML tools have the potential to change the game when it comes to machine learning. They are constantly improving and adding new features to allow organizations and individuals to take advantage of machine learning capabilities.

## 3. Related work

Our MLR is a significant effort to provide a comprehensive overview of the field of AutoML research, separating itself from previous studies by its systematic protocol, consolidating key advantages and difficulties of AutoML. In the following, we review the related work in the field of AutoML.

### *3.1. Practical Challenges and Historical Perspectives*

Several researchers have focused on examining the practical challenges of implementing AutoML and providing historical context for its development. These studies offer valuable information on the evolution and current state of AutoML technology.

Elshawi and Sakr [25] provide useful information on the practical challenges of AutoML execution. The authors highlight key issues in the deployment of AutoMLsolutions, such as scalability, optimization techniques, time budgeting, and user-friendliness. Escalante [26] presents a historical perspective on AutoML and its development over the past decade. The authors also emphasize the role of academic challenges in the advancement of the field. The paper concludes by identifying open issues and research opportunities in AutoML, including explainability, feature engineering, handling non-tabular data, and large-scale applications. Santu et al. [21] introduce a seven-tier classification for AutoML systems based on autonomy levels, highlighting the limitations of current solutions that still require significant human involvement. The authors propose a roadmap towards fully autonomous AutoML systems, which would allow domain experts to directly engage with machine learning. The key challenges identified include developing a formal language for the expression of prediction tasks, methods to identify promising tasks, and approaches to the summary and recommendation of results. The paper emphasizes the need for interdisciplinary research to advance AutoML towards complete automation of the end-to-end machine learning process.

### 3.2. Focused Reviews and Analyses

Some studies have reviewed AutoML in specific domains or aspects of the machine learning pipeline. These reviews offer valuable insights but lack a comprehensive overview of the benefits and limitations of AutoML.

Thirunavukarasu et al. [27] compile a list of the clinical uses of AutoML, look at the strengths and weaknesses of the platforms that were used, judge the reliability of the research that tested AutoML, and compare the performance of these platforms with models traditionally created. However, their focus is limited to healthcare. Branco et al. [28] investigated the application of AutoML to electrical biosignal problems, scrutinizing articles from six databases centered on technology and machine learning from 2018 to 2022. They highlight the current challenges in the field, offer insights into biosignal comprehension, and pinpoint the top AutoML solutions. Nonetheless, their focus remains restricted to healthcare. Baymurzina et al. [29] examine the most current research on Neural Architecture Search (NAS), which represents a very specialized field of AutoML tools, and highlight various important concepts and issues that are associated with this topic. Hence, their focus is specifically on NAS algorithms and tools. Valle et al. [30] aim to discover and evaluate AutoML research in the context of multi-label classification and multi-target regression through a systematic literature review. However, their research questions do not involve the benefits and limitations of AutoML tools as a whole. Wen and Li [31] answer questions related to the benefits and limitations of AutoML, but in the area of spatial decision support systems. The main goal of their paper is to analyze the benefits of using AutoML tools in spatial decision support systems.

### 3.3. Broader Reviews and Analyses

Some studies have attempted to provide a broader perspective on AutoML. Nagarajah and Poravi [32] review the current state of AutoML, hyperparameter tuning, and meta-learning. They analyze many methods and evaluate them based on the algorithms they support, the features they provide, and how well they work in practice. However, they did not use a systematic approach when conducting the research. Khalid et al. [33] conducted a systematic literature review of the challenges related to declarative machine learning AutoML solution (which allow users to express their intents through high-level abstractions). They included only 'white' literature (*i.e.*, official peer-reviewed articles published in academic journals or conference proceedings) published until May 2022. They used a database-only search strategy, excluded AutoML solutions that were not declarative and provided limited details on their analysis procedures. Barbudo et al. [34] review the literature on AutoML from 2014 to 2021. The exclusion criteria filtered out papers lacking clear evidence of a blind, peer-review process and those not published in conferences ranked A* or A by the CORE ranking system, as well as papers from non-JCR indexed journals. This study has four research questions. Initially, it seeks to identify commonly used AutoML terms from original research. Secondly, the article takes a quantitative look at the research trajectory inside AutoML to see how it has evolved. Third, it explores the various phases of the knowledge discovery process covered by different AutoML tasks and the various techniques used. Finally, the study identifies emerging trends and unexplored areas, thereby pinpointing potential paths for future research within AutoML. Their study does not focus on revealing the reported benefits and limitations.

## 4. Methodology

### *4.1. Goal and Research Questions*

The goal of this work is to synthesize evidence on the benefits and limitations of AutoML solutions, focusing on those currently adopted in industry. Accordingly, the research questions addressed by this study are:

**RQ1:** *What are the main benefits of state-of-the-practice AutoML tools?*

**RQ2:** *What are the main limitations of state-of-the-practice AutoML tools?*

To address RQ1 and RQ2, we conducted a multivocal literature review [35]. The result of this literature analysis is a comprehensive catalog that highlights the most frequently cited drawbacks and advantages of using AutoML tools.

### 4.2. *Multivocal Literature Review*

A multivocal literature review includes both peer-reviewed academic publications (white literature) and informal documents (grey literature) such as whitepapers and blog posts. Garousi *et al.* [35] showed the benefits of adding grey literature to Systematic Literature Reviews (SLRs) in software engineering. Software engineers rely heavily on grey literature to remain current in their field and share their expertise. Consequently, restricting reviews to academic publications may overlook valuable insights from practitioners. For this reason, we have decided to include grey literature in our research. We will explain our MLR protocol in the following sections.

### 4.3. *MLR protocol*

In May 2023, we searched for articles in three distinct search engines. To obtain scientific publications, we applied the strategy described by Wohlin *et al.* [36], which has shown successful in identifying relevant primary studies in several investigations [37, 36]. This strategy consists of conducting a database search on Scopus[1], applying the study selection process to retrieved candidate papers to gather a fair and representative seed set, and then complementing the primary study identification with iterative snowballing until saturation is reached. To get the grey literature, we conducted an online search using the Google[2] search engine and the Gartner[3] knowledge database.

Hereafter, we provide detailed description of the entire MLR protocol, including the search strategy, selection criteria, data collection, data extraction, and data analysis.

#### 4.3.1. *Search strategy*

Our MLR adopted two distinct search approaches to cover both white and grey literature. These approaches were designed to ensure a thorough exploration of the investigated topic. In the following, we outline the specific strategies we used for each type of literature.

**Search strategy for the white literature**

For the white literature, we followed the search strategy combining database search with snowballing described by Wohlin *et al.* [36]. Hence, we defined a

[1]https://www.scopus.com/
[2]https://www.google.com/
[3]https://www.gartner.com

concise search string to be applied on Scopus in order to retrieve a representative seed set for snowballing. Given that we wanted to retrieve investigations related to AutoML and observe benefits and limitations, the search string for the white literature was defined as shown in Figure 2.

("automl")

**AND**

("benefits" **OR** "challenges" **OR** "limitations")

Figure 2: Final search query for the white literature.

It is noteworthy that, as suggested by the adopted search strategy [36], the string intentionally aimed at representativeness and not completeness, as we wanted to identify a seed set to make use of forward and backward snowballing iterations. This strategy has shown itself effective in identifying relevant primary studies [36].

**Search strategy for the grey literature**

For the grey literature, we used Google Search with a more comprehensive search string compared to the one used for the white literature. This approach was necessitated by the inherent challenges of applying snowballing techniques to the grey literature, which often lacks structured citations and is not systematically indexed. To compensate for these limitations and ensure a thorough exploration of relevant content, we broadened the scope of our search terms by including a wider range of synonyms. Moreover, to focus our search on industry-relevant AutoML solutions, we extended our query to include specific mentions of the AutoML leaders identified in the Gartner Magic Quadrant for Cloud AI Developer Services [8], namely Microsoft, Google, IBM, and AWS. This addition helped filter search results towards discussions of established, industry-ready AutoML solutions, while also excluding content primarily focused on experimental tools that would fall outside the scope of our study. The final search query for the grey literature can be seen in Figure 3.

To complement our grey literature search, we also queried the database of articles offered by Gartner, a well-known research and advisory firm. For this database, we employed a very simple query string, "automl", because

("automl" **OR** "automated machine learning" **OR** "automatic machine learning")
**AND**
("benefits" **OR** "advantages" **OR** "pros" **OR** "strengths" **OR**
"opportunities" **OR** "challenges" **OR** "limitations" **OR** "issues" **OR**
"difficulties" **OR** "cons" **OR** "disadvantages" **OR** "pitfalls" **OR**
"downsides" **OR** "weaknesses" **OR** "drawback")
**AND**
(Microsoft **OR** Google **OR** IBM **OR** AWS)

Figure 3: Final search query for the grey literature.

we noticed that the number of search results did not change when we used a more detailed query.

By implementing these distinct search strategies, we aimed at creating a well-rounded and exhaustive review of both white and grey literature on the benefits and limitations of industry-relevant AutoML tools.

#### *4.3.2. Selection criteria*

Aligned with our research objective, we aimed to include in our review both white and grey literature discussing the benefits and limitations of industry-relevant AutoML solutions —i.e., AutoML tools currently adopted in industry or considered for adoption in real-world industrial applications. To select relevant material for subsequent analysis, we defined a list of exclusion criteria (EC), to be applied to the search results. These criteria are reported in Table 1, along with their rationale.

Among the various exclusion criteria, EC5 deserves a more in-depth explanation. To ensure the industrial relevance of the analyzed material, we decided to base our study on articles mentioning at least one of the leaders in the field of AutoML. To identify these leaders, we took advantage of the Gartner Magic Quadrant, which is both a research method and a graphic representation to rank companies in certain technology markets. It rates vendors based on how well they understand and can implement their vision in a certain industry or sector. In the Magic Quadrant, vendors are collocated into a two-dimensional grid with their ability to execute on the y-axis and their completeness of vision on the x-axis. The quadrant defines four groups: Leaders, Challengers, Visionaries, and Niche Players. The leaders are vendors who usually have a clearly defined vision and are good at putting it

| Search Results Exclusion Criteria | | |
|---|---|---|
| | **Criterion** | **Rationale** |
| **EC1** | Not published between 2017 and 2022. | This criterion narrows the literature down to recent sources, ensuring relevance and up-to-date information. |
| **EC2** | Not written in English. | This ensures that the literature reviewed is in a language accessible to most researchers. |
| **EC3** | Not a journal article, conference paper, book chapter, blog post,* PhD thesis,* technical report,* business report,* white paper,* or another form of substantial technical written content.* | This restriction ensures the inclusion of comprehensive, substantive content from reputable academic and professional sources. It excludes superficial or brief mentions of AutoML tools, such as those found in social media posts or marketing blurbs. Moreover, it excludes all forms of non-written content, such as videos, podcasts, etc. |
| **EC4** | Document not available for our institutions. | This criterion helps in streamlining the literature review process by focusing on sources that the researchers can readily access. |
| **EC5** | It does not include Google, Amazon, Microsoft, or IBM AutoML tools. | This specific criterion ensures that the literature reviewed is aligned with the major AutoML vendors, as per the Gartner Magic Quadrant for Cloud AI Developer Services. |
| **EC6** | Other AutoML-related publications that do not emphasize AutoML benefits, challenges, and limitations. | Our focus is on the benefits, challenges, and limitations of using AutoML tools. |

Table 1: The exclusion criteria and their rationale. Items marked with an asterisk ('*') apply to grey literature only.

into action. People see them as market leaders because they consistently set new standards and develop new ideas. According to this categorization, in our review, we specifically considered the four leaders emerged in the Cloud AI Developer Services magic quadrant [8], i.e., Google, AWS, IBM, and Microsoft (see Figure 4). Our assumption was that articles mentioning at least one of the AutoML solutions from these leaders are more likely to discuss or


Figure 4: Gartner Magic Quadrant for Cloud AI Developer Services [8].

compare state-of-the-practice AutoML solutions.

*4.3.3. Data collection*

For the white literature, we applied the query search in the Scopus database and applied the exclusion criteria. From the remaining articles, we started applying forward and backward snowballing. We applied a hybrid approach (i.e., a combination of database search and snowballing technique) involving backward and forward snowballing iterations to curate our selection of relevant studies. As mentioned by Wohlin et al. [36], the hybrid search strategy is of significant importance in systematic literature studies due to its effectiveness.

Our approach extended beyond Scopus, encompassing Semantic Scholar[1]

[1]https://www.semanticscholar.org/

and Google Scholar[2] for the snowballing process. We started by using Google Scholar for citation retrieval, as it consistently offered a more extensive range of citations than Scopus and Semantic Scholar. After the first round of snowballing, we transitioned to Semantic Scholar because it allows data access through an API, which streamlined the data collection process.

In total, we conducted three rounds of snowballing. We concluded our search in the third round, as no additional articles that met our criteria could be identified, reaching snowballing saturation. Our initial pool of articles comprised approximately 5,700 papers, which can be found in our online open-science repository [38]. Before implementing our exclusion criteria, we performed a validation check on the papers. This entailed checking for duplicates, inspecting references classified as grey literature, and verifying that references or citations contained the keywords "automl", "automated machine learning" or "automatic machine learning" within their titles, abstracts, or keywords.

During the snowballing process, we excluded grey literature, recognizing that a separate Google Search was concurrently conducted to address this specific category of sources. Additionally, to maintain the focus on AutoML, we eliminated papers that did not feature the designated keywords in their title, abstract, or keywords. This decision was motivated by the observation that several articles under analysis pertained to general Machine Learning topics rather than Automated Machine Learning.

For the grey literature, we employed the Google Search query outlined in Section 4.3.1. Upon analyzing the results, we evaluated the exclusion criteria and determined that we would stop the search after three consecutive pages – each containing ten results – with no articles being selected.

#### *4.3.4. Data extraction*

We carefully extracted the data to understand the pros and cons of AutoML tools. Part of the process involved collecting pertinent data and evaluating each research article. Here we detail the main data we collected from each article:

**Code**: We introduced a source tracking identification code, which allowed us to easily link the extracted data to the specific article from which it originated. This code served as a critical reference point for the origin of

[2] https://scholar.google.com/

our findings.

**Title**: The title of each article to quickly identify them.

**Year**: We kept track of the year each article was published so we could see how AutoML tools have changed over time and see if any limitations or benefits have become more apparent.

**Extract**: Parts of the articles' text that address the strengths and weaknesses of AutoML tools. Our subsequent qualitative analysis relied on these extracted passages.

**Type of AutoML tool**: We categorized the mentioned AutoML tools as either commercial, open source, a combination of both, or unspecified. This categorization provided valuable context on the accessibility/availability of the tools and their main benefits and challenges.

**Method**: We distinguished between articles that made unsupported claims about AutoML tools and those that provided evidence-based assessments. This differentiation allowed us to assess the rigor and credibility of the information.

**Examples of referenced tools**: We compiled a list of AutoML tools that were referred to or discussed for each article – this list of tools served as a reference point for identifying trends and popular choices in the field.

By capturing this information from each article, we were able to compile a dataset that forms the foundation of our analysis. The extracted data is also available in our online open-science repository [38].

*4.3.5. Data analysis*

To improve our catalog, we conducted a thematic synthesis [39] of the advantages and disadvantages of AutoML tools. More precisely, we assigned codes to each recognized advantage and drawback. Each code was designed to represent snippets with similar characteristics, making it easier to classify them into broader topics. The process of translating codes into themes required careful consideration of how codes could be combined to create comprehensive overarching themes. As we moved further away from the text, the level of abstraction increased, which improved the ability to apply the concepts to a wider range of situations.

This was not a single step; instead, it was an iterative process. During the various iterations, certain codes defined in previous cycles were incorporated into other codes, reclassified, or eliminated. Advancing in the translation process required rearranging and reclassifying encoded material into various, and occasionally innovative, codes. The process ended once the saturation point was reached in identifying the potential themes that emerged from the data. Mind maps were used to organize the many codes into coherent themes.

The article's second author conducted the initial coding and thematic analysis. The resulting collection of codes and themes was independently peer-reviewed by the other three authors (each one reviewed one-third of the codes) and afterward subjected to collaborative discussion and improvement, including contributions from all authors. The ultimate codes and themes were established by reaching a consensus among the four authors.

### *4.4. MLR protocol application*

Our initial query on the Scopus database with the defined search query for the white literature got 274 results. After applying the exclusion criteria, 24 articles remained. Then, we started the first snowballing iteration. For a comprehensive overview of our three snowballing rounds and the systematic application of our exclusion criteria, refer to Figure 5.

After the first round of snowballing, we had 2,620 articles, of which 1,665 were excluded for being duplicated, grey literature (intentionally not collected during the white literature search strategy), or because they did not contain the word "AutoML" in the title, abstract, or keywords sections. We applied the exclusion criteria in this phase for 955 articles and 23 articles remained.

After the second round of snowballing, we had 2,342 articles, of which 1,461 were excluded because they were duplicated, grey literature, or did not contain the word "AutoML" in the title, abstract, or keywords sections. In this phase, we applied the exclusion criteria for 881 articles and 7 articles remained.

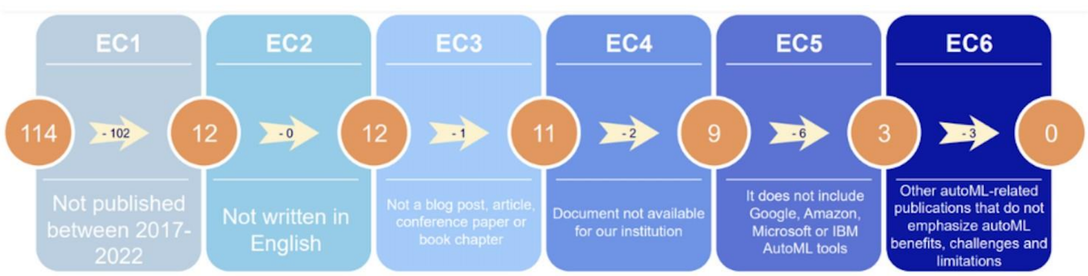
After the third round of snowballing, we had 654 articles, of which 540 were excluded because they were duplicated, grey literature, or did not contain the word "AutoML" in the title, abstract, or keywords sections. We applied the exclusion criteria in this phase for 114 articles and as we included no additional studies, snowballing saturation was reached. In total, 54 white literature studies passed our selection phase.

For the grey literature, Google Search retrieved 737,000 results (in May 2023). We analyzed 199 items from this total to check the exclusion criteria, and we stopped searching further after analyzing three consecutive search pages (with ten items per page) with no relevant articles. A total of 41 items passed the selection process. For the grey literature from the Gartner database, we retrieved 184 articles, of which 117 were excluded after checking for duplicates and the exclusion criteria. 67 articles passed this selection


Application of the exclusion criteria in the first Scopus search results.


Application of the exclusion criteria in the first snowballing round.


Application of the exclusion criteria in the second snowballing round.




Application of the exclusion criteria in the third snowballing round.

Figure 5: Application of the exclusion criteria in the Scopus search results and snowballing.

process. Therefore, a total of 108 grey literature sources passed our selection phase. Details on the number of search results retrieved from each search engine for the grey literature are reported in Figure 6 and Table 2.

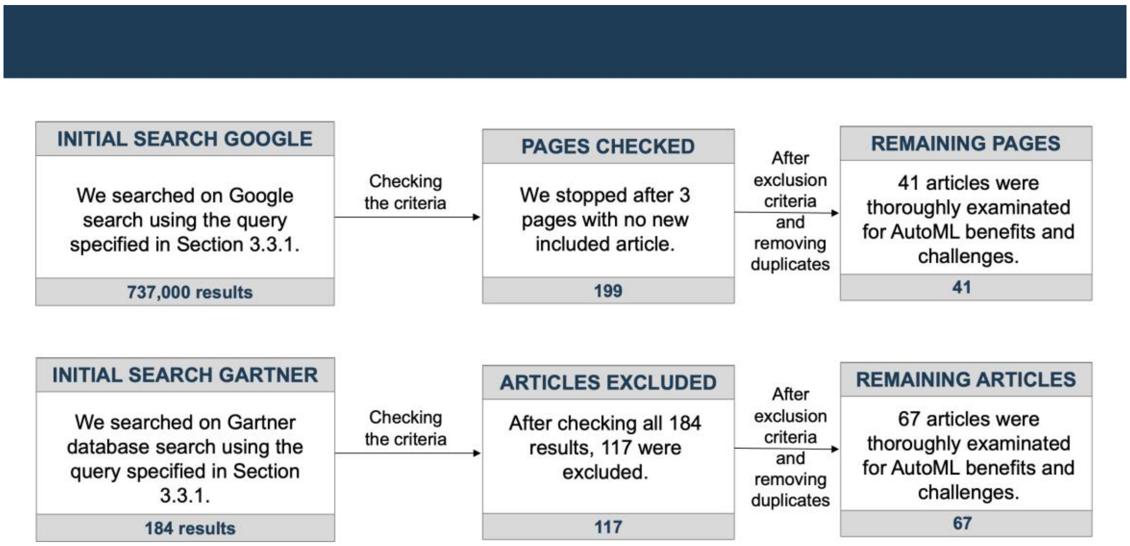



Figure 6: Application of the exclusion criteria in the grey literature (Google Search and Gartner knowledge database).

## 5. Results of the Multivocal Literature Review

### *5.1. Overview of the selected articles*

In this MLR, various articles covering different years were included. The increase in the number of publications on AutoML indicates how popular and useful AutoML tools have become in academia and industry, particularly in recent years. The distribution of the articles analyzed per year is illustrated in Figure 7. To ease the understanding of what kind of evidence we are referring to, in this results section, we cite papers from the white literature with a W prefix and papers from grey literature with a G prefix.

The scientific articles from the white literature examined cover a wide variety of AutoML application domains, demonstrating how versatile AutoML is across various scenarios. Among the several domains covered, we highlight the following.

| Number of excluded articles for each criterion | | | |
|---|---|---|---|
| | **Criterion** | **Excluded Google** | **Excluded Gartner** |
| **EC1** | Not published between 2017 and 2022 | 89 | 18 |
| **EC2** | Not written in English. | - | - |
| **EC3** | Not a blog post, article, conference paper, or book chapter. | 55 | - |
| **EC4** | Document not available for our institutions. | - | - |
| **EC5** | It does not include Google, Amazon, Microsoft, or IBM AutoML tools. | 9 | 78 |
| **EC6** | Other AutoML-related publications that do not emphasize AutoML benefits, challenges, and limitations. | 5 | 1 |

Table 2: The exclusion criteria and the number of articles excluded from the grey literature.

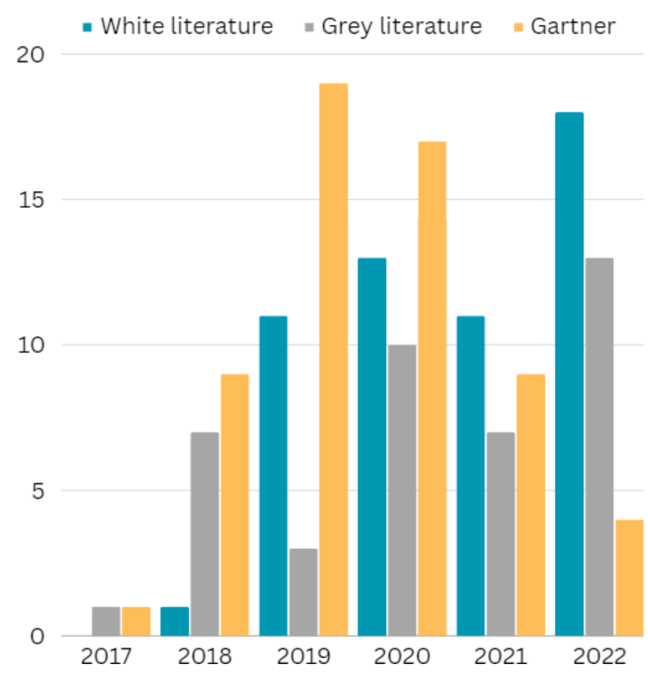



Figure 7: Distribution of the included articles per year.

**Healthcare**: Many of the reviewed articles are related to healthcare analytics, pathology, and clinical decision-making processes, suggesting a strong emphasis on the use of AutoML for these tasks [W40, W18, W41, W42, W24, W43, W44, W16, W45, W22, W23].

**Computer vision**: Demonstrating the application of AutoML algorithms to tasks that involve image analysis, object detection, and picture recognition in different settings [W24, W16, W46].

**Manufacturing**: Showing how AutoML can improve production processes, quality assurance, and preventive maintenance [W47, W14, W48].

**Water quality**: A domain that showcases the use of AutoML in environmental science, emphasizing data-driven approaches to water quality problems [W49].

**Internet of Things (IoT)**: Using AutoML to detect anomalies and perform predictive analytics in IoT applications, for example, smart grids, intelligent vehicles, smart homes, smart agriculture, and smart healthcare [W17].

**Anomaly detection**: Showing how AutoML can be employed for tasks such as fraud detection, intrusion detection, and healthcare system monitoring [W50].

**Sentiment analysis**: Demonstrating the versatility of AutoML methodologies to comprehend and evaluate textual data for sentiment-related tasks [W51].

### *5.2. AutoML tools*

To provide a more comprehensive overview of the AutoML landscape, in Table 3 we offer a detailed comparison of the most relevant AutoML solutions mentioned in our literature review, including both commercial and open-source offerings.

The table breaks down the characteristics of each tool in several dimensions, including infrastructure support, solution type, supported data types, and various stages of the machine learning pipeline. This comparison reveals similarities and differences among the tools. For example, while all tools support data cleaning, feature engineering, model training, and evaluation, there are notable differences in areas such as data labeling and model monitoring.

Cloud-based solutions from major providers like Google, Microsoft, and IBM tend to offer more comprehensive features, including model deployment and monitoring capabilities. These platforms also typically provide web- and API-based interfaces, enhancing their accessibility for different types of

Table 3: A breakdown of the characteristics of the most relevant AutoML solutions mentioned in the literature review (* marks open source solutions).

| **Tool** | **Infra struct.** | **Solution** | **Data types** | **Data clean.** | **Data label.** | **Feature engin.** | **Model traninig** | **Model eval.** | **Model deploym.** | **Model monitor.** |
|---|---|---|---|---|---|---|---|---|---|---|
| AutoGluon* | On prem. | API | img, tab, time series, txt | ✓ | | ✓ | ✓ | ✓ | ✓ | |
| AutoKeras* | On prem. | API | img, tab, txt | ✓ | | ✓ | ✓ | ✓ | | |
| Auto-sklearn* | On prem. | API | tab, time series, txt | ✓ | | ✓ | ✓ | ✓ | | |
| Amazon SageMaker AutoPilot / Canvas | Cloud | API, Web | tab, time series, txt | ✓ | | ✓ | ✓ | ✓ | ✓ | |
| Google Cloud AutoML | Cloud | API, Web | img, tab, time series, txt, vid | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| H2o Driverless AI | Cloud, On prem. | API, Web | img, tab, txt | ✓ | | ✓ | ✓ | ✓ | ✓ | |
| IBM Watson AutoAI | Cloud, On prem. | API, Web | tab, time series, txt | ✓ | | ✓ | ✓ | ✓ | ✓ | ✓ |
| Ludwig AI AutoML* | On prem. | API, CLI | audio, img, tab, time series, txt | ✓ | | ✓ | ✓ | ✓ | ✓ | |
| MS Azure AutoML | Cloud | API, Web | img, tab, time series, txt | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| TPOT* | On prem. | API | tab, time series, txt | ✓ | | ✓ | ✓ | ✓ | | |

users. In comparison, open-source solutions like AutoGluon, Auto-sklearn, and TPOT, while robust in core ML tasks, generally have more limited deployment and monitoring features.

Interestingly, the table highlights a trade-off between flexibility and completeness. On-premises solutions offer greater control and customization, but may lack some of the end-to-end capabilities of cloud-based alternatives. For example, Google Cloud AutoML and Microsoft Azure AutoML stand out for their comprehensive coverage across all analyzed dimensions, including data labeling and model monitoring, which are less common among other tools. Additionally, we note that, while some features may not be directly integrated into AutoML products, they are often available as separate but complementary services within the same ecosystem. For example, although Amazon SageMaker AutoPilot does not include any support automated data labeling or model monitoring, these functionalities are available through other products in the SageMaker suite (i.e., SageMaker Ground Truth for data labeling and SageMaker Model Monitor for monitoring). This modular approach allows greater flexibility and customization in the building of end-to-end ML pipelines.

This analysis provides insight into the current state of AutoML tools, their strengths, and potential areas for improvement. It also underscores the diversity of the AutoML ecosystem, with different tools catering to various needs and use cases.

### *5.3. AutoML benefits*

The codes and themes that emerged from our thematic analysis of the benefits of AutoML are schematically represented in Figure 8. In the following paragraphs, we will go through each theme and provide a description and a few examples for each underlying code.

#### *5.3.1. Infrastructure*

This first theme highlights the ability of AutoML solutions to adapt to different computational needs and seamlessly integrate into various infrastructure setups, enabling organizations to deploy and scale their ML initiatives efficiently.

***Provide scaling efficiency***. This code is supported by nine articles, five from the white literature [W52, W43, W24, W13, W53], and four from the grey literature [G54, G55, G56, G57]. AutoML facilitates efficient scaling of machine learning initiatives, making it easy to handle larger datasets and more complex models. For instance, Das et al. [W13] highlighted that Autopilot streamlines the model building process by offering an automatic hardware recommendation feature; this dynamic allocation of computational resources for each algorithm, dataset, and feature preprocessing pipeline aims to mitigate out-of-memory errors, ensuring seamless operations. By the same token, Ghosh et al. [W43] noted that AutoML platforms offer scalability and enable users to enhance existing models by adding complexity to classifiers. This scalability is especially helpful when updating models to accommodate changing needs and complicated data. Furthermore, as mentioned by Elshawi et al. [W53], AutoML provides users with the flexibility to use the service locally or take advantage of the performance and scalability offered by cloud services. Similarly, in the study by Noshiri et al. [W52], Microsoft Azure ML Studio is recognized for its high scalability, allowing the deployment of in-

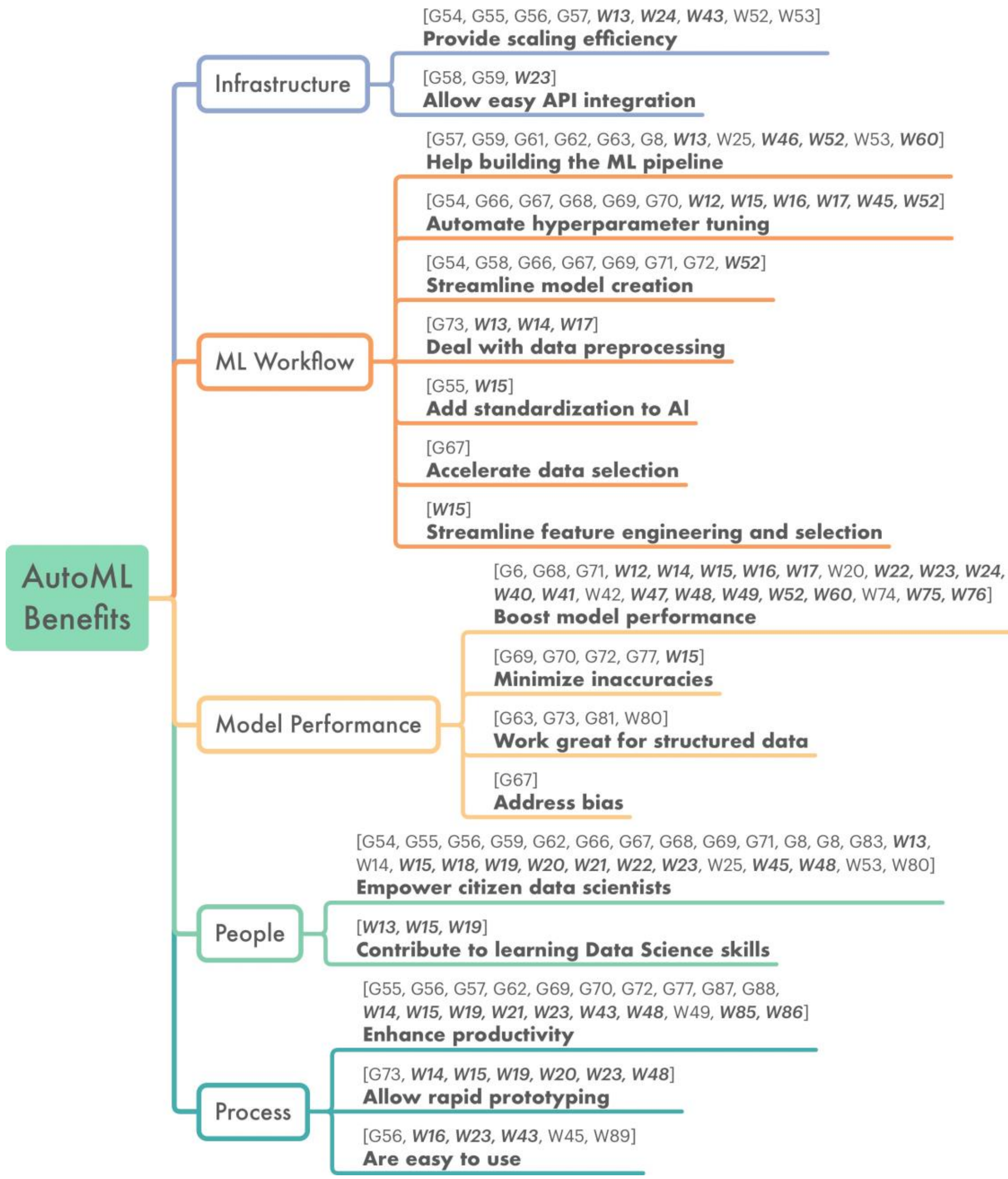



Figure 8: The benefits of using Automated Machine Learning tools. The bold text references are evidence-based extractions; non-bold references indicate unsupported claims.

stances on AWS as well as on-premises. This multi-faceted scalability ensures seamless AutoML operation across various computational environments.

***Allow easy API integration***. Three articles underpin this code, two from the grey literature [G58, G59] and one peer-reviewed paper [W23]. AutoML tools can offer easy integration with various APIs, facilitating the incorporation of ML models into existing workflows and software systems. As detailed by Touma et al. [W22], trained models can be quickly integrated into various applications, including online and mobile apps, without the need for a complex or lengthy setup.

#### *5.3.2. ML workflow*

This theme captures how AutoML tools streamline and enhance the entire machine learning workflow, automating a wide range of tasks from data selection to feature engineering and hyperparameter tuning.

***Help building the ML pipeline***. Twelve articles contribute to this code: six scholarly sources [W25, W52, W46, W13, W53, W60] and six grey literature items [G61, G8, G59, G62, G57, G63]. AutoML can help build end-to-end machine learning pipelines, from data pre-processing to model deployment. This comprehensive support ensures a systematic and integrated approach to machine learning development. According to Xin et al. [W15], AutoML tools not only standardize the ML workflow, but also enhance reproducibility, code maintainability, and knowledge sharing, streamlining collaborative efforts in machine learning projects. For example, as articulated by Das et al. [W13], platforms such as Amazon SageMaker Studio provide data science teams with a unified web-based visual interface, consolidating all the steps of the machine learning process in a single environment. Furthermore, as highlighted by Baker et al. [64], AutoML services play an integral role in the construction of fully integrated MLOps pipelines, reducing the burden on developers to integrate disparate tools and ensuring seamless compatibility among components. Such automated pipelines significantly increase developers' productivity, expediting the delivery of new and enhanced application functionalities. As mentioned by Bonderud [65], AutoML streamlines the complexities associated with building, testing, and deploying novel ML frameworks, thereby simplifying the processes essential for addressing line-of-business challenges. Touma et al. [W22] even found that ophthalmology residents and fellows without coding skills could build effective deep learning models using AutoML. Interestingly, these AutoML-generated models outperformed custom models built by AI specialists.

***Automate hyperparameter tuning***. This code is backed by twelve references: 6 from the white literature [W15, W52, W17, W45, W16, W12] and 6 from the grey literature [G66, G67, G68, G54, G69, G70]. Xin et al. [W15] found that participants in their study commonly use AutoML for tasks such as hyperparameter tuning and model selection during the modeling phase of the ML workflow. Mullen et al. [G58] noted that AutoML services can automatically improve model performance and accuracy by fine-tuning their hyperparameters. According to Wan et al. [W16], AutoML Vision harnesses Google's neural architecture search technology to automatically identify the most effective neural network architecture and hyperparameters, streamlining the model optimization process. Likewise, Schwen et al. [W45] noted that tools such as AutoGluon enable users to define network architectures as well as a thorough parameterization of hyperparameters, offering optimal control over model configurations. In a case study assessing the impact of AutoML at GNP Seguros – an insurance company – the experimented AutoML tool achieved an exceptional success rate of 99.2% by adeptly selecting the most suitable classification algorithm and fine-tuning its hyperparameters, as detailed by Chauhan et al. [W12].

***Streamline model creation***. Eight sources support this code: one academic publication [W52] and seven grey literature articles [G58, G66, G67, G54, G71, G69, G72]. AutoML streamlines the complex process of developing machine learning models, empowering users to build, validate, and deploy them with minimal manual intervention. For instance, Mullen et al. [G58] state that AutoML services make it easier to create datasets for model training; once these datasets are ready, AutoML services rapidly evaluate various ML methods to find the best one for building the desired ML model. By the same token, Baker et al. [G66] noted that AutoML tools analyze data to select optimal methods for model building and improvement.

***Deal with data preprocessing***. This code is supported by four articles: three peer-reviewed sources [W14, W17, W13] and one grey literature piece [G73]. AutoML streamlines various data preprocessing tasks, encompassing the management of missing values, feature scaling, and categorical variable encoding. This not only expedites the process, but also guarantees a uniform and error-free input for machine learning models. As highlighted by Krauß et al. [W14], the use of Auto-sklearn substantially reduced data preparation efforts, requiring only basic data cleaning such as handling NaNs, null columns, and type conversion. Similarly, Das et al. [W13] report that Amazon SageMaker Autopilot adeptly identifies imbalanced binary classification datasets and adjusts the ML pipeline accordingly, resulting in notable improvements in prediction accuracy.

***Add standardization to AI***. A couple sources support this code, one from the white literature [W15] and one from the grey literature [G55]. AutoML platforms can help standardize the machine learning (ML) workflow. According to Xin et al. [W15], AutoML "*Standardizes the ML workflow for better reproducibility, code maintainability, knowledge sharing. Another benefit of the black-box nature of Auto-ML tools is that by having a predetermined search space that does not change, there is more standardization of the ML development process, leading to better comparisons across models, code maintainability, and effortless knowledge transfer.*"

***Accelerate data selection***. One source from the grey literature support this code [G67]. AutoML can accelerate the data selection process by quickly identifying and utilizing relevant datasets. This acceleration is particularly valuable when dealing with large and diverse datasets. According to Batchu et al. [G67], "*today, AutoML offers the following benefits: it reduces the time to identify the best data sources and hyperparameter tuning settings.*"

***Streamline feature engineering and selection***. One source from the white literature support this code [W15]. AutoML simplifies the feature engineering and selection process, allowing users to identify and incorporate relevant features efficiently. This is crucial for optimizing model performance by focusing on the most influential variables. According to Xin et al. [W15], "*feature engineering and feature selection are among the most automated data preprocessing tasks.*"

*5.3.3. Model performance*

This theme encompasses the ways in which AutoML tools contribute to enhancing the overall performance and accuracy of machine learning models. By automating critical aspects of the modeling process, AutoML technologies aim to produce models that are more accurate, robust, and reliable than those developed through manual methods alone.

***Boost model performance***. Twenty articles from the white literature support this code [W47, W14, W49, W15, W42, W23, W74, W52, W41, W17, W16, W24, W22, W12, W40, W20, W48, W75, W60, W76], as well as three sources from the grey literature [G68, G71, G69]. AutoML contributes to enhanced model performance by automating the selection of optimal algorithms, features, and configurations, resulting in models that are better suited to the underlying data. In the study conducted by Xin et al. [W15], many participants, especially ML engineers, emphasize the capability of AutoML tools to rapidly produce superior models. In the study by Unadkat et al. [W23], code-free ML systems outperformed traditional ML object detection systems by using multiple object identification models and selecting the one that exhibits the highest performance. This approach resulted in notable performance gains, underscoring the efficiency of AutoML-driven model selection. Moreover, as detailed by Zeng and Zhang [W24], their AutoML Vision model exhibited slight but noteworthy improvements over previously published models, demonstrating the continuous evolution and refinement achievable through AutoML technologies. In another study documented by Touma et al. [W22], the AutoML model showcased exceptional performance metrics, boasting values in recall (81%), precision (71%), and F1 score (79%), emphasizing its efficacy in achieving a balance between accuracy and robustness. Furthermore, as mentioned by Chauhan et al. [W12], an accuracy rate of 98.1% was attained using AutoML, surpassing the performance of any manually trained models used previously, underlining the substantial performance enhancements facilitated by AutoML. In a separate investigation by Luo and Kindratenko [W76], employing IBM Visual Insights with AutoML, precision and recall rates reached 90% and 100%, respectively, showcasing the exceptional accuracy and reliability achievable through AutoML-driven approaches in specific use cases.

***Minimize inaccuracies***. This code is supported by a total of five sources, one from the white literature [W15] and four from the grey literature [G69, G77, G70, G72]. By automating repetitive tasks and leveraging standardized procedures, AutoML contributes to the creation of more reliable and precise models, ultimately reducing the potential for inaccuracies arising from human intervention. This is stated, for instance, in [G78], in which the authors note that AutoML significantly minimizes the likelihood of inaccuracies due to bias or human errors.

***Work great for structured data***. Four sources underpin this code: one peer-reviewed paper [W79] and three grey literature articles [G73, G80, G63]. AutoML excels particularly in the management of structured data, highlighting its effectiveness in supervised learning tasks, particularly with small to medium-sized datasets characterized by structured formats.

***Address bias***. One grey literature article supports this code [G67]. Some AutoML tools incorporate features that address bias in machine learning models, contributing to more equitable and fair predictions. As highlighted by Batchu et al. [G67], these tools streamline the process by recommending the algorithm most suitable for a specific use case while simultaneously addressing algorithmic bias.

#### *5.3.4. People*

This theme explores the human-centric impact of AutoML tools, focusing on how these technologies democratize access to machine learning capabilities and foster skill development in data science.

***Empower citizen data scientists***. This code is supported by 27 sources: 14 scholarly articles [W25, W21, W14, W15, W23, W79, W45, W13, W22, W53, W20, W48, W19, W18] and 13 grey literature items [G66, G67, G68, G8, G59, G62, G54, G55, G71, G69, G81, G56, G82]. AutoML empowers individuals without extensive machine learning expertise, allowing citizen data scientists to harness the capabilities of ML for their specific use cases. Elshawi and Sakr [W25] highlight that Google AutoML facilitates the training of a diverse range of machine learning models in various domains, offering a user-friendly experience with minimal technical complexity. According to Santu et al. [W21], increased automation enables domain experts to effectively use machine learning technologies. This sentiment is echoed by Unadkat et al. [W23], who emphasize Google AutoML's target users from non-technical backgrounds. Moreover, Das et al. [W13] underscore how Autopilot demystifies machine learning for end users lacking expertise in the field, offering a starting point for applying the predictive capabilities of ML to business problems. It helps users understand the tangible value of ML in specific scenarios, bypassing the costlier and riskier alternative of hiring professional data scientists. In a study by Schwen et al. [W45], classifiers built with AutoML tools performed comparably to the findings in the literature. In particular, these classifiers were created using generic presets with minimal interaction, diverging from prior work that employed task-specific network architectures and optimized training procedures. Additionally, as described by Borkowski et al. [W40], both Google Cloud AutoML and Apple Create ML contribute to the democratization of machine learning. These systems demonstrated robust performance in trained lung and colon cancer diagnostic models.

***Contribute to learning data science skills***. Three peer-reviewed publications support this code [W15, W13, W19]. AutoML can also serve as an educational tool, helping users learn data science concepts and techniques through hands-on experience with automated processes. For instance, Xin et al. [W15] noted that users who gained access to the search history of AutoML tools reported significant learning experiences. This included insights into new modeling techniques, the implementation of specific ML algorithms, understanding of model architecture, evaluation of model performance across distinct tasks, and understanding of model resource consumption. This first-hand exploration through AutoML not only fosters familiarity with diverse data science techniques, but also provides valuable insights into the practical application of machine learning, contributing significantly to users' skill development in this field.

#### *5.3.5. Process*

This theme explores how AutoML tools optimize the machine learning development process. By automating time-consuming tasks, facilitating rapid prototyping, and offering user-friendly interfaces, these tools streamline workflows and reduce barriers to entry in machine learning projects.

***Enhance productivity***. A total of 20 references support this code: ten from the white literature [W21, W14, W49, W83, W15, W84, W23, W43, W48, W19] and ten from the grey literature [G85, G62, G55, G86, G69, G77, G56, G57, G70, G72]. By automating repetitive and time-consuming tasks, AutoML can significantly improve overall efficiency and productivity in the machine learning development process. As noted by Crisan and Fiore-Gartland [W20], people with high technical expertise, such as data scientists, can use AutoML systems to accelerate routine tasks, thus improving the speed and efficiency of their workflows. Furthermore, according to Santu et al. [W21], AutoML tools substantially increase the productivity of data scientists by automating a considerable portion of manual work, allowing them to focus on more complex aspects of model development and analysis. This aligns with the observations by Venkata Vara Prasad et al. [W49], who noted that AutoML plays a vital role in democratizing AI by alleviating redundant human labor, thus improving efficiency in both time and results. Xin et al. [W15] emphasized that a primary benefit reported by many participants in their study – especially ML engineers – is the ability of AutoML tools to accelerate model-building processes while ensuring superior model quality. Likewise, Kramer et al. [W48] observed that AutoML's ability to significantly reduce the time required for data analysis, algorithm training, and optimization further underscores its role in expediting the overall machine learning pipeline.

***Allow rapid prototyping***. This code is backed by seven articles, six of which are academic publications [W14, W15, W23, W20, W48, W19] and only one a grey literature piece [G73]. AutoML is particularly advantageous in the prototyping phase of an ML project, allowing rapid experimentation and iteration in model development. It can potentially reduce deployment costs by streamlining the development process, making machine learning more accessible without substantial financial investments. According to the findings of Krauß et al. [W14], `Auto-sklearn` serves as an initial platform for data scientists, offering both a foundational groundwork for manual implementation and a springboard for further enhancements to their solutions. Additionally, as highlighted by Xin et al. [W15], AutoML's influence has significantly reduced the entry barrier by empowering users to develop prototypes fast. These prototypes serve as invaluable tools for assessing the feasibility and potential impact of machine learning applications. In a study conducted by Crisan and Fiore-Gartland [W20], participants actively used AutoML technologies to rapidly prototype and design viable solutions for data preparation and analysis, underscoring its pivotal role in accelerating the ideation phase. Moreover, as mentioned by Kramer et al. [W48], AutoML plays a dual role by serving as a prototyping tool while seamlessly integrating into various business processes, demonstrating its versatility and applicability across different domains.

***Are easy to use***. Evidence supporting this code is drawn from six articles: five from the white literature [W23, W87, W43, W45, W16] and one from the grey literature [G56]. Thanks to user-friendly interfaces, AutoML tools can help make machine learning accessible to a wider audience, fostering collaboration between domain experts and data scientists. As noted by Unadkat et al. [W23], code-free ML systems present significant advantages, including ease of use, cost-effectiveness, and a visually intuitive interface that displays true/false positives and negatives. Furthermore, as highlighted by Hayashi et al. [W87], AutoML tools stand out for their user-friendly interfaces, which require minimal machine learning expertise to effectively train models. This characteristic empowers a wider spectrum of users, allowing them to engage in model development and analysis without extensive technical knowledge. Consistently, as mentioned by Wan et al. [W16], the user-friendly interface of `AutoML Vision` shows great potential in clinical practice, helping physicians in decision making processes. The intuitive design and accessibility of such tools contribute significantly to their adoption and usability across diverse domains, ultimately fostering collaboration between domain experts and technical specialists.

### *5.4. AutoML limitations*

The codes and themes that emerged from our thematic analysis on the limitations of AutoML are schematically represented in Figure 9. In the following paragraphs, we will go through each theme and provide a description and a few examples for each underlying code.

#### *5.4.1. Data*

This theme explores the limitations of AutoML tools in handling various data-related challenges. Despite their many advantages, these tools face constraints when processing large-scale datasets, navigating complex scenarios, and supporting certain specialized data types.

***Have data size constraints***. This code is supported by seven articles: four from the white literature [W25, W17, W53, W26] and three from the grey literature [G62, G56, G88]. Challenges arise when dealing with large datasets, as AutoML can face limitations in processing large volumes of data efficiently. As observed by Escalante [W26], the resolution of large-scale problems remains an ongoing challenge for contemporary AutoML solutions. As further discussed by Yang and Shami [W17], the application of AutoML models to large-scale datasets proves to be challenging due to the need for multiple training iterations to identify the optimal solution.

***Do not work well for complex scenarios***. This code is supported by nine articles, including five academic publications [W14, W49, W15, W42, W51] and four grey literature items [G89, G67, G62, G69]. AutoML's effectiveness may be limited in complex scenarios, where the intricate nature of

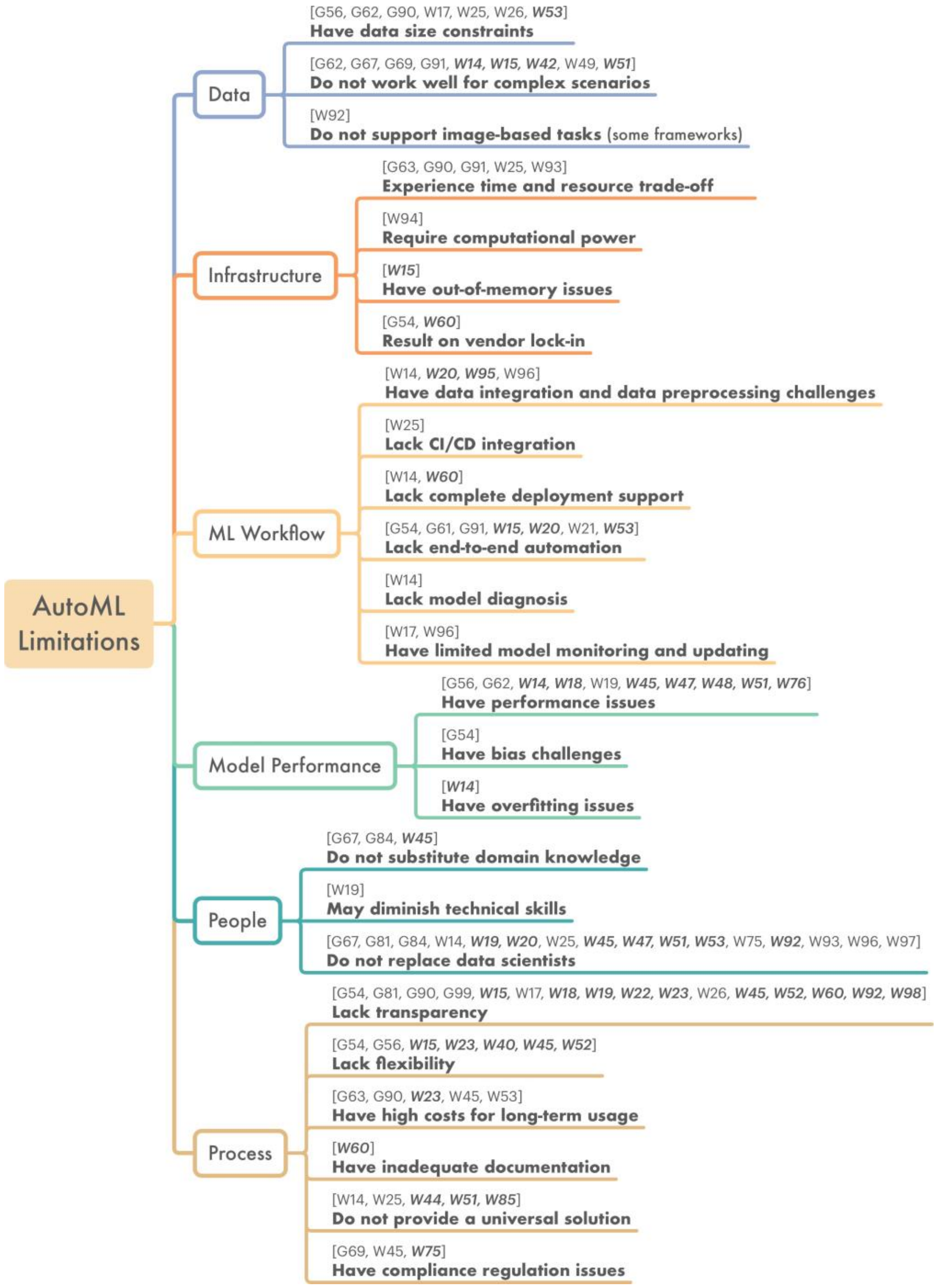



Figure 9: The limitations of using Automated Machine Learning tools. The bold text references are evidence-based extractions; non-bold references indicate unsupported claims.

data and the lack of clear problem structures entail significant challenges. As highlighted by Xin et al. [15], human intervention often compensates for the shortcomings of AutoML, thereby improving its overall performance. However, these limitations become clear when AutoML confronts non-standard use cases and domains, struggling to adapt its predefined frameworks to the unique complexities presented in such scenarios. For instance, Waring et al. [W42] showed how prevailing AutoML approaches exhibit limitations, particularly in managing the scale and diversity of data within biomedical environments.

***Do not support image-based tasks (some frameworks)***. This code is supported by a single research paper [W90]. Limited support for image-based tasks restricts AutoML's applicability in image-reliant domains. As Siriborvornratanakul [W90] notes, leading AutoML frameworks still lack comprehensive support for certain image-based tasks like segmentation.

*5.4.2. Infrastructure*

This theme delves into the infrastructure-related challenges that organizations face when implementing and operating AutoML solutions. While these tools offer significant benefits, they also present notable hurdles in terms of resource management, computational requirements, and technological dependencies.

***Experience time and resource trade-off***. Five articles contributed to this code: two from the white literature [W25, W91] and three from the grey literature [G89, G63, G88]. AutoML processes often involve a trade-off between time and computational resources, necessitating careful consideration of resource allocation to achieve the desired results efficiently. As articulated by Elshawi and Sakr [W25], a larger time budget corresponds to prolonged waiting periods and increased consumption of computing resources — which in turn results in higher costs, especially when using cloud-based resources. In contrast, a smaller time budget reduces waiting periods and costs, but decreases the likelihood of obtaining the optimal recommendation. This trade-off necessitates strategic decision-making to balance optimal outcomes and computational costs.

***Require computational power***. This code is supported by a single article from the white literature [W92]. Symeonidis et al. [W92] emphasize that AutoML processes require considerable computational resources to function effectively, especially for tasks such as hyperparameter tuning and model training.

***Have out-of-memory issues***. Also this code is supported by a single research paper [W15]. Resource consumption challenges extend to memory constraints, as large datasets or complex models can lead to memory issues during AutoML processes. A prevalent concern of participants in the study by Xin et al. [W15] using OSS AutoML solutions is that compute-intensive workloads frequently result in system failures. Specifically, participants reported that encountering limitations in main memory capacity constituted one of the primary technical challenges while using AutoML.

***Result on vendor lock-in***. Two articles underpin this code, one from the white literature [W60] and one from the grey literature[G54]. The potential for vendor lock-in emerges as a challenge, with limited interoperability between AutoML solutions, restricting organizations from easily transitioning between platforms. Xin and Meertens [G54] observed that AutoML solutions are often the source of frustrations typically associated with cloud services, such as a lack of customizability, susceptibility to vendor lock-in, and an opaque operational process.

#### *5.4.3. ML workflow*

This theme explores the limitations of AutoML tools across various stages of the machine learning workflow. While these tools excel in certain aspects of the ML pipeline, they often fall short in providing comprehensive end-to-end solutions. From challenges in data integration and preprocessing to gaps in deployment support, model diagnosis, and ongoing monitoring, AutoML tools present several areas for improvement.

***Have data integration and data preprocessing challenges***. This code is supported by four articles, all of which are from the white literature [W14, W20, W93, W94]. Integrating data from various sources can be a challenge for AutoML, particularly when dealing with heterogeneous datasets that require careful preprocessing and harmonization. As highlighted by Krauß et al. [W14], the current landscape lacks systems capable of effectively automating the data integration phase. Successful integration requires a comprehensive knowledge of the types, structures, and nuances of the data sources involved.

***Lack CI/CD integration***. A single academic publication supports this code [W25]. Integration challenges with continuous integration/continuous deployment (CI/CD) pipelines hinder the incorporation of AutoML into agile development workflows. According to Santu et al. [W21], AutoML solutions predominantly concentrate on hyperparameter tuning and feature engineering, often neglecting comprehensive considerations from a software engineering and integration standpoint.

***Lack complete deployment support***. Two white literature papers support this code [W14, W60]. Often, AutoML tools lack support for deploying models in diverse production environments, hindering the transition from development to real-world ML applications. Alamin and Uddin [W60] found that a significant portion of AutoML practitioners, approximately 13%, encounter difficulties in model deployment. This issue is critical as AutoML is also expected to streamline model deployment and mitigate stability-related problems in production.

***Lack end-to-end automation***. This code is supported by seven articles, including four peer-reviewed papers [W21, W15, W53, W20] and three articles from the grey literature [G61, G89, G54]. AutoML's focus on specific aspects of the machine learning pipeline results in partial end-to-end support, requiring manual intervention for certain stages of model development and deployment. Krauß et al. [W14] observe that existing AutoML systems encompass various ML pipeline steps, but none of them cover the entire spectrum. Likewise, according to Xin et al. [W15], AutoML focuses mainly on automating model training, leaving users responsible for data pre-processing and post-processing tasks.

***Lack model diagnosis***. A single peer-reviewed paper underpins this code [W14]. Diagnosing and understanding model behavior, particularly in complex scenarios, remains challenging, limiting AutoML-generated models' trustworthiness. In particular, Krauß et al. [W14] note that the comprehensive diagnosis of models remains a challenging area that requires further attention and improvement.

***Have limited model monitoring and updating***. Two papers from the white literature support this code [W17, W94]. AutoML tools often lack robust model monitoring and updating mechanisms, crucial for adapting to evolving data patterns and maintaining performance in production environments. Wang et al. [W94] note that while AutoML excels in modeling and data analysis tasks, it lacks automation for labor-intensive activities like data preparation and ongoing model monitoring.

#### *5.4.4. Model performance*

This theme addresses the limitations of AutoML tools in achieving consistently optimal model performance across various scenarios. Despite the promise of automation in machine learning, AutoML systems face several challenges that can impact the quality and reliability of the models they produce. These issues range from suboptimal performance in certain tasks to difficulties in addressing bias and overfitting problems.

***Have performance issues***. Evidence for this code is supported by ten articles, eight from the white literature [W47, W14, W45, W48, W19, W18, W76, W51] and two from the grey literature [G62, G56]. AutoML models may not always achieve optimal performance due to suboptimal automated choices in various tasks, including algorithm selection, hyperparameter tuning, and feature selection. For instance, Krauß et al. [W14] found that manual data preparation performed by data scientists resulted in superior model performance. Luo and Kindratenko [W76] observed that specific tasks, such as differentiating between images of viral and bacterial pneumonia, pose considerable challenges for AutoML-generated models, leading to less favorable outcomes in performance metrics. Similarly, in the study by Faes et al. [W18], AutoML-generated deep learning models exhibited poor performance in certain multi-label classification tasks, possibly due to peculiarities within the training datasets.

***Have bias challenges***. A single grey literature article supports this code [G54]. Despite AutoML's aim to streamline model development, the identification and correction of biases remains an open challenge. Xin and Meertens [G54] note that while AutoML providers are taking steps to embed anti-bias processes and explainability, significant improvements are still needed. Revealing the model's inner workings and identifying potential sources of suspect outcomes are pivotal in fostering trust and, in case of issues, facilitating remedial actions. The mere involvement of humans does not ensure the absence of bias, so model transparency is essential to mitigate biases in decision-making processes.

***Have overfitting issues***. A single paper from the white literature supports this code [W14]. Despite automated hyperparameter tuning, AutoML models may still be prone to overfitting, especially when dealing with complex datasets or a limited amount of training data. As noted by Krauß et al. [W14], the community reports overfitting as an issue associated with Auto-sklearn, often occurring when an excessive amount of time is allocated to the training process.

*5.4.5. People*

This theme examines how AutoML tools, despite their advantages, cannot fully substitute domain-specific expertise, may potentially impact the development of technical skills, and do not eliminate the need for skilled data scientists.

***Do not substitute domain knowledge***. Three sources support this code, one from the white literature [W45] and two from the grey literature [G67, G82]. AutoML tools do not substitute the need for domain knowledge, and successful model development still requires deep understanding of the intricacies of a problem's specific domain. For instance, Schwen et al. [W45] emphasize that combining domain knowledge with ML expertise remains essential for connecting diagnostic tasks, assessment metrics, and the associated machine learning activities.

***May diminish technical skills***. A single academic paper supports this code [W19]. Wang et al. [W19] report skepticism about AutoML's widespread adoption potentially weakening data scientists' technical skills. The participants in their study expressed concerns about future professionals overrelying on automated tools, potentially undermining the development of essential technical competencies. This shift might hinder the development of in-depth technical knowledge and skills crucial for innovation in machine learning. Balancing the use of automated tools with technical expertise cultivation is vital for the sustained advancement and evolution of data science.

***Do not replace data scientists***. This code is supported by sixteen articles: thirteen from the white literature [W25, W47, W14, W45, W53, W20, W75, W19, W51, W91, W95, W94, W90] and three from the grey literature [G67, G80, G82]. While AutoML enhances the efficiency and capabilities of existing data science teams, it falls short of enabling complete ML autonomy for companies lacking specialized in-house talent. Its inability to replace ML experts represents a limitation in its potential to democratize AI, particularly given the current scarcity of skilled ML professionals. AutoML tools cannot replace the role of data scientists; rather, they complement their work by automating specific tasks. Skilled data scientists remain essential for effective model development and interpretation. As noted by Polzer and Thalmann [W75], AutoML primarily serves exploratory purposes, while the creation of comprehensive AI systems still largely relies on the expertise of AI professionals. Siriborvornratanakul [W90] highlights that the use of human knowledge significantly reduces the computational resources and time required to obtain a high-performance model from the vast search space considered in AutoML solutions. Crisan and Fiore-Gartland [W20] also emphasize that the practical use of AutoML in real-world settings requires substantial human effort for effective deployment and utilization. Wang et al. [W94] observed that, in practice, current AutoML systems predominantly target technical personas like data scientists and AI/ML Ops engineers. Likewise, Lee et al. [W95] acknowledge that, despite the automation provided by AutoML, many real-world scenarios still require human supervision.

#### *5.4.6. Process*

This theme explores the process-related challenges that organizations face when implementing and using AutoML solutions. From transparency and flexibility issues to concerns about long-term costs, inadequate documentation, and regulatory compliance, AutoML tools face a range of obstacles in their practical application. These challenges underscore the complexity of integrating AutoML into existing processes and highlight the need for careful consideration in their implementation.

***Lack transparency***. A total of sixteen sources support this code, including twelve papers from the white literature [W15, W23, W52, W17, W45, W22, W60, W19, W18, W26, W90, W96] and four grey literature items [G97, G54, G80, G88]. AutoML's black-box nature poses a significant challenge, as the lack of transparency in its decision-making processes limits user understanding and trust in automated outcomes. As expressed by Xin et al. [W15], the lack of configurability and transparency is evident in tools such as Google Cloud AutoML. Despite leveraging proprietary Google Research technology to enhance model performance, this platform restricts user input on model types and fails to offer visibility into model internals, impeding users' ability to understand model functioning. Similarly, as articulated by Touma et al. [W22], the lack of insight into the model architecture and hyperparameters within AutoML tools restricts users from understanding the classification mechanisms or customizing performance parameters. Furthermore, as highlighted by Schwen et al. [W45], AutoML Vision lacks logs or reports detailing initialization, network architecture search, or hyperparameter optimization, adding to the opaqueness surrounding model development processes.

***Lack flexibility***. This code is supported by seven references: five form the white literature [W15, W23, W52, W45, W40] and two from the grey literature [G54, G56]. AutoML's predefined algorithms and workflows may lack the flexibility needed for certain specialized tasks or unconventional problem domains, limiting its applicability in diverse scenarios. As highlighted by Xin et al. [W15], customizability ranks notably low among participants' assessments of AutoML tools. Interestingly, the dissatisfaction with customizability stemmed from both insufficient options for customization and an overwhelming degree of flexibility, contributing to the overall lower rating in this aspect. Moreover, as elucidated by Schwen et al. [W45], AutoML Vision's limitations are evident in its restrictive approach, allowing users to select only one among three presets and specifying a runtime budget based on the dataset size. The limited customization options prevent users from tailoring the tool to meet specific requirements of diverse datasets or specialized tasks.

***Have high costs for long-term usage***. Five papers support this code: three academic publications [W23, W45, W53] and two grey literature articles [G63, G88]. The adoption of AutoML can incur high costs, both in terms of licensing fees and computational resources required, potentially discouraging smaller organizations with limited budgets. As elucidated by Elshawi et al. [W53], the user-set time budget emerges as a critical parameter in AutoML systems, significantly impacting the system's ability to explore various options within the search space and thereby the likelihood of building optimal models. However, a larger time budget amplifies waiting periods and escalates computational resource consumption, directly translating into higher costs —especially when using cloud-based resources. For instance, Unadkat et al. [W23] note that while code-free ML systems are user-friendly and demand minimal technical expertise, cost concerns hinder widespread adoption among neurosurgeon-scientists, the target population in their study. Continuous cloud-based model usage incurs hourly charges, regardless of active training or testing.

***Have inadequate documentation***. A single academic paper supports this code [W60]. Insufficient software engineering practices in AutoML challenge code maintainability, scalability, and collaborative integration, hindering smooth incorporation of AutoML into established development workflows. In particular, in their study, Alamin and Uddin [W60] found that widespread API misuse often occurs due to inadequate documentation or insufficient expertise of software engineers across all phases of the machine learning life cycle. For instance, errors in API calls for the deployment of Google AutoML's natural language model reveal documentation deficiencies.

***Do not provide a universal solution***. Five papers support this code, all from the white literature [W25, W14, W83, W44, W51]. AutoML tools lack universal applicability across all domains, datasets, and ML scenarios. Truong et al. [W83] note that no single tool consistently outperforms others on multiple tasks, highlighting AutoML's varying effectiveness across different applications. In line with this observation, Mustafa and Rahimi Azghadi [W44] show that AutoML results can vary even when using the same dataset.

***Have compliance regulation issues***. This code is supported by three articles: two from the white literature [W45, W75] and one from the grey literature [G69]. Schwen et al. [W45] highlight potential compatibility issues between cloud-based tools like AutoML Vision and local data protection regulations. On the other hand, Polzer and Thalmann [W75] stress the need for better documentation of insights obtained during the exploration of AutoML use cases. In general, ensuring AutoML applications' regulatory compliance remains challenging, requiring more comprehensive measures to ensure alignment with evolving standards.

## 6. Discussion

### *6.1. RQ1 – On the Advantages of AutoML tools*

Our analysis of the primary advantages of AutoML tools, as addressed in RQ1 ("What are the main benefits of state-of-the-practice AutoML tools?"), reveals several impactful benefits associated with their use. AutoML tools showcase proficiency in streamlining multiple facets of the machine learning workflow, demonstrating their ability to optimize model performance through various means.

Our study presents a comprehensive list of 18 primary advantages associated with the utilization of AutoML tools. This distinguishes it from the aforementioned systematic literature review (SLR) on AutoML tools by Barbudo et al. [34], which explores different research questions, not explicitly focusing on reported benefits.

First, AutoML tools excel at simplifying feature engineering, aiding in the identification and integration of pertinent features critical to enhancing model performance. In addition, these tools expedite data selection, enabling efficient utilization of diverse datasets, which is particularly advantageous when handling extensive and varied data sources.

AutoML significantly contributes to data preprocessing by automating tasks such as handling missing values, scaling features, and encoding categorical variables. This automation not only saves time but also helps to have standardized, error-free data input for models.

Additionally, AutoML simplifies the overall model creation process by automating the building, validation, and deployment stages. It optimizes model configurations by automating hyperparameter tuning, consequently enhancing model accuracy and efficiency.

Beyond technical benefits, AutoML tools are pivotal in democratizing machine learning by offering user-friendly interfaces and automating complex processes, enabling individuals without extensive machine learning expertise to interact with ML models and perform data analysis. Furthermore, these tools incorporate features to mitigate bias in machine learning models, contributing to more reliable predictions.

The cumulative advantages highlighted underscore the substantial impact of AutoML in optimizing the machine learning workflow. These tools serve as catalysts for efficiency, accelerating processes, improving model performance, increasing productivity, aiding scalability, and facilitating rapid prototyping, thus broadening the approachability and effectiveness of artificial intelligence methods.

Practitioners can use the evidence to support more informed decisions on adopting AutoML tools. For instance, they could prioritize tools that offer customizable feature engineering and robust data preprocessing pipelines. Furthermore, tools with advanced hyperparameter tuning capabilities and support for diverse data types can be more effective in improving model prediction performance. Additionally, selecting AutoML tools that emphasize bias mitigation and transparency can be critical in ensuring ethical and reliable use in sensitive applications.

### 6.2. RQ2 – On the Limitations of AutoML tools

AutoML tools play a crucial role in democratizing machine learning, but their widespread adoption faces several challenges. In response to RQ2 ("What are the main limitations of state-of-the-practice AutoML tools?"), our analysis of multiple articles uncovered significant obstacles that shed light on the complexities and limitations inherent in AutoML. We provide a catalog of 25 limitations of AutoML tools qualitatively organized into themes (Figure 9).

One substantial challenge lies in the common lack of transparency and of effective bias mitigation within AutoML's decision-making processes, which impacts the understanding and trust in the generated models. Issues such as vendor lock-in and limited interoperability may pose additional important hurdles. AutoML's reliance on high-quality input data and struggles with diverse datasets and unstructured data compound these limitations.

Furthermore, Established AutoML algorithms are often constrained in their adaptability to specific tasks, potentially leading to reduced effectiveness in complex scenarios. Performance deficiencies, including suboptimal model outputs and potential overfitting, may undermine the reliability and applicability of AutoML-generated models.

It is crucial to note that AutoML does not replace human expertise but enhances efficiency, necessitating skilled users to leverage its capabilities alongside their own expertise and insights. Another related risk is the potential erosion of technical skills due to over-reliance on automation, further emphasizing these shortcomings.

Considering this evidence, practitioners could favor tools that best mitigate the limitations that have the most impact on their specific usage context. For example, tools that incorporate explainability techniques can help overcome transparency challenges by providing insights into model decisions. In environments where vendor lock-in and limited interoperability are concerns, opting for standard-compliant tools can enhance flexibility and facilitate smoother integration, particularly in production environments.

### 6.3. Future directions

While our goal was to synthesize evidence, producing (and assessing) practitioner tool selection guidelines considering the gathered evidence on the potential advantages and limitations for specific scenarios could be part of future work.

Furthermore, considering the evolution of the AutoML tools, the MLR revealed limitations that we hereafter map into potential areas for future enhancement and research.

*Scalability and efficient handling of large-scale data.* Addressing the challenges associated with large datasets is crucial. Future advancements should focus on developing innovations that enhance AutoML's scalability and efficiency in processing large-scale data without compromising performance. This includes strategies to optimize computational resources, reduce processing times, and streamline iterative model training on large datasets.

*Enhanced support for complex scenarios and diverse domains.* Enhancements are needed to improve AutoML's adaptability to complex scenarios and unsupervised learning tasks. Tailoring AutoML frameworks to effectively handle the scale and diversity of data in specialized domains, such as biomedical research or other complex fields, is important. Further research should aim to bridge the gap between predefined frameworks and the unique complexities present in diverse domains.

*Expanded capabilities for unstructured data.* Advancements in AutoML frameworks should prioritize comprehensive support for unstructured data types, including but not limited to images and text. While our review particularly highlighted the need for improved capabilities in image-based tasks such as segmentation, research efforts should focus on augmenting general AutoML's abilities to accommodate and process a wider range of input data types. This expansion would enhance AutoML's applicability across various domains that rely on complex, non-tabular data sources. Special attention should be given to improving these capabilities in open-source tools, as our review indicated a particular need in this area.

*Optimized resource allocation and infrastructure management.* Future developments should aim to strike a balance between time, computational resources, and costs in AutoML processes. Innovations in resource management strategies, cost-effective use of computational power, and mitigation of out-of-memory issues are essential to streamline AutoML operations.

*Improvements in data integration, preprocessing, and CI/CD integration.* Innovations in automating data integration from heterogeneous sources, robust preprocessing methods, and seamless integration with CI/CD pipelines are crucial. Addressing these challenges requires advanced capabilities to handle diverse data sources and frameworks.

*Comprehensive model deployment support and end-to-end automation.* Enhancements in AutoML should focus on comprehensive support for deploying models across diverse production environments. Moreover, efforts are needed to achieve more significant steps in end-to-end automation, covering all stages of the machine learning pipeline without requiring extensive manual intervention.

*Enhanced model diagnosis, monitoring, and updating.* Future developments should prioritize robust mechanisms to diagnose model behavior, real-time monitoring, and automated updating. Ensuring that models adapt to evolving data patterns is crucial for sustained performance in real-world applications.

*Performance optimization and bias mitigation.* Research efforts should aim to address issues related to suboptimal model performance, overfitting, and effective bias mitigation. The achievement of better model generalization and fairness in decision-making processes remains critical.

*Balancing technical expertise and automation.* Striking a balance between leveraging automation and nurturing technical expertise in data science is important. Future directions should focus on ensuring that AutoML tools complement, rather than replace, the role of skilled data scientists, fostering continuous innovation and understanding in the field.

*Enhanced transparency, flexibility, and cost considerations.* Improving transparency in AutoML decision-making processes, improving flexibility for customization, and addressing cost implications are crucial to broader accessibility and usability across diverse organizations and problem domains.

*Compliance and regulation alignment.* Future AutoML developments should ensure alignment with evolving regulatory standards, facilitating compatibility with various data protection regulations and documentation requirements.

In summary, addressing these specific challenges and prioritizing these future directions could significantly improve the efficiency, applicability, and ethical implementation of AutoML in various real-world scenarios.

## 7. Threats to validity

We aimed to cover a broad spectrum of AutoML content from academic and non-peer-reviewed literature. Therefore, we employed an efficient search strategy for the white literature [36] and described our search strategies for white and grey literature in detail, making the details of each filtering step available in our open science repository [38]. Despite employing comprehensive search strategies and examining multiple sources, there remains the risk of overlooking relevant information. Furthermore, the subjective nature of the exclusion criteria may have influenced the completeness of our data collection. To mitigate this threat, we peer-reviewed the application of the search strategy and the exclusion criteria.

An additional limitation arises from our decision to include in our results only the sources that mention AutoML solutions from Google, Microsoft, IBM, and Amazon (see EC5 in Table 1). However, our decision to focus on the four leading solutions from Gartner's Magic Quadrant was only considered as an initial selection criterion of the relevant results. In fact, our review includes an analysis of the benefits and challenges of any other tools mentioned in the selected sources, such as open source solutions. Although the goal of this paper was not to provide an exhaustive review of AutoML, we acknowledge this as a limitation that may have potentially excluded some valuable insights from lesser-known or emerging AutoML solutions.

Another potential limitation of our study stems from our selection criteria for grey literature. We focused exclusively on substantial written materials such as blog posts, articles, and white papers, while excluding short-form content (e.g., social media posts, advertisement blurbs) and non-written formats (e.g., podcasts, videos). This decision was made to ensure depth of analysis, maintain consistency with academic literature, and facilitate systematic review. However, we acknowledge that this approach may have excluded some valuable insights, particularly those shared through multimedia channels or in more concise formats.

A further limitation of our study is the exclusion of post-processing steps from our analysis of the typical ML workflow. Our simplified ML workflow (see Figure 1), based on Amershi et al.'s 9-stage workflow [9] and the CRISP-DM process model [10], does not explicitly include post-processing as a distinct stage. Nonetheless, we recognize that post-processing activities, such as report generation and result visualization, can play a key role in some AutoML applications. Future research could explore the role and importance of post-processing in AutoML, potentially uncovering additional benefits and limitations concerning related activities.

Another threat concerns publication bias. Although negative research is important because it demonstrates what does not work, scientific literature has a publication bias toward positive results [98]. Negative research is less cited, published, and is generally considered less scientifically interesting [98]. This threat is inherent in literature reviews and cannot be mitigated. However, we still found an expressive number of limitations (25).

Finally, regarding the data analysis, while conducting the thematic synthesis on the advantages and limitations of AutoML tools, we recognize some constraints that could impact the reliability of our findings. In our thematic synthesis, we assigned codes to the benefits and limitations, with the goal of transforming them into overarching themes. This process allowed us to condense a large amount of material into more concise elements. However, this process introduced subjectivity and potential biases. To reduce this threat, we peer-reviewed all the codes and carefully discussed them, involving the entire team of authors. Furthermore, to provide complete transparency and enable auditing of our qualitative coding, we made the coded data available in our open science repository.

## 8. Conclusion

This study investigates the benefits and limitations of AutoML solutions identified in the white and grey literature, with a particular emphasis on tools being considered for adoption in industry. The study employed comprehensive and transparent search strategies and qualitative thematic analysis procedures, revealing 18 reported benefits and 25 limitations. We visually organized the benefits and limitations into a catalog grouped by themes using mind maps. In general, AutoML tools streamline machine learning workflows, simplify tasks, empower users, improve model performance, increase efficiency, scalability, and accelerate prototyping. However, there are still ongoing difficulties, including the extent of workflow coverage, the limited ability to replace human expertise, problems with transparency and handling diverse data, and potential performance drawbacks.

While AutoML simplifies the process of creating machine learning pipelines, the effectiveness and breadth coverage of these tools can vary. One key point is that these tools enhance human expertise rather than replace it, requiring skilled users to use their capabilities. Gaining a comprehensive understanding of the AutoML landscape is essential for optimizing machine learning progress and choosing the tools appropriately.

Hence, the findings can be used by practitioners to consider trade-offs between such benefits and limitations to conduct effective evaluations of AutoML solution options. Furthermore, they can be used by researchers to steer future research addressing the current limitations.

## References


[1] M. I. Jordan, T. M. Mitchell, Machine learning: Trends, perspectives, and prospects, Science 349 (2015) 255–260.

[2] A. Luckow, K. Kennedy, M. Ziolkowski, E. Djerekarov, M. Cook, E. Duffy, M. Schleiss, B. Vorster, E. Weill, A. K. et al, Artificial intelligence and deep learning applications for automotive manufacturing, 2018 IEEE Int'l Conf. on Big Data (Big Data) (2018) 3144–3152.

[3] W. Markow, S. Braganza, B. Taska, S. M. Miller, D. Hughes, The quant crunch: How the demand for data science skills is disrupting the job market, https://www.ibm.com/downloads/cas/3RL3VXGA, 2017.

[4] B. Chakravorti, A. Bhalla, R. S. Chaturvedi, C. Filipovic, 50 global hubs for top ai talent", https://hbr.org/2021/12/50-global-hubs-for-top-ai-talent, 2021.

[5] D. Zhang, S. Mishra, E. Brynjolfsson, J. Etchemendy, D. Ganguli, B. Grosz, T. Lyons, J. C. N. J. Manyika, M. S. et al., The ai index 2021 annual report, 2021. doi:10.48550/arXiv.2103.06312.

[6] linkedin, Linkedin 2018 emerging jobs report, 2018. URL: https://economicgraph.linkedin.com/research/linkedin-2018-emerging-jobs-report.

[7] F. Hutter, L. Kotthoff, J. Vanschoren, Automated machine learning: methods, systems, challenges, Springer Link (2019).

[8] V. Baker, S. Sicular, E. Brethenoux, A. Batchu, M. Fang, Magic quadrant for cloud ai developer services, gartner.com/document/4014812 G00746705 (2022).

[9] S. Amershi, A. Begel, C. Bird, R. DeLine, H. Gall, E. Kamar, N. Nagappan, B. Nushi, T. Zimmermann, Software engineering for machine learning: A case study, 2019 IEEE/ACM 41st International Conference on Software Engineering: Software Engineering in Practice (ICSE-SEIP) (2019).

[10] C. Schröer, F. Kruse, J. M. Gómez, A systematic literature review on applying crisp-dm process model, Procedia Computer Science 181 (2021) 526–534.

[11] N. Nahar, H. Zhang, G. Lewis, S. Zhou, C. Kästner, A meta-summary of challenges in building products with ml components – collecting experiences from 4758+ practitioners, in: 2023 IEEE/ACM 2nd International Conference on AI Engineering – Software Engineering for AI (CAIN), 2023, pp. 171–183. doi:10.1109/CAIN58948.2023.00034.

[12] K. Chauhan, S. Jani, D. Thakkar, R. Dave, J. Bhatia, S. Tanwar, M. S. Obaidat, Automated machine learning: The new wave of machine learning, 2nd International Conference on Innovative Mechanisms for Industry Applications (ICIMIA) (2020) 205–212. doi:10.1109/ICIMIA48430.2020.9074859.

[13] P. Das, V. Perrone, N. Ivkin, T. Bansal, Z. Karnin, H. Shen, I. Shcherbatyi, Y. Elor, W. Wu, A. Zolic, T. Lienart, A. Tang, A. Ahmed, J. B. Faddoul, R. Jenatton, F. Winkelmolen, P. Gautier, L. Dirac, A. Perunicic, L. Rouesnel, Amazon sagemaker autopilot: A white box automl solution at scale, 2020. URL: http://arxiv.org/abs/2012.08483.

[14] J. Krauß, B. M. Pacheco, H. M. Zang, R. H. Schmitt, Automated machine learning for predictive quality in production, Procedia CIRP 93 (2020). doi:10.1016/j.procir.2020.04.039.

[15] D. Xin, E. Y. Wu, D. J.-L. Lee, N. Salehi, A. Parameswaran, Whither automl? understanding the role of automation in machine learning workflows, 2021. URL: http://arxiv.org/abs/2101.04834.

[16] K. W. Wan, C. H. Wong, H. F. Ip, D. Fan, P. L. Yuen, H. Y. Fong, M. Ying, Evaluation of the performance of traditional machine learning algorithms, convolutional neural network and automl vision in ultrasound breast lesions classification: A comparative study, Quantitative

Imaging in Medicine and Surgery 11 (2021). doi:10.21037/qims-20-922.

[17] L. Yang, A. Shami, Iot data analytics in dynamic environments: From an automated machine learning perspective, Engineering Applications of Artificial Intelligence 116 (2022). doi:10.1016/j.engappai.2022.105366.

[18] L. Faes, S. K. Wagner, D. J. Fu, X. Liu, E. Korot, J. R. Ledsam, T. Back, R. Chopra, N. Pontikos, C. Kern, G. Moraes, M. K. Schmid, D. Sim, K. Balaskas, L. M. Bachmann, A. K. Denniston, P. A. Keane, Automated deep learning design for medical image classification by healthcare professionals with no coding experience: A feasibility study, The Lancet Digital Health 1 (2019). doi:10.1016/S2589-7500(19)30108-6.

[19] D. Wang, J. D. Weisz, M. Muller, P. Ram, W. Geyer, C. Dugan, Y. Tausczik, H. Samulowitz, A. Gray, Human-ai collaboration in data science: Exploring data scientists' perceptions of automated ai, Proceedings of the ACM on Human-Computer Interaction 3 (2019) 1–24. doi:10.1145/3359313.

[20] A. Crisan, B. Fiore-Gartland, Fits and starts: Enterprise use of automl and the role of humans in the loop, 2021. URL: http://arxiv.org/abs/2101.04296.

[21] S. K. K. Santu, M. M. Hassan, M. J. Smith, L. Xu, C. Zhai, K. Veeramachaneni, Automl to date and beyond: Challenges and opportunities, 2022. URL: https://doi.org/10.1145/3470918. doi:10.1145/3470918.

[22] S. Touma, F. Antaki, R. Duval, Development of a code-free machine learning model for the classification of cataract surgery phases, Scientific Reports 12 (2022) 2398. doi:10.1038/s41598-022-06127-5.

[23] V. Unadkat, D. J. Pangal, G. Kugener, A. Roshannai, J. Chan, Y. Zhu, N. Markarian, G. Zada, D. A. Donoho, Code-free machine learning for object detection in surgical video: A benchmarking, feasibility, and cost study, Neurosurgical Focus 52 (2022) E11. doi:10.3171/2022.1.FOCUS21652.

[24] Y. Zeng, J. Zhang, A machine learning model for detecting invasive ductal carcinoma with google cloud automl vision, Computers in Biology and Medicine 122 (2020). doi:10.1016/j.compbiomed.2020.103861.

[25] R. Elshawi, S. Sakr, Automated machine learning: Techniques and frameworks, Big Data Management and Analytics 390 (2020).

[26] H. J. Escalante, Automated machine learning - A brief review at the end of the early years, in: N. Pillay, R. Qu (Eds.), Automated Design of Machine Learning and Search Algorithms, Natural Computing Series, Springer, 2021, pp. 11–28. URL: https://doi.org/10.1007/978-3-030-72069-8_2. doi:10.1007/978-3-030-72069-8_2.

[27] A. J. Thirunavukarasu, K. Elangovan, L. Gutierrez, R. Hassan, Y. Li, T. F. Tan, H. Cheng, Z. L. Teo, G. Lim, D. S. Wei, Clinical performance of automated machine learning: a systematic review, 2022. doi:10.1101/2023.10.26.23297599.

[28] J. A. C. Branco, B. T. Marques, L. A. Cruz, R. P. Magalhães, J. A. F. de Macêdo, Automatic machine learning applied to electrical biosignals: A selective review, 2023 IEEE 19th International Conference on e-Science (e-Science) (2023). doi:10.1109/e-Science58273.2023.10254793.

[29] D. Baymurzina, E. Golikov, M. Burtsev, A review of neural architecture search, Neurocomputing 474 (2022) 82–93.

[30] A. M. D. Valle, R. G. Mantovani, R. Cerri, A systematic literature review on automl for multi-target learning tasks, Artificial Intelligence Review 56 (2023). doi:10.1007/s10462-023-10569-2.

[31] R. Wen, S. Li, Spatial decision support systems with automated machine learning: A review, 2022. doi:10.3390/ijgi12010012.

[32] T. Nagarajah, G. Poravi, A review on automated machine learning (automl) systems, 2019 IEEE 5th International Conference for Convergence in Technology (I2CT) (2019) 1–6. doi:10.1109/I2CT45611.2019.9033810.

[33] E. T. Khalid, A. J. Y. Aldarwish, A. A.Yassin, Challenges in automl and declarative studies using systematic literature review, Applied Data Science and Analysis (2023) 118–125.

[34] R. Barbudo, S. Ventura, J. Romero, Eight years of automl: categorisation, review and trends, Knowl Inf Syst 65 (2023) 5097–5149. doi:10.1007/s10115-023-01935-1.

[35] V. Garousi, M. Felderer, M. V. Mäntylä, Guidelines for including grey literature and conducting multivocal literature reviews in software engineering, Information and Software Technology 106 (2019) 101–121. doi:10.48550/arXiv.1707.02553.

[36] C. Wohlin, M. Kalinowski, K. Felizardo, E. Mendes, Successful combination of database search and snowballing for identification of primary studies in systematic literature studies, Information and Software Technology 147 (2022).

[37] E. Mourão, J. F. Pimentel, L. Murta, M. Kalinowski, E. Mendes, C. Wohlin, On the performance of hybrid search strategies for systematic literature reviews in software engineering, Information and software technology 123 (2020) 106294.

[38] K. Azevedo, L. Quaranta, F. Calefato, M. Kalinowski, Artifacts: A multivocal literature review on the benefits and limitations of automated machine learning tools, https://doi.org/10.5281/zenodo.10515779, 2024.

[39] D. S. Cruzes, T. Dybå, Recommended steps for thematic synthesis in software engineering, ESEM '11: Proceedings of the 2011 International Symposium on Empirical Software Engineering and Measurement (2011). doi:10.1109/ESEM.2011.36.

[40] A. A. Borkowski, C. P. Wilson, S. A. Borkowski, L. B. Thomas, L. A. Deland, S. J. Grewe, S. M. Mastorides, Google auto ml versus apple create ml for histopathologic cancer diagnosis: Which algorithms are better?, 2019. URL: https://arxiv.org/abs/1903.08057.

[41] A. Sakagianni, G. Feretzakis, D. Kalles, C. Koufopoulou, V. Kaldis, Setting up an easy-to-use machine learning pipeline for medical decision support: A case study for covid-19 diagnosis based on deep learning

with ct scans, Studies in Health Technology and Informatics 272 (2020) 13–16. doi:10.3233/SHTI200481.

[42] J. Waring, C. Lindvall, R. Umeton, Automated machine learning: Review of the state-of-the-art and opportunities for healthcare, Artificial Intelligence in Medicine 104 (2020) 101822. doi:10.1016/j.artmed.2020.101822.

[43] T. Ghosh, S. Tanwar, S. Chumber, K. Vani, Classification of chest radiographs using general purpose cloud-based automated machine learning: Pilot study, Egyptian Journal of Radiology and Nuclear Medicine 52 (2021) 120. doi:10.1186/s43055-021-00499-w.

[44] A. Mustafa, M. Rahimi Azghadi, Automated machine learning for healthcare and clinical notes analysis, Computers 10 (2021) 24. doi:10.3390/computers10020024.

[45] L. O. Schwen, D. Schacherer, C. Geißler, A. Homeyer, Evaluating generic automl tools for computational pathology, Informatics in Medicine Unlocked 29 (2022). doi:10.1016/j.imu.2022.100853.

[46] R. Takezoe, X. Liu, S. Mao, M. T. Chen, Z. Feng, S. Zhang, X. Wang, Deep active learning for computer vision: Past and future, APSIPA Transactions on Signal and Information Processing 12 (2023). doi:10.1561/116.00000057.

[47] J. Bender, M. Trat, J. Ovtcharova, Benchmarking automl-supported lead time prediction, Procedia Computer Science 200 (2022). doi:10.1016/j.procs.2022.01.246.

[48] K. J. Kramer, N. Behn, M. Schmidt, The potential of automl for demand forecasting, 3rd Conference on Production Systems and Logistics (2022) 682–692. doi:10.15488/12162.

[49] D. Venkata Vara Prasad, P. Senthil Kumar, L. Y. Venkataramana, G. Prasannamedha, S. Harshana, S. Jahnavi Srividya, K. Harrinei, S. Indraganti, Automating water quality analysis using ml and auto ml techniques, Environmental Research 202 (2021) 111720. doi:10.1016/j.envres.2021.111720.

[50] M. Bahri, F. Salutari, A. Putina, M. Sozio, Automl: State of the art with a focus on anomaly detection, challenges, and research directions, International Journal of Data Science and Analytics 14 (2022). doi:10.1007/s41060-022-00309-0.

[51] K. T. Y. Mahima, T. N. D. S. G., K. De, Evaluation of sentiment analysis based on automl and traditional approaches, International Journal of Advanced Computer Science and Applications 12 (2021). doi:10.14569/IJACSA.2021.0120277.

[52] N. Noshiri, M. Khorramfar, T. Halabi, Machine learning-as-a-service performance evaluation on multi-class datasets, 2021 IEEE International Conference on Smart Internet of Things (SmartIoT) (2021) 332–336. doi:10.1109/SmartIoT52359.2021.00060.

[53] R. Elshawi, M. Maher, S. Sakr, Automated machine learning: State-of-the-art and open challenges, 2019. URL: http://arxiv.org/abs/1906.02287.

[54] D. Xin, R. Meertens, Automl: the promise vs. reality according to practitioners, https://www.infoq.com/articles/auto-ml-promise-reality/, 2022.

[55] thenewstack, 4 Major Benefits of Using AutoML over Hand Coding - The New Stack, 2022. URL: https://thenewstack.io/4-major-benefits-of-using-automl-over-hand-coding/.

[56] sapsales, SAP Sales Forecast using Automated Machine Learning {Azure AutoML No-code/Low-code} on Microsoft Power BI | SAP Blogs, 2021. URL: https://blogs.sap.com/2021/01/12/sap-sales-forecast-using-automated-machine-learning-azure-automl-no-code-on-microsoft-power-bi/.

[57] biztechmagazine, What is AutoML and Why Should Your Business Consider It - BizTech Magazine, 2020. URL: https://biztechmagazine.com/article/2020/08/automated-machine-learning-whats-its-role-future-work-perfcon.

[58] A. Mullen, S. Alaybeyi, V. Baker, A. Chandrasekaran, A. Linden, M. Revang, S. Sicular, Predicts 2020: Artificial intelligence — the road to pro-

duction, Gartner (2020). URL: https://www.gartner.com/document/3975770.

[59] V. Baker, B. Elliot, S. Sicular, A. Mullen, E. Brethenoux, Critical capabilities for cloud ai developer services, https://www.gartner.com/document/3999739, 2021.

[60] M. A. A. Alamin, G. Uddin, Challenges and barriers of using low code software for machine learning, 2022. URL: http://arxiv.org/abs/2211.04661.

[61] M. Revang, V. Baker, A. Linden, R. Sallam, M. Bhat, A. Mullen, Predicts 2019: The democratization of ai, https://www.gartner.com/document/3894386, 2019.

[62] P. Bansal, Automated Machine Learning | Azure | Pros And Cons | K21Academy, 2022. URL: https://k21academy.com/microsoft-azure/dp-100/automated-machine-learning-azure-pros-and-cons/.

[63] futurice, What is AutoML and how can it be applied to practice?, 2021. URL: https://futurice.com/blog/what-is-automl-and-how-can-it-be-applied-to-practice.

[64] V. Baker, S. Sicular, E. Brethenoux, A. Batchu, M. Fang, Magic quadrant for cloud ai developer services, Gartner (2022). URL: https://www.gartner.com/document/4014812.

[65] D. Bonderud, Automated machine learning: What's its role in the future of work?, https://biztechmagazine.com/article/2020/08/automated-machine-learning-whats-its-role-future-work-perfcon, 2020.

[66] V. Baker, A. Mullen, B. Elliot, Market guide for hosted ai services, Gartner (2018). URL: https://www.gartner.com/document/3892295.

[67] A. Batchu, T. Klinect, A. Thomas, Hype cycle for open-source software, 2020, 2020. URL: https://www.gartner.com/document/3987447.

[68] S. Vashisth, A. Linden, J. Hare, P. den Hamer, Hype cycle for data science and machine learning, 2020, https://www.gartner.com/document/3988118, 2020.

[69] forbytes, What Is AutoML? Meaning, Use, Examples of Tools - Forbytes, 2022. URL: https://forbytes.com/blog/what-is-automl/.

[70] D. Wadhwa, Pros and Cons of Automated Machine Learning, 2019. URL: https://www.topbug.net/blog/2019/04/17/pros-and-cons-of-automated-machine-learning/.

[71] N. Kreiling, AutoML - A Comparison of cloud offerings, 2020. URL: https://www.scieneers.de/automl-a-comparison-of-cloud-offerings/.

[72] analyticsinsight, How Automated Machine Learning Addresses the Challenges of Traditional ML Models? | Analytics Insight, 2020. URL: https://www.analyticsinsight.net/how-automated-machine-learning-addresses-the-challenges-of-traditional-ml-models/.

[73] altexsoft, AutoML: Capabilities and Limitations of Automated Machine Le, 2021. URL: https://www.altexsoft.com/blog/automl/.

[74] R. S. M. Lakshmi Patibandla, V. S. Srinivas, S. N. Mohanty, C. Ranjan Pattanaik, Automatic machine learning: An exploratory review, 2021 9th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO) (2021) 1–9. doi:10.1109/ICRITO51393.2021.9596483.

[75] A. K. Polzer, S. Thalmann, The impact of automl on the ai development process, Proceedings of the 2022 Pre-ICIS SIGDSA Symposium 13 (2022).

[76] S. Luo, V. Kindratenko, Hands-on with ibm visual insights, Computing in Science & Engineering 22 (2020) 108–112. doi:10.1109/MCSE.2020.3009765.

[77] datatechvibe, Why Does Your Business Need AutoML?, 2021. URL: https://datatechvibe.com/ai/why-does-your-business-need-automl/.

[78] automl2022, Automl - the future of machine learning, https://insidebigdata.com/2022/12/28/automl-the-future-of-machine-learning/, 2022.

[79] S. Brändle, M. Hanussek, M. Blohm, M. Kintz, Evaluation of representation models for text classification with automl tools, 2021. URL: http://arxiv.org/abs/2106.12798.

[80] N. Kamdar, Text Classification using AutoML Tables|Google Cloud Platform, 2022. URL: https://medium.com/google-cloud/text-classification-using-automl-tables-google-cloud-platform-43816c4afb93.

[81] informationage, How AutoML is helping to bridge the artificial intelligence skills gap, 2018. URL: https://www.information-age.com/automated-machine-learning-ai-skills-gap-11766/.

[82] A. Rosebrock, Auto-Keras and AutoML: A Getting Started Guide, 2019. URL: https://pyimagesearch.com/2019/01/07/auto-keras-and-automl-a-getting-started-guide/.

[83] A. Truong, A. Walters, J. Goodsitt, K. Hines, C. B. Bruss, R. Farivar, Towards automated machine learning: Evaluation and comparison of automl approaches and tools, 2019 IEEE 31st International Conference on Tools with Artificial Intelligence (ICTAI) (2019) 1471–1479. doi:10.1109/ICTAI.2019.00209.

[84] S. B. Lee, X. Gui, E. Hamilton, Application of automl in the automated coding of educational discourse data, Interdisciplinarity in the Learning Sciences: Proceedings of the 14th International Conference of the Learning Sciences, ICLS (2020).

[85] R. Sallam, D. Feinberg, M. Beyer, W. R. Schulte, A. Linden, J. Unsworth, S. Sicular, N. Heudecker, E. Zaidi, A. Ronthal, E. Brethenoux, P. den Hamer, A. Woodward, Top 10 data and analytics technology trends that will change your business, 2019. URL: https://www.gartner.com/document/3906812.

[86] ibm2023, IBM Wants To Make Artificial Intelligence Fair And Transparent With AI OpenScale, 2018. URL: https://www.forbes.com/sites/janakirammsv/2018/10/21/ibm-wants-to-make-artificial-intelligence-fair-and-transparent-with-ai-openscale/?sh=2b11b16a1251.

[87] M. Hayashi, K. Tamai, Y. Owashi, K. Miura, Automated machine learning for identification of pest aphid species (hemiptera: Aphididae), Applied Entomology and Zoology 54 (2019) 487–490. doi:10.1007/s13355-019-00642-0.

[88] getindata, Automated Machine Learning (AutoML) with BigQuery ML. Start Machine Learning easily and validate if ML is worth investing in or not. - GetInData, 2022. URL: https://getindata.com/blog/automated-machine-learning-automl-bigquery-ml-start-machine-learning-easily-validate-ml-worth-investing/.

[89] G. Simsek, Introduction to Automated Machine Learning (AutoML), 2019. URL: https://softwareengineeringdaily.com/2019/05/15/introduction-to-automated-machine-learning-automl/.

[90] T. Siriborvornratanakul, Human behavior in image-based road health inspection systems despite the emerging automl, Journal of Big Data 9 (2022) 96. doi:10.1186/s40537-022-00646-8.

[91] Y.-W. Chen, Q. Song, X. Hu, Techniques for automated machine learning, 2019. URL: http://arxiv.org/abs/1907.08908.

[92] G. Symeonidis, E. Nerantzis, A. Kazakis, G. A. Papakostas, Mlops - definitions, tools and challenges, 2022. URL: http://arxiv.org/abs/2201.00162.

[93] F. Majidi, M. Openja, F. Khomh, H. Li, An empirical study on the usage of automated machine learning tools, 2022. URL: http://arxiv.org/abs/2208.13116.

[94] D. Wang, Q. V. Liao, Y. Zhang, U. Khurana, H. Samulowitz, S. Park, M. Muller, L. Amini, How much automation does a data scientist want?, 2021. URL: http://arxiv.org/abs/2101.03970.

[95] D. J.-L. Lee, S. Macke, D. Xin, A. Lee, S. Huang, A. Parameswaran, A human-in-the-loop perspective on automl: Milestones and the road ahead, IEEE Data Engineering Bulletin (2019).

[96] J. Drozdal, J. Weisz, D. Wang, G. Dass, B. Yao, C. Zhao, M. Muller, L. Ju, H. Su, Trust in automl: Exploring information needs for establishing trust in automated machine learning systems, Proceedings of the 25th International Conference on Intelligent User Interfaces (2020) 297–307. doi:10.1145/3377325.3377501.

[97] N. Schlüter, Can automated machine learning outperform handcrafted models?, 2019. URL: https://towardsdatascience.com/can-automated-machine-learning-outperform-handcrafted-models-feff83e0dc97.

[98] M. N, F. E, E. M, L. JS, N. KA., Negativity towards negative results: a discussion of the disconnect between scientific worth and scientific culture, Dis Model Mech 7 (2014) 171–3. doi:10.1242/dmm.015123.