Winter 12-10-2022

The impact of AutoML on the AI development process

Anna Katharina Polzer

Stefan Thalmann

The impact of AutoML on the AI development process

Completed Research

Anna Katharina Polzer
University of Graz
anna.polzer@uni-graz.at

Stefan Thalmann
University of Graz
stefan.thalmann@uni-graz.at

Abstract

Artificial intelligence (AI)-based decision support systems (DSS) have been promoted for their high potential for many application domains. However, the adoption of such AI-based DSS is still low in practice. The AI-expert-centric development process and domain experts' challenges in exploring suitable AI use cases and communicating their requirements are major barriers for adoption. Automated Machine Learning (AutoML) is one approach to democratise AI, which could empower domain experts to explore AI use cases and better communicate with AI experts. However, as of now, little focus has been put on how AutoML is being used in the context of AI development processes. In this literature review, we investigate AutoML use cases and structure the results according to the CRISP-DM model. Based on the uncovered benefits and challenges of AutoML, we propose a research agenda with five major future research streams.

Keywords

Artificial Intelligence, Machine Learning, AutoML, Decision Support Systems

Introduction

Advancements in the field of artificial intelligence (AI) and especially in machine learning (ML) have had an important impact on the development of decision support systems (DSS) (Thalmann, 2018). However, while AI systems[1] are adopted in several application domains of DSS, there are still many use cases of AI in the context of DSS which lag behind. Challenges in AI adoption range from regulatory barriers (Königstorfer & Thalmann, 2020, 2022)technological and organisational to social and data-related barriers (Dwivedi et al., 2021). A pivotal point in developing AI systems is incorporating domain expertise into the design process of AI-based DSS systems. So far, the design process is centred around AI experts not only responsible for the development process but also for defining the scope of AI systems (Ozkaya, 2020). While domain experts such as e.g. engineers or operators have an in-depth understanding of the application domain of an AI-based DSS, they lack knowledge regarding AI capabilities in the complex development process. Thus, the development process usually requires an iterative back and forth between domain and AI experts to align the different knowledge requirements (Zöller & Huber, 2021). As a result, domain experts face challenges in identifying suitable use cases for AI-based DSS. Further communication barriers between domain experts and developers have often been cited as a critical challenge in the development of AI systems (Bauer et al., 2020; Enholm et al., 2021; Kirschbaum et al., 2022; Westenberger et al., 2022).

One stream of technological advancements that could empower domain experts and facilitate communication in AI projects is automated Machine Learning (AutoML). AutoML is a subfield of AI research that subsumes the methods aiming to automate at least to some extent all stages of the design and development of AI systems (Hutter et al., 2019). As such, AutoML aims to reduce tedious and repetitive tasks in the development of AI systems and offer easier access to powerful ML algorithms to people without in-depth AI/ML knowledge and, in doing so, democratise AI to a broader audience (Hutter et al., 2019).

[1] In this paper, the term AI systems will be used to refer to systems based on ML techniques rather than knowledge-driven ones (e.g. expert systems).

Therefore, AutoML can be seen as an entry point for domain experts to familiarise themselves with ML models and AI-based DSS systems. Similar to productivity tools such as spreadsheet applications, domain experts could experiment and create first prototypical solutions for potential use cases in their daily work routines using AutoML tools without involving AI experts. The model created could facilitate communication in the professional development of AI systems (Karmaker et al., 2022). Although several commercial and open source AutoML applications and frameworks exist, e.g., auto-sklearn, H2O or Google Cloud AutoML, little research has focused on the utilisation perspective of AutoML in the context of the AI development process. As such this literature review aims to answer the research question:

What are the current application cases of AutoML, and what are associated benefits and challenges?

To answer this research, question a literature review was conducted. The next section focusses on the challenges in AI development and the role of AutoML in the AI development process. The third section describes the literature review's methodology, followed by the discussion of the results in the fourth section. Based on the findings from the literature review, the research agenda for AutoML is presented in the fifth section, followed by the paper's conclusion.

Background

Challenges in AI development

Although a myriad of AI use cases has been noted in recent years across different industries and application domains (Bertolini et al., 2021; Loureiro et al., 2021), many companies still struggle to apply AI systems more broadly and outside of proof of concept and pilot stages (Dzhusupova et al., 2022). While traditional software engineering and in a broader sense IT project management has a long history of formalised development processes (Sommerville, 2011), these traditional processes cannot be transferred without adjustments to the development of AI systems (Ozkaya, 2020). Specific characteristics of AI, such as the opaqueness (Guidotti et al., 2019), the probabilistic nature of learning-based AI systems (Laato et al., 2022), the necessity for high-quality data for the AI system development (Enholm et al., 2021; Z. Wan et al., 2020) are major issues for the technical development process. However, regulatory and ethical challenges also arise as such systems' testing and validation is considered very challenging (Baier et al., 2019; Ishikawa & Yoshioka, 2019; Z. Wan et al., 2020).

Still, not only technical and regulatory issues pose a challenge to the development of AI systems. In their investigation of AI readiness, Jöhnk et al. (2021) identified five categories of organisational factors that influence the success of AI adoption. While technical aspects such as data quality and accessibility necessarily influence the success of AI development, the authors also highlight that strategic alignment, culture, resources and knowledge are important (Jöhnk et al., 2021). Similarly, Baier et al. (2019) showcase that besides technical challenges lacking know-how and communication with the users of the AI systems are major challenges in the deployment and operation of AI systems in practice. Additionally, Bauer et al. (2020) found that a primary challenge in the development and implementation of AI systems in small and medium-sized enterprises (SMEs) is the lack of basic ML capabilities and know-how for the use case definition as well as the implementation. As such, the lack of AI/ML knowledge is a frequently noted challenge for investigating and implementing potential AI use cases (Bauer et al., 2020; Enholm et al., 2021; Kirschbaum et al., 2022; Westenberger et al., 2022). Nevertheless, the success of AI use cases highly depends on the skills and knowledge of organisations' employees (Enholm et al., 2021). It is critical not just for the identification and exploration of potential use cases, but also in the development of AI systems, as AI development processes usually require an iterative back and forth between domain experts who want to satisfy a specific business need and AI experts having the data science and AI development skills but are lacking knowledge of the application domain (Zöller & Huber, 2021).

However, this issue has received little attention so far, as existing literature on the identification and adoption of AI use cases often assumes that companies already have some understanding and knowledge of AI (Kirschbaum et al., 2022). Some researchers suggest that AutoML might provide an easy way for domain experts to explore possible use cases (Crisan & Fiore-Gartland, 2021). This could be especially beneficial as domain experts provide expertise and context that is highly relevant in the context of DSS, yet AutoML cannot provide it (Xin et al., 2021). However, due to the easy and unrestricted access to powerful ML algorithms, also concerns regarding the possibility of automating bad decisions by non-experts have been raised (Crisan & Fiore-Gartland, 2021; D. Wang et al., 2019). As such also important issues concerning fairness, accountability and transparency of AI systems might get overlooked.

Additionally, research on appropriate AI development processes and AI project management in general is limited (Ishikawa & Yoshioka, 2019; Laato et al., 2022). Laato et al. (2022) investigated how to integrate ML into established software development approaches and concluded that existing software engineering approaches like the waterfall model or Scrum are considerably modified for ML use cases in practice. However, the authors also highlighted that more cyclical approaches might be promising to achieve a shared understanding and commitment between domain experts and developers (Laato et al., 2022). As such, AutoML might be utilised to overcome the current challenges of AI development processes.

Crisp-DM

As mentioned before no common approach for AI development projects or data science projects in general exists (Kolyshkina & Simoff, 2021). However, for the organisation of the results of our analysis of AutoML use cases, we draw upon the structure of the Cross-Industry Standard Process for Data Mining (CRISP-DM) model. Introduced in the late 1990s, CRISP-DM is an industry-independent and application agnostic process model for data mining (Chapman et al., 2000; Wirth & Hipp, 2000). Due to some drawbacks and limitations of the CRISP-DM model such as its inherent rigidity (Walcott & Ali, 2021) or the lack of updates that would help to better encompass newer technologies (Kolyshkina & Simoff, 2021; Schafer et al., 2018), several enhancement of the CRISP-DM model and novel methodologies such as the improved Analytics Solutions Unified Method for Data Mining (ASUM-DM) (Angée et al., 2018) or Lean Design Thinking Methodology for Machine Learning and Modern Data Projects (LDTM) (Ahmed et al., 2018) have been proposed. However, the CRISP-DM model is still one of the most used methodologies and is considered an industry standard for data science projects (Ahmed et al., 2018; Kolyshkina & Simoff, 2021).

The CRISP-DM model breaks down the data science project into six phases: *Business Understanding, Data Understanding, Data Preparation, Modelling, Evaluation,* and *Deployment* (Chapman et al., 2000). In the *Business Understanding* phase, project objectives and requirements are gathered. Key steps of this phase include determining the business objectives, situation assessment, and project planning. In the *Data understanding* phase, data is collected and explored including steps like the description and exploration of data as well as verifying the data quality. In the third phase, *Data Preparation,* the data is prepared for model training which includes steps such as selecting data by defining inclusion and exclusion criteria, cleaning and imputation of data, construction, merging and integration, as well as formatting data. The *Modelling* phase consists of four steps: selection of the modelling technique, development of the test design and the model and assessment of the model based on evaluation criteria. The *Evaluation* phase consists of planning the deployment, monitoring, maintenance, reporting, and documentation.

AutoML

AutoML is a subfield of AI research that subsumes the methods aiming to automate the design and development of ML systems (Hutter et al., 2019). Even as AutoML aims to make AI accessible for non-ML experts (domain experts), applying AutoML still requires human involvement in several vital steps, as well as basic knowledge of AI (Xin et al., 2021). AutoML research aims to democratise AI to non-experts by enabling domain experts to automatically build ML applications without the need to rely on data scientists or having in-depth statistical or AI/ML knowledge (He et al., 2021; Hutter et al., 2019). As such Karmaker et al. (2022) defined two major user groups of AutoML tools, namely domain experts who have in-depth knowledge of the domain in which ML is applied but limited knowledge of ML itself and data scientists who have in-depth knowledge of ML but limited knowledge of the domain where it is applied. Reaching the overall goal of AutoML to democratise AI would require research to focus on the utilisation and the design requirement perspective of AutoML tools, especially including the perspective of domain expert needs. However, as of now, the majority of AutoML research focuses on the development and improvements of algorithms to automate the ML pipeline, like hyperparameter optimisation (HPO), neural architecture search (NAS) or Combined Algorithm Selection and Hyper-parameter tuning (CASH). Reviews from the technical perspective of AutoML were published by He et al. (2021), Elshawi et al. (2019) and Zöller & Huber (2021).

Currently, several mature AutoML tools exist that rival and sometimes even outperform human experts (Hutter et al., 2019). Xin et al. (2021) categorise existing AutoML tools into three groups, *Open-Source Software, Cloud Provider Solutions* and *AutoML Platforms* (Xin et al., 2021)*:*

- *Open-Source Software* is characterised by a high level of flexibility which can easily be integrated into custom code. However, these tools usually run on their own computational resources and usually lack support in post-processing steps such as evaluation, deployment and monitoring of models. AutoML tools of this category are, for instance, libraries such as TPOT, autoKeras or H2O.
- Prominent AutoML tools of the category *Cloud Provider Solutions* are, for instance, Google Cloud AutoML, Microsoft Azure or Amazon SageMaker Autopilot. These AutoML tools usually do not require computation resources as they are hosted by the cloud provider. Further, as they tend to encompass the entire ML pipeline and usually provide a no-code user interface (UI) such AutoML tools require less to no programming expertise. However, the system's internals are opaque to the user and thus often less configurable and transparent.
- Compared to *Cloud Provider Solutions,* AutoML tools of the *AutoML Platform* group position themselves as turnkey AutoML solutions that provide more technical support and customizability in each stage of the ML pipeline, especially regarding model interpretability and deployment options. AutoML of this category is, for instance, DataRobot andH20Driverless.

However, an in-depth investigation of how this plethora of AutoML tools can be utilised in practice is scarce. Xanthopoulos et al. (2020) provided an evaluation of AutoML tools from a user’s perspective. Further, existing research studied the perception of AI experts on automation in AI development, highlighting the benefits and limitations of utilising AutoML (D. Wang et al., 2019; Xin et al., 2021). Similarly, Crisan & Fiore-Gartland (2021) investigated the utilisation of AutoML in enterprises and identified three usage scenarios for AutoML; (1) the automation of routine tasks, (2) the fast exploration of potential data science solutions and (3) the ability to build AI systems by domain experts. Additionally, Karmaker et al. (2022) provide a review of which tasks of the ML pipeline have been automated as of now and how a potential end-to-end ML pipeline for AI development would look like. Although, integrating AutoML as a facilitator into the AI development process has been stated in several studies (Bauer et al., 2020; Karmaker et al., 2022; D. Wang et al., 2019; Zöller & Huber, 2021), to the best of our knowledge no literature review exists that investigates the utilisation of AutoML by domain experts in the context of the AI development process.

Methodology

To answer the research question, a structured literature review, according to Webster and Watson was conducted, which consists of the three steps *Identifying relevant literature*, *Structuring the review* and *Theoretical development* (Webster & Watson, 2002).

As the goal of this literature review was to showcase the utilisation of AutoML across domains, the literature search was performed in the Web of Science and Scopus databases. These two databases have been selected because both have an extensive corpus of multidisciplinary research, while strict selection criteria must be satisfied for articles to be listed in the databases. Due to the novelty of the topic, the query focused on papers published no later than January 2017 and listed in both databases by March 2022.

To exhaust the exploration of any existing literature on AutoML use cases relevant keywords and search strings were defined. The main keywords included variations on the term AutoML such as “automated machine learning” or “automated ML”. As keywords that would indicate a specific application of AutoML terms and variations of the word “tool”, “software”, “application”, “library” or “package” were used. This resulted in 350 papers from the Scopus database and 126 papers in the Web of Science database. In the next step, an abstract scan was carried out to identify whether the extracted papers utilise an AutoML framework or application of AutoML was a central aspect of the paper. As such, all papers that focused on the development or optimisation of AutoML from a technical perspective were excluded. One paper was excluded because it was the only one that did not utilise AutoML for a supervised ML problem and thus did not fit into the comparison of the identified AutoML use cases. Further also short papers were excluded. After a forward-backwards scan, a total of 46 papers were considered for in-depth content analysis.

In a second step, the review was structured in a concept-centric way. As a basis for this, the Crisp-DM process model has been utilised as a template for analysing the identified AutoML use cases. This allowed to analyse the identified literature based on an ideal-typical procedure of conducting a ML use case and to showcase the differences in the approaches and documentation of the AutoML use cases. As suggested by Webster and Watson, this literature review also focused on the identification of knowledge gaps which became evident in the analysis of the literature (Webster & Watson, 2002). A discussion and elaboration on possible future research directions is presented in the *Research agenda* section.

Discussion of results

In this section, the results are discussed according to the phases of the CRISP-DM model. Regarding the fifth phase, Schröer et al. (2021) found in their investigation of reported CRISP-DM applications that the deployment phase is missing in most of the reviewed studies. This also became apparent in this literature review. Which is why reports on the last phase of the CRISP-DM model – *Deployment* – have been omitted from the discussion of results. All articles analysed in this literature review are: Agrapetidou et al., 2021; Antaki et al., 2020; Antaki et al., 2021; Arrogante-Funes et al., 2021; Ashraf et al., 2021; Bruzón et al., 2021; Cui et al., 2020; Czub et al., 2021; Dafflon et al., 2020; Faes et al., 2019; Fayez & Kurnaz, 2021; Feretzakis et al., 2021; Gomathi et al., 2022; Han et al., 2021; Howard et al., 2020; Ikemura et al., 2021; Ito et al., 2021; Jiménez et al., 2020; Karaglani et al., 2020; Koh et al., 2021; Korot et al., 2021; Leduc & Assaf, 2020; Li et al., 2021; Mahima et al., 2021; Mena et al., 2022; Orlenko et al., 2020; Ou et al., 2021; Panagopoulou et al., 2021; Peng et al., 2022; Sawaki et al., 2019; Schulze-Brüninghoff et al., 2021; Sills et al., 2021; Smith et al., 2022; Stojadinovic et al., 2021; Su et al., 2020; Sun et al., 2021; Tran et al., 2020; Tsamardinos et al., 2020; Tsiakmaki et al., 2020; K. W. Wan et al., 2021; S. Wang et al., 2020; P. Yang et al., 2021; H.-S. Yang et al., 2021; Q. Zhang et al., 2020; C. Zhang & Ye, 2021; Zhu et al., 2021.

Business Understanding

Although AutoML aims at the democratisation of data science, the literature review showed an especially strong utilisation of existing AutoML tools in medicine and life sciences, where we found 24 out of the 46 analysed papers. The other prominent fields were environmental sciences (e.g., Arrogante-Funes et al., 2021; Jiménez et al., 2020; Li et al., 2021; Sun et al., 2021), engineering (e.g., Q. Zhang et al., 2020; C. Zhang & Ye, 2021), chemistry (Tsamardinos et al., 2020; P. Yang et al., 2021) or energy (Ashraf et al., 2021; Mena et al., 2022). Still, most other use cases stem from a multidisciplinary background (e.g., Sawaki et al., 2019; Schulze-Brüninghoff et al., 2021).

The main objective of more than 80% of the reviewed use cases was to investigate possible new application cases of AutoML and ML models in their domain. As such, several studies examined the suitability of AutoML for predictions on domain-specific data and problem settings. For instance, Cui et al. (2020) investigated the suitability of AutoML for predicting scores of patients with nasopharyngeal carcinoma, while Koh et al. (2021) investigated the suitability of utilising AutoML for plant classification via remote sensing data. Additionally, several studies stated that they aimed to get faster results through the utilisation of AutoML, which could be beneficial, especially in the medical domain, for the timely treatment of patients (e.g., Feretzakis et al., 2021; Sills et al., 2021) or could provide a diagnosis based on data that could be gathered with less invasive procedures on the patients (Peng et al., 2022). Besides investigating new use cases for AutoML and ML in general, several studies focused on the investigation of the feasibility of AutoML tools to be utilised by domain experts without AI expertise (e.g., Antaki et al., 2020; Faes et al., 2019; Ou et al., 2021) or the comparison between traditionally engineered AI systems and results of the utilisation of AutoML (e.g., Mahima et al., 2021; S. Wang et al., 2020). Current AutoML applications are mainly developed for supervised ML problems. Approximately three-quarters of the reviewed papers focused on classification problems in their use cases. Only in nine articles AutoML was utilised for regression problems. Tsiakmaki et al. (2020) as well as Koh et al. (2021) used AutoML for classification and regression models. Additionally, S. Wang et al. (2020) utilised AutoML for classification and object detection in the context of the detection of breast lesions.

In summary, the analysis of the *Business Understanding* phase showcased that current AutoML tools are used in many different backgrounds and for diverse tasks. Deriving from the mostly missing stated intent, most of the published papers did not expect their use cases to be implemented for a broader user base. As such, the reported findings could mostly be considered an exploration of how to utilise AutoML for domain-specific problems or early proof of concept reports. This is also reflected in the reported use of AutoML by AI experts, where AutoML is often used in the early AI development stages for exploration (Crisan & Fiore-

Gartland, 2021; Xin et al., 2021). Furthermore, the analysis of the data mining goals found rather vague reported goals, which could be an indicator that the analysed use cases had a rather investigative purpose and that the authors might not necessarily have in-depth AI expertise to formulate the desired data mining goals clearly. Thus, also the lack of in-depth documentation in this step of the CRISP-DM is understandable, as most publications were more concerned with providing a first proof of concept.

Data understanding

The investigation of the *Data Understanding* phase showed that most papers utilised already existing data sets. In the medical domain, the initial data collection was usually extracted as retrospective studies (e.g., Feretzakis et al., 2021; Ito et al., 2021). About a third of the examined papers used publicly available data (e.g., Han et al., 2021; Zhu et al., 2021) or data sets that were made available specifically for the purpose of the studies (Howard et al., 2020; C. Zhang & Ye, 2021). In only eight investigated AutoML use cases, the data was collected specifically for the purpose of the papers (e.g., Leduc & Assaf, 2020; Sawaki et al., 2019).

Around half of the AutoML use cases were applied to structured data (e.g., Ashraf et al., 2021; Tsiakmaki et al., 2020), while the second most used data format was image data in approx. a third of all use cases (e.g., Antaki et al., 2021; Li et al., 2021). However, there are only a few use cases on other data types, such as text (Howard et al., 2020), multitemporal data (Arrogante-Funes et al., 2021) or video (Smith et al., 2022).

While most of the analysed papers describe the surface properties of the initial data sets similar to the suggestions of CRISP-DM, the level of detail varies widely. As such, some only reference the origin of the data set used, but do not provide any description of what subset of the data was used (Han et al., 2021). In contrast to the description of the data sets' surface information, just around two-thirds of all examined papers additionally provide some exploration of the data sets used, which in most cases consist of basic descriptive statistics of the most important feature (e.g., Stojadinovic et al., 2021; Zhu et al., 2021). Only a couple of papers provide more in-depth results of data exploration on the initial data set (e.g., Bruzón et al., 2021; Czub et al., 2021). Just around a quarter of the analysed use cases address data quality issues. Karaglani et al. (2020) mentioned the quality and completeness in open data sets as a possible quality issue. In cases where image data was used the main point of concern was the inadequate quality of images (e.g., Schulze-Brüninghoff et al., 2021; Su et al., 2020) as well as the potential of observer bias during the selection and annotation of the data (Ito et al., 2021).

In summary, the analysis showed that the data used for the AutoML use cases were rarely specifically collected or prepared. Even in cases where the utilised data was described in more detail, the focus of the descriptions and explorations were more in line with the scientific practices in the respective domains and less regarding the specific ML problem to be addressed. More detailed examination and reflection on the data is found in notably few use cases and mostly in more sensitive application domains like health sciences. Thus, from the analysis of the *Data Understanding* phase, no clear conclusion could be drawn whether AutoML would need more or fewer data instances than traditional ML approaches. Especially as contradictory statements have been found in the analysed use cases (Leduc & Assaf, 2020; Sawaki et al., 2019; Tran et al., 2020). As we found only a few instances in which data quality issues were addressed, it seems that data quality verification was not at the forefront of the analysis. Although not necessarily true, a reason for this might be that the knowledge about such quality issues is not present in domain expert users of AutoML and data quality verification is usually not a part of the provided AutoML tools.

Data preparation

The description of the data preparation in the reviewed publications was kept rather short or in case of seven articles completely absent (e.g., Agrapetidou et al., 2021; Han et al., 2021). More than half of all reviewed papers at least provide some basic description of how the data was selected. Especially in cases where image data was used, a description of how the data was selected and how areas of interest were defined was provided (e.g., Jiménez et al., 2020; Sawaki et al., 2019). For other data formats, authors usually listed their reasoning for including and excluding data (e.g., Panagopoulou et al., 2021; C. Zhang & Ye, 2021) or described how the features for modelling were selected (e.g., Cui et al., 2020; Howard et al., 2020). Thirteen reviewed papers provided information on the data cleaning procedure. Most described in various levels of detail how they dealt with missing values or duplicates (e.g., Czub et al., 2021; Sills et al., 2021). Most reviewed articles do not provide information on the construction step of the *Data Preparation* phase. The data preparation step was described in detail in few cases where additional software was used to transform or construct data (e.g., Peng et al., 2022; Schulze-Brüninghoff et al., 2021). Regarding documentation on data formatting, around a third of papers reported on conducted scaling or normalisation of their data (e.g., Feretzakis et al., 2021; Panagopoulou et al., 2021).

Most investigated papers described a form of automated feature engineering by the utilised AutoML tools, this may have caused a scarce description of the conducted data preparation. The overall tenor was that little effort had been spent on data preparation and feature extraction steps. A reason for this might be that, as mentioned in the *Business Understanding* section, the primary focus of the majority of the analysed use cases was of an exploratory nature. Thus, little effort was spent on optimising and documenting this step of the AI development process. However, especially in the context of data preparation, AutoML may not provide the best solutions as it has been noted to be incomplete or insufficient at times (Xin et al., 2021). Another possible explanation for the absent documentation on the *Data Preparation* phase might be lacking feedback from the AutoML tools themselves. However, several AutoML tools offer additional information on demand (Drozdal et al., 2020).

Modelling

The modelling part of the AI development process is a key aspect of all AutoML applications and is thus mostly automated. AutoML tools provide default settings which can be customized if necessary.

The selection of the modelling technique (selection of ML model) is left to the AutoML tool in almost all use cases. However, there are noticeable variations in the developed test designs. Almost all use cases split the data for training and testing purposes according to common ML rule of thumb, in the range of 80-70% of data for the training data to 20-30% for the test dataset (e.g., Koh et al., 2021; Sills et al., 2021), or utilise cross-validation to get more stable indicators on the performance of the developed prediction models (e.g., Cui et al., 2020; Czub et al., 2021). Cross-validation is also already automatically implemented in some AutoML tools like TPOT (e.g., Su et al., 2020; Tsamardinos et al., 2020) or JadBio (e.g., Agrapetidou et al., 2021; Karaglani et al., 2020). Only a few studies report test designs that deviate from such traditional approaches, but usually do not provide rationale on why a particular training/test split and evaluation procedure was chosen (e.g., Ito et al., 2021; C. Zhang & Ye, 2021). Additionally to the internal evaluation of the model performance, only a handful of studies examined the generalizability of their models on additional/external data (e.g., Gomathi et al., 2022; Panagopoulou et al., 2021). In terms of evaluation of the AutoML results, more than half of the reviewed articles compared the results to similar manually generated ML models (e.g., Koh et al., 2021; Mahima et al., 2021) or existing non-ML based prediction or classification models (e.g., Cui et al., 2020; Panagopoulou et al., 2021). However, some papers did not provide any information on the test design at all (Gomathi et al., 2022; Leduc & Assaf, 2020).

In general, the models were assessed on standard metrics such as accuracy and sensitivity for classification problems or variations on error metrics like RMSE and MAE for regression models. While the metrics on which the models were evaluated were rather standardised, the depth of the documentation on the assessment varied noticeably. In most cases, the models were assessed on a combination of metrics which were stated but not necessarily further interpreted (e.g., Faes et al., 2019; Korot et al., 2021). By contrast, around a quarter of the reviewed used cases provided a detailed description of the generated models and how they selected the final model (e.g., Su et al., 2020; Zhu et al., 2021). In this context, the issue of limited customizability of GUI-based AutoML tools was criticised as it limited the comparability between models (e.g., Antaki et al., 2021; H.-S. Yang et al., 2021). Besides the reporting of the models' performances on standard metrics, almost a third of the publications further investigated the generated AutoML models with some post-hoc explainable AI (XAI) approach (Adadi & Berrada, 2018) such as the analysis of feature importance or partial dependence plots (e.g., Antaki et al., 2021; Bruzón et al., 2021). It is noticeable that especially users of the H2O AutoML tool used XAI approaches in their articles, as six out of eight total H2O users provided an assessment of the developed AutoML models beyond standard performance metrics (e.g., Karaglani et al., 2020; Sills et al., 2021). However, it is unclear whether the XAI analysis was included because H2O as an AutoML tool provides easy access or whether H2O as an AutoML tool was chosen because it includes easily available XAI approaches.

Summing up, few articles discussed the modelling results and their implications in detail. With regard to the affiliations of the authors a reason for this might be that most use cases seem to be written and published by domain experts (e.g., Korot et al., 2021; Schulze-Brüninghoff et al., 2021; C. Zhang & Ye, 2021) that do not necessarily have in-depth knowledge on AI development and evaluation and thus focus on the performance metrics provided by the utilised AutoML tools. If this is the case, the question arises of what kind of feedback or guidance domain experts need to evaluate developed ML models and the relation these might have to their problem sets. Although recognised as an issue (Crisan & Fiore-Gartland, 2021; Drozdal et al., 2020), little research has yet been done on guidance needs for domain experts using AutoML. Another issue raised in several publications is the lack of transparency in AutoML tools, confirming the assessment of Xanthopoulos et al. (2020) that the results of AutoML should be explained, visualised and interpreted. Some AutoML applications do provide basic explainability components that can help users analyse the results of the AutoML process. Still, the knowledge of the existence and how to interpret such explanations for the developed AutoML would require the domain experts to at least have a familiarity with AI development and its evaluation. However, there is a wide range of individual preferences or different domains of use that needs to be considered for the design of user guidance (Drozdal et al., 2020). As such, research on the transparency as well as the explanation needs, especially for domain experts, is lacking.

Evaluation

The Evaluation step of the CRISP-DM aims to assess to which degree the earlier defined (business) objectives have been reached. Our analysis of the reviewed papers showed overall relatively positive results regarding the predefined goals. Almost all reviewed articles concluded that their main objectives were, in some capacity, reached (e.g., Ou et al., 2021; Q. Zhang et al., 2020). However, most of the reviewed papers noted limitations or announced that they would work on further improvements of the models as issues and challenges became apparent. Most often, data-related issues were addressed, where the provision of more standardised or more in-depth data for future improvements was noted (e.g., Karaglani et al., 2020; Leduc & Assaf, 2020). Antaki et al. (2020) concluded that using AutoML by domain experts without further support is questionable, as in their study, the data preparation steps were conducted by a data scientist instead of domain experts. Only Ashraf et al. (2021) explicitly stated that they chose to proceed with a manual AI development approach as the manually developed AI's performance showcased better performance results than AutoML. As another challenge in the utilisation of AutoML, Koh et al. (2021) highlighted the significant computational costs associated with AutoML. This concern applies especially to OSS tools which usually require suitable computational resources from the users. While this can be compensated with cloud-based solutions, these usually come with the drawbacks of less transparency and less customizability (Xin et al., 2021). As such, also limitations in model transparency and explainability have been frequently mentioned (e.g., Dafflon et al., 2020; Korot et al., 2021). Furthermore, uncertainties regarding regulatory compliance of AutoML have been pointed out, especially in sensitive application domains such as the energy (Dafflon et al., 2020) or medicine (Ikemura et al., 2021).

In conclusion, AutoML as a tool for exploration or as an early proof of concept on how ML could be used for domain-specific issues resulted in predominantly positive results. AutoML is not used for developing AI-based DSS systems that would be deployed in work processes or professional contexts. Rather, AutoML was used as exploration tool for domain experts providing them valuable insights into the development of AI systems in general as well about ML-opportunities in their domains. As suggested by D. Wang et al. (2019), AutoML can thus be utilised for educational purposes, as domain experts seem to develop a better understanding of what and how different challenges may affect the outcome of the ML development project when they use AutoML for initial explorations of an ML task.

Research agenda

Building on the insights of the *Discussion of results* and the identified open issues for AI development through AutoML, we present five thematic blocks for future research.

Basic AI/ML training for domain experts

As seen in the analysis of the reviewed AutoML use cases, AutoML is often used by domain experts to conduct an initial exploration of whether AI-based DSS systems are even suitable approaches for their problem settings. AutoML provides an easy-to-use starting point for domain experts to get first-hand experience in developing AI systems. However, the analysis of the documented AutoML use cases also showcased that important issues for AI development, such as data quality or data preparation, are often overlooked, which leads to concerns regarding the fundamental knowledge of domain experts regarding AI/ML. Similar concerns have already been raised that AutoML might lead to automating bad decisions by non-experts (Crisan & Fiore-Gartland, 2021). D. Wang et al. (2019) even reported that data scientists worry that through AutoML the bar might be lowered too much for uneducated users to conduct good data science. However, as of now, little research has been conducted on what knowledge is needed by domain experts to utilise AutoML as also little attention has been given to the question for which purposes domain experts might even use AutoML. While several studies investigated the use of AutoML by AI experts (Crisan & Fiore-Gartland, 2021; D. Wang et al., 2019; Xin et al., 2021), not much insight is available on the usage of AutoML by domain experts. However, the required knowledge of domain experts to utilise AutoML as a tool for explorative tasks differs considerably from the knowledge needed to build AI applications for deployment in real-life settings. As such, basic training concepts should be developed taking different utilisation purposes of AutoML into account.

User guidance for domain experts using AutoML

The literature review further identified challenges of interpreting the results or identifying possible quality issues along the AutoML pipeline. It cannot be expected of domain experts to acquire sufficient knowledge by using initial training courses to understand all pitfalls along the ML development pipeline. This became evident during the analysis of the *Data Understanding* and *Data Preparation* phases of the reviewed literature as considerations and documentation on important issues such as data quality or the influence of imbalanced data classes were the exception rather than the rule. Q. Yang et al. (2018) already illuminated some technical pitfalls and possible guidance opportunities. Further, explainability turned out to be a critical factor for domain experts. Thus, it seems desirable to include XAI components enabling domain experts to understand the solution (Gashi et al., 2022) or to even provide causal discovery mechanisms to further investigate the AI model and its interdependencies (Vuković & Thalmann, 2022). However, further research is needed that focuses on identifying domain experts' guidance needs and investigating how they could be met. In this literature review, clear guidance needs were identified regarding data understanding and preparation, as well as the interpretation of results. As such, guidance concepts that provide some kind of data quality check or guide the users towards evaluating the suitability of the data to reach the ML goal would be necessary. However, the development of such guidance concepts is very complex as user information and guidance needs are very diverse, and no one-size-fits-all approach will be able to satisfy everyone's information and guidance requirements (Drozdal et al., 2020; Xanthopoulos et al., 2020). In this context, it is also important not to overload domain experts as this could hinder the utilisation of AutoML. As such Q. Yang et al. (2018) suggest to embed safeguards and corrective measures directly into the utilised tools, while Xin et al. (2021) propose an adaptive UI for AutoML tools that take into account the diverse skills and expertise of its users. However, a thorough investigation of the suitability of such guidance approaches is still missing. Thus, future research needs to focus on developing and evaluating guidance concepts for the utilisation of AutoML by domain experts.

Embedding AutoML into the AI development process

The literature review revealed that AutoML is often utilised as an exploration tool by domain experts. As such, it has also been noted that after the first proof-of-concept done with AutoML by domain experts, professional AI development teams are intended to take over the development of an AI system which should be deployed in a professional manner. One key question in this regard is how the insight gained by the domain experts during the exploration with AutoML can be transferred to the AI development team. This transfer could be beneficial for the entire AI development process, but particularly in the context of requirements engineering, which is often regarded as challenging in AI system development (Felderer & Ramler, 2021) and further exacerbated by the insufficient AI knowledge of domain experts (Ishikawa & Yoshioka, 2019; Westenberger et al., 2022). The potential of AutoML for educational purposes has already been recognised by D. Wang et al. (2019), who argue that AutoML could take the role of a teacher that educates people less versed in AI development. Thus, utilising AutoML for the exploration of AI use cases might support the common understanding of domain experts and AI developers regarding contextual definitions or data requirements and further help clarify and define the AI system requirements. However, no research has been found that would investigate approaches of integrating AutoML into the AI development process or that would connect the challenges and possibilities of integrating AutoML to the existing body of knowledge in IS development.

Verifying the regulatory compliance of AutoML applications

Although the reviewed literature showcased that AutoML is mostly used in explorative studies, the aim of AutoML is not just to automate repetitive and time-consuming tasks in ML development but also the democratisation of AI to non-experts by enabling domain experts to automatically build ML applications without the need to rely on data scientist (Zöller & Huber, 2021). As such, the vision is to use AutoML also as a productivity tool in practice. However, considering upcoming AI regulations around the world, such as the legal framework proposed by the European Commission (2021) or the AI Governance Framework developed by Singapore's Personal Data Protection Commission Singapore (2020), further questions arise on how AutoML fits in these frameworks and which requirements might be imposed on AI applications developed through AutoML. As such, Xin et al. (2021) noted that AI experts switch to manual AI development in sensible use cases. The main reason is that AI experts need to be able to reason and justify their design decision. Further, not only the development and the resulting AI system are subject to regulations, also the embedding of the system in business processes and decision-making needs to fulfil the requirement of human oversight (Kloker et al., 2022). Thus, the question remains which regulatory requirements will be imposed on AutoML-based AI systems and whether and how they might restrict the democratisation efforts of AutoML. Further, it is important to investigate how to train and guide users regarding regulatory requirements. As such, more effort must be put into assuring the reliability and safe utilisation of AutoML tools that corresponds to the level of proficiency of its user to achieve regulatory compliance.

Documenting the AutoML exploration

A fact that has become evident through the analysis of the reviewed literature is that there is no standardised approach to documenting the AI development process. This is less surprising as the analysed use cases mostly stemmed from various disciplines that focused on reporting their exploration of domain specific AutoML use cases rather than adhering to documentation standards for AI systems. However, adequate documentation of the carried-out exploration steps could be a relevant artefact when professional AI development teams are charged with developing full-fledged AI applications. Although documentation standards for AI systems are just being developed, several suggestions on how to document AI systems have been published (Liu et al., 2020; Mitchell et al., 2019). In this regard, Königstorfer & Thalmann (2021) identified five requirements for AI documentation: *descriptions of the application domain*, *training data set* and *design decisions*, and *understandable documentation for knowledgeable third parties* and the *balancing of the benefits and efforts of documenting artificial intelligence*. However, the opaqueness and lack of transparency and customizability of some AutoML tools might make such documentation difficult, especially regarding the description of design decisions. In this literature review, it became apparent that users of specific AutoML tools were constrained to the information provided (Antaki et al., 2021; H.-S. Yang et al., 2021). Furthermore, the question remains which additional aspects of AutoML will be required in the documentation of AutoML use cases, particularly also regarding regulatory conformity.

Conclusion

This paper presents the results of a literature review about the utilization of AutoML. The central insight is that in most cases, AutoML has been utilised as a tool for the exploration of various domain-specific problems. Thus, AutoML provides easy access for domain experts to develop AI-based DSS systems. However, AutoML is used mainly for exploratory purposes, while full-fledged AI systems are still expected to be developed by AI experts. As the vast majority of the reviewed use cases were satisfied with the utilisation of AutoML, further utilization and spread seems likely.

Based on our insights we propose a research agenda. Further research should focus on basic AI training as well as user guidance for domain experts using AutoML. Further, the embedding of AutoML into AI development processes and how AutoML can facilitate the knowledge transfer between domain and AI experts is a promising avenue for future research. Research in this direction might mitigate common challenges identified in the AI development process, such as the lack of AI knowledge in domain experts and the resulting communication issues. An important issue that also needs further attention is the documentation of insights gathered during the exploration of AutoML use cases. Finally, the verification of regulatory compliance of AutoML applications is still an open issue.

References

Adadi, A., & Berrada, M. (2018). Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI). *IEEE Access*, *6*, 52138–52160. https://doi.org/10.1109/ACCESS.2018.2870052

Agrapetidou, A., Charonyktakis, P., Gogas, P., Papadimitriou, T., & Tsamardinos, I. (2021). An AutoML application to forecasting bank failures. *Applied Economics Letters*, *28*(1), 5–9. https://doi.org/10.1080/13504851.2020.1725230

Ahmed, B., Dannhauser, T., & Philip, N. (2018). A Lean Design Thinking Methodology (LDTM) for Machine Learning and Modern Data Projects. In *2018 10th Computer Science and Electronic Engineering (CEEC)* (pp. 11–14). IEEE. https://doi.org/10.1109/CEEC.2018.8674234

Angée, S., Lozano-Argel, S. I., Montoya-Munera, E. N., Ospina-Arango, J.-D., & Tabares-Betancur, M. S. (2018). Towards an Improved ASUM-DM Process Methodology for Cross-Disciplinary Multi-organization Big Data & Analytics Projects. In L. Uden, B. Hadzima, & I.-H. Ting (Eds.), *Communications in Computer and Information Science. Knowledge Management in Organizations: 13th International Conference, KMO 2018, Žilina, Slovakia, August 6-10, 2018, Proceedings* (Vol. 877, pp. 613–624). Springer International Publishing. https://doi.org/10.1007/978-3-319-95204-8_51

Antaki, F., Coussa, R. G., Kahwati, G., Hammamji, K., Sebag, M., & Duval, R. (2021). Accuracy of automated machine learning in classifying retinal pathologies from ultra-widefield pseudocolour fundus images. *The British Journal of Ophthalmology*. Advance online publication. https://doi.org/10.1136/bjophthalmol-2021-319030

Antaki, F., Kahwati, G., Sebag, J., Coussa, R. G., Fanous, A., Duval, R., & Sebag, M. (2020). Predictive modeling of proliferative vitreoretinopathy using automated machine learning by ophthalmologists without coding experience. *Scientific Reports*, *10*(1), 19528. https://doi.org/10.1038/s41598-020-76665-3

Arrogante-Funes, P., Bruzón, A. G., Arrogante-Funes, F., Ramos-Bernal, R. N., & Vázquez-Jiménez, R. (2021). Integration of Vulnerability and Hazard Factors for Landslide Risk Assessment. *International Journal of Environmental Research and Public Health*, *18*(22). https://doi.org/10.3390/ijerph182211987

Ashraf, W. M., Uddin, G. M., Farooq, M., Riaz, F., Ahmad, H. A., Kamal, A. H., Anwar, S., El-Sherbeeny, A. M., Khan, M. H., Hafeez, N., Ali, A., Samee, A., Naeem, M. A., Jamil, A., Hassan, H. A., Muneeb, M., Chaudhary, I. A., Sosnowski, M., & Krzywanski, J. (2021). Construction of Operational Data-Driven Power Curve of a Generator by Industry 4.0 Data Analytics. *Energies*, *14*(5), 1227. https://doi.org/10.3390/en14051227

Baier, L., Jöhren, F., & Seebacher, S. (2019). Challenges in the Deployment and Operation of Machine Learning in Practice. In *Ecis 2019 proceedings* (pp. 8–14). Association for Information Systems.

Bauer, M., van Dinther, C., & Kiefer, D. (2020). Machine Learning in SME: An Empirical Study on Enablers and Success Factors. In *AMCIS 2020 Proceedings*. https://aisel.aisnet.org/amcis2020/adv_info_systems_research/adv_info_systems_research/3

Bertolini, M., Mezzogori, D., Neroni, M., & Zammori, F. (2021). Machine Learning for industrial applications: A comprehensive literature review. *Expert Systems with Applications*, *175*, 114820. https://doi.org/10.1016/j.eswa.2021.114820

Bruzón, A. G., Arrogante-Funes, P., Arrogante-Funes, F., Martín-González, F., Novillo, C. J., Fernández, R. R., Vázquez-Jiménez, R., Alarcón-Paredes, A., Alonso-Silverio, G. A., Cantu-Ramirez, C. A., & Ramos-Bernal, R. N. (2021). Landslide Susceptibility Assessment Using an AutoML Framework. *International Journal of Environmental Research and Public Health*, *18*(20). https://doi.org/10.3390/ijerph182010971

Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., & Wirth, R. (2000). CRISP-DM 1.0: Step-by-step data mining guide. *SPSS Inc*, *9*(13), 1–73.

Crisan, A., & Fiore-Gartland, B. (2021). Fits and Starts: Enterprise Use of AutoML and the Role of Humans in the Loop. In Y. Kitamura (Ed.), *ACM Digital Library, Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (pp. 1–15). Association for Computing Machinery. https://doi.org/10.1145/3411764.3445775

Cui, C., Wang, S [Shunxin], Zhou, J., Dong, A., Xie, F., Li, H., & Liu, L. (2020). Machine Learning Analysis of Image Data Based on Detailed MR Image Reports for Nasopharyngeal Carcinoma Prognosis. *BioMed Research International*, *2020*, 8068913. https://doi.org/10.1155/2020/8068913

Czub, N., Pacławski, A., Szlęk, J., & Mendyk, A. (2021). Curated Database and Preliminary AutoML QSAR Model for 5-HT1A Receptor. *Pharmaceutics*, *13*(10). https://doi.org/10.3390/pharmaceutics13101711

Dafflon, J., Pinaya, W. H. L., Turkheimer, F., Cole, J. H., Leech, R., Harris, M. A., Cox, S. R., Whalley, H. C., McIntosh, A. M., & Hellyer, P. J. (2020). An automated machine learning approach to predict brain age from cortical anatomical measures. *Human Brain Mapping*, *41*(13), 3555–3566. https://doi.org/10.1002/hbm.25028

Drozdal, J., Weisz, J., Wang, D., Dass, G., Yao, B., Zhao, C., Muller, M., Ju, L., & Su, H. (2020). Trust in AutoML. In F. Paternò, N. Oliver, C. Conati, L. D. Spano, & N. Tintarev (Eds.), *Proceedings of the 25th International Conference on Intelligent User Interfaces* (pp. 297–307). ACM. https://doi.org/10.1145/3377325.3377501

Dwivedi, Y. K., Hughes, L., Ismagilova, E., Aarts, G., Coombs, C., Crick, T., Duan, Y., Dwivedi, R., Edwards, J., Eirug, A., Galanos, V., Ilavarasan, P. V., Janssen, M., Jones, P., Kar, A. K., Kizgin, H., Kronemann, B., Lal, B., Lucini, B., . . . Williams, M. D. (2021). Artificial Intelligence (AI): Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy. *International Journal of Information Management*, *57*, 101994. https://doi.org/10.1016/j.ijinfomgt.2019.08.002

Dzhusupova, R., Bosch, J., & Olsson, H. H. (2022). The Goldilocks Framework: Towards Selecting the Optimal Approach to Conducting AI Projects. In *2022 IEEE/ACM 1st International Conference on AI Engineering–Software Engineering for AI (CAIN))* (pp. 124–135).

Elshawi, R., Maher, M., & Sakr, S. (2019). *Automated Machine Learning: State-of-The-Art and Open Challenges*.

Enholm, I. M., Papagiannidis, E., Mikalef, P., & Krogstie, J. (2021). Artificial Intelligence and Business Value: A Literature Review. *Information Systems Frontiers*. Advance online publication. https://doi.org/10.1007/s10796-021-10186-w

European Commission. (2021). *Proposal for a Regulation of the European Parliament and the Council: Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts*. https://ec.europa.eu/newsroom/dae/redirection/document/75788

Faes, L., Wagner, S. K., Fu, D. J., Liu, X., Korot, E., Ledsam, J. R., Back, T., Chopra, R., Pontikos, N., Kern, C., Moraes, G., Schmid, M. K., Sim, D., Balaskas, K., Bachmann, L. M., Denniston, A. K., & Keane, P. A. (2019). Automated deep learning design for medical image classification by health-care professionals with no coding experience: a feasibility study. *The Lancet Digital Health*, *1*(5), e232-e242. https://doi.org/10.1016/S2589-7500(19)30108-6

Fayez, M., & Kurnaz, S. (2021). Novel method for diagnosis diseases using advanced high-performance machine learning system. *Applied Nanoscience*. Advance online publication. https://doi.org/10.1007/s13204-021-01990-6

Felderer, M., & Ramler, R. (2021). *Quality Assurance for AI-based Systems: Overview and Challenges*.

Feretzakis, G., Sakagianni, A., Loupelis, E., Kalles, D., Skarmoutsou, N., Martsoukou, M., Christopoulos, C., Lada, M., Petropoulou, S., Velentza, A., Michelidou, S., Chatzikyriakou, R., & Dimitrellos, E. (2021). Machine Learning for Antibiotic Resistance Prediction: A Prototype Using Off-the-Shelf Techniques and Entry-Level Data to Guide Empiric Antimicrobial Therapy. *Healthcare Informatics Research*, *27*(3), 214–221. https://doi.org/10.4258/hir.2021.27.3.214

Gashi, M., Vuković, M., Jekic, N., Thalmann, S., Holzinger, A., Jean-Quartier, C., & Jeanquartier, F. (2022). State-of-the-Art Explainability Methods with Focus on Visual Analytics Showcased by Glioma Classification. *BioMedInformatics*, *2*(1), 139–158. https://doi.org/10.3390/biomedinformatics2010009

Gomathi, S., Kohli, R., Soni, M., Dhiman, G., & Nair, R. (2022). Pattern analysis: predicting COVID-19 pandemic in India using AutoML. *World Journal of Engineering*, *19*(1), 21–28. https://doi.org/10.1108/WJE-09-2020-0450

Guidotti, R., Monreale, A., Ruggieri, S., Turini, F., Giannotti, F., & Pedreschi, D. (2019). A Survey of Methods for Explaining Black Box Models. *ACM Computing Surveys*, *51*(5), 1–42. https://doi.org/10.1145/3236009

Han, T., Gois, F. N. B., Oliveira, R., Prates, L. R., & Porto, M. M. d. A. (2021). Modeling the progression of COVID-19 deaths using Kalman Filter and AutoML. *Soft Computing*, 1–16. https://doi.org/10.1007/s00500-020-05503-5

He, X [Xin], Zhao, K., & Chu, X. (2021). AutoML: A survey of the state-of-the-art. *Knowledge-Based Systems*, *212*, 106622. https://doi.org/10.1016/j.knosys.2020.106622

Howard, D., Maslej, M. M., Lee, J., Ritchie, J., Woollard, G., & French, L. (2020). Transfer Learning for Risk Classification of Social Media Posts: Model Evaluation Study. *Journal of Medical Internet Research*, *22*(5), e15371. https://doi.org/10.2196/15371

Hutter, F., Kotthoff, L., & Vanschoren, J. (2019). *Automated Machine Learning*. Springer International Publishing. https://doi.org/10.1007/978-3-030-05318-5

Ikemura, K., Bellin, E., Yagi, Y., Billett, H., Saada, M., Simone, K., Stahl, L., Szymanski, J., Goldstein, D. Y., & Reyes Gil, M. (2021). Using Automated Machine Learning to Predict the Mortality of Patients With COVID-19: Prediction Model Development Study. *Journal of Medical Internet Research*, *23*(2), e23458. https://doi.org/10.2196/23458

Ishikawa, F., & Yoshioka, N. (2019). How Do Engineers Perceive Difficulties in Engineering of Machine-Learning Systems? - Questionnaire Survey. In *2019 IEEE/ACM Joint 7th International 5 2019* (pp. 2–9). https://doi.org/10.1109/CESSER-IP.2019.00009

Ito, Y., Unagami, M., Yamabe, F., Mitsui, Y., Nakajima, K., Nagao, K., & Kobayashi, H. (2021). A method for utilizing automated machine learning for histopathological classification of testis based on Johnsen scores. *Scientific Reports*, *11*(1), 9962. https://doi.org/10.1038/s41598-021-89369-z

Jiménez, B., Maya, C., Velásquez, G., Barrios, J. A., Pérez, M., & Román, A. (2020). Helminth Egg Automatic Detector (HEAD): Improvements in development for digital identification and quantification of Helminth eggs and its application online. *MethodsX*, *7*, 101158. https://doi.org/10.1016/j.mex.2020.101158

Jöhnk, J., Weißert, M., & Wyrtki, K. (2021). Ready or Not, AI Comes— An Interview Study of Organizational AI Readiness Factors. *Business & Information Systems Engineering*, *63*(1), 5–20. https://doi.org/10.1007/s12599-020-00676-7

Karaglani, M., Gourlia, K., Tsamardinos, I., & Chatzaki, E. (2020). Accurate Blood-Based Diagnostic Biosignatures for Alzheimer's Disease via Automated Machine Learning. *Journal of Clinical Medicine*, *9*(9). https://doi.org/10.3390/jcm9093016

Karmaker, S. K., Hassan, M. M., Smith, M. J., Xu, L., Zhai, C., & Veeramachaneni, K. (2022). AutoML to Date and Beyond: Challenges and Opportunities. *ACM Computing Surveys*, *54*(8), 1–36. https://doi.org/10.1145/3470918

Kirschbaum, J., Posselt, T., & Roth, A. (2022). Use-Case-Based Innovation For Artificial Intelligence–An Ontological Approach. In *ECIS 2022 Research-in-Progress Papers*. https://aisel.aisnet.org/ecis2022_rip/64/

Kloker, A., Fleiß, J., Koeth, C., Kloiber, T., Ratheiser, P., & and Thalmann, S. (2022). Caution or Trust in AI? How to design XAI in sensitive Use Cases? In *AMCIS 2022 Proceedings*. https://aisel.aisnet.org/amcis2022/sig_dsa/sig_dsa/16/

Koh, J. C., Spangenberg, G., & Kant, S. (2021). Automated Machine Learning for High-Throughput Image-Based Plant Phenotyping. *Remote Sensing*, *13*(5), 858. https://doi.org/10.3390/rs13050858

Kolyshkina, I., & Simoff, S. (2021). Interpretability of Machine Learning Solutions in Public Healthcare: The CRISP-ML Approach. *Frontiers in Big Data*, *4*, 660206. https://doi.org/10.3389/fdata.2021.660206

Königstorfer, F., & Thalmann, S. (2020). Applications of Artificial Intelligence in commercial banks – A research agenda for behavioral finance. *Journal of Behavioral and Experimental Finance*, *27*, 100352. https://doi.org/10.1016/j.jbef.2020.100352

Königstorfer, F., & Thalmann, S. (2021). Software documentation is not enough! Requirements for the documentation of AI. *Digital Policy, Regulation and Governance*, *23*(5), 475–488. https://doi.org/10.1108/DPRG-03-2021-0047

Königstorfer, F., & Thalmann, S. (2022). AI Documentation: A path to accountability. *Journal of Responsible Technology*, *11*, 100043. https://doi.org/10.1016/j.jrt.2022.100043

Korot, E., Pontikos, N., Liu, X., Wagner, S. K., Faes, L., Huemer, J., Balaskas, K., Denniston, A. K., Khawaja, A., & Keane, P. A. (2021). Predicting sex from retinal fundus photographs using automated deep learning. *Scientific Reports*, *11*(1), 10286. https://doi.org/10.1038/s41598-021-89743-x

Laato, S., Mäntymäki, M., Minkkinen, M., Birkstedt, T., Islam, A., & Dennehy, D. (2022). Integrating machine learning with software development lifecycles: Insights from experts. In *ECIS 2022 Research Papers*. https://aisel.aisnet.org/ecis2022_rp/118/

Leduc, E., & Assaf, G. J. (2020). Road visualization for smart city: Solution review with road quality qualification. *Internet of Things*, *12*, 100305. https://doi.org/10.1016/j.iot.2020.100305

Li, K.-Y., Burnside, N. G., Lima, R. S. de, Peciña, M. V., Sepp, K [Karli], Cabral Pinheiro, V. H., Lima, B. R. C. A. de, Yang, M.-D., Vain, A., & Sepp, K [Kalev] (2021). An Automated Machine Learning Framework in Unmanned Aircraft Systems: New Insights into Agricultural Management Practices Recognition Approaches. *Remote Sensing*, *13*(16), 3190. https://doi.org/10.3390/rs13163190

Liu, X., Rivera, S. C., Moher, D., Calvert, M. J., & Denniston, A. K. (2020). Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: The CONSORT-AI Extension. *BMJ (Clinical Research Ed.)*, *370*, m3164. https://doi.org/10.1136/bmj.m3164

Loureiro, S. M. C., Guerreiro, J., & Tussyadiah, I. (2021). Artificial intelligence in business: State of the art and future research agenda. *Journal of Business Research*, *129*, 911–926. https://doi.org/10.1016/j.jbusres.2020.11.001

Mahima, K. T., T.N.D.S.Ginige, & Zoysa, K. de (2021). Evaluation of Sentiment Analysis based on AutoML and Traditional Approaches. *International Journal of Advanced Computer Science and Applications*, *12*(2). https://doi.org/10.14569/IJACSA.2021.0120277

Mena, P., Borrelli, R. A., & Kerby, L. (2022). Nuclear Reactor Transient Diagnostics Using Classification and AutoML. *Nuclear Technology*, *208*(2), 232–245. https://doi.org/10.1080/00295450.2021.1905470

Mitchell, M., Wu, S., Zaldivar, A., Barnes, P., Vasserman, L., Hutchinson, B., Spitzer, E., Raji, I. D., & Gebru, T. (2019). Model Cards for Model Reporting. In *Proceedings of the Conference on Fairness, Accountability, and Transparency* (pp. 220–229). ACM. https://doi.org/10.1145/3287560.3287596

Orlenko, A., Kofink, D., Lyytikäinen, L.-P., Nikus, K., Mishra, P., Kuukasjärvi, P., Karhunen, P. J., Kähönen, M., Laurikka, J. O., Lehtimäki, T., Asselbergs, F. W., & Moore, J. H. (2020). Model selection for metabolomics: Predicting diagnosis of coronary artery disease using automated machine learning. *Bioinformatics (Oxford, England)*, *36*(6), 1772–1778. https://doi.org/10.1093/bioinformatics/btz796

Ou, C., Liu, J., Qian, Y., Chong, W., Liu, D., He, X [Xuying], Zhang, X., & Duan, C.-Z. (2021). Automated Machine Learning Model Development for Intracranial Aneurysm Treatment Outcome Prediction: A Feasibility Study. *Frontiers in Neurology*, *12*, 735142. https://doi.org/10.3389/fneur.2021.735142

Ozkaya, I. (2020). What Is Really Different in Engineering AI-Enabled Systems? *IEEE Software*, *37*(4), 3–6. https://doi.org/10.1109/MS.2020.2993662

Panagopoulou, M., Karaglani, M., Manolopoulos, V. G., Iliopoulos, I., Tsamardinos, I., & Chatzaki, E. (2021). Deciphering the Methylation Landscape in Breast Cancer: Diagnostic and Prognostic Biosignatures through Automated Machine Learning. *Cancers*, *13*(7). https://doi.org/10.3390/cancers13071677

Peng, W.-L., Zhang, T.-J., Shi, K., Li, H.-X., Li, Y., He, S., Li, C., Xia, D., Xia, C.-C., & Li, Z.-L. (2022). Automatic machine learning based on native T1 mapping can identify myocardial fibrosis in patients with hypertrophic cardiomyopathy. *European Radiology*, *32*(2), 1044–1053. https://doi.org/10.1007/s00330-021-08228-7

Personal Data Protection Commission Singapore. (2020). *Model Artificial Intelligence Governance Framework*. https://www.pdpc.gov.sg/Help-and-Resources/2020/01/Model-AI-Governance-Framework

Sawaki, R., Sato, D., Nakayama, H., Nakagawa, Y., & Shimada, Y. (2019). ZF-AutoML: An Easy Machine-Learning-Based Method to Detect Anomalies in Fluorescent-Labelled Zebrafish. *Inventions*, *4*(4), 72. https://doi.org/10.3390/inventions4040072

Schafer, F., Zeiselmair, C., Becker, J., & Otten, H. (2018). Synthesizing CRISP-DM and Quality Management: A Data Mining Approach for Production Processes. In *2018 IEEE International Conference on Technology Management, Operations and Decisions (ICTMOD)* (pp. 190–195). IEEE. https://doi.org/10.1109/ITMC.2018.8691266

Schröer, C., Kruse, F., & Gómez, J. M. (2021). A Systematic Literature Review on Applying CRISP-DM Process Model. *Procedia Computer Science*, *181*, 526–534. https://doi.org/10.1016/j.procs.2021.01.199
Schulze-Brüninghoff, D., Wachendorf, M., & Astor, T. (2021). Potentials and Limitations of WorldView-3 Data for the Detection of Invasive Lupinus polyphyllus Lindl. in Semi-Natural Grasslands. *Remote Sensing*, *13*(21), 4333. https://doi.org/10.3390/rs13214333
Sills, M. R., Ozkaynak, M., & Jang, H. (2021). Predicting hospitalization of pediatric asthma patients in emergency departments using machine learning. *International Journal of Medical Informatics*, *151*, 104468. https://doi.org/10.1016/j.ijmedinf.2021.104468
Smith, R., Julian, D., & Dubin, A. (2022). Deep neural networks are effective tools for assessing performance during surgical training. *Journal of Robotic Surgery*, *16*(3), 559–562. https://doi.org/10.1007/s11701-021-01284-7
Sommerville, I. (2011). *Software engineering* (9th ed.). Pearson.
Stojadinovic, M., Milicevic, B., & Jankovic, S. (2021). Improved predictive performance of prostate biopsy collaborative group risk calculator when based on automated machine learning. *Computers in Biology and Medicine*, *138*, 104903. https://doi.org/10.1016/j.compbiomed.2021.104903
Su, X., Chen, N., Sun, H., Liu, Y., Yang, X., Wang, W., Zhang, S., Tan, Q., Su, J., Gong, Q., & Yue, Q. (2020). Automated machine learning based on radiomics features predicts H3 K27M mutation in midline gliomas of the brain. *Neuro-Oncology*, *22*(3), 393–401. https://doi.org/10.1093/neuonc/noz184
Sun, A. Y., Scanlon, B. R., Save, H., & Rateb, A. (2021). Reconstruction of GRACE Total Water Storage Through Automated Machine Learning. *Water Resources Research*, *57*(2). https://doi.org/10.1029/2020WR028666
Thalmann, S. (2018). Data driven decision support. *It - Information Technology*, *60*(4), 179–181. https://doi.org/10.1515/itit-2018-0017
Tran, L. M., Mocle, A. J., Ramsaran, A. I., Jacob, A. D., Frankland, P. W., & Josselyn, S. A. (2020). Automated Curation of CNMF-E-Extracted ROI Spatial Footprints and Calcium Traces Using Open-Source AutoML Tools. *Frontiers in Neural Circuits*, *14*, 42. https://doi.org/10.3389/fncir.2020.00042
Tsamardinos, I., Fanourgakis, G. S., Greasidou, E., Klontzas, E., Gkagkas, K., & Froudakis, G. E. (2020). An Automated Machine Learning architecture for the accelerated prediction of Metal-Organic Frameworks performance in energy and environmental applications. *Microporous and Mesoporous Materials*, *300*, 110160. https://doi.org/10.1016/j.micromeso.2020.110160
Tsiakmaki, M., Kostopoulos, G., Kotsiantis, S., & Ragos, O. (2020). Implementing AutoML in Educational Data Mining for Prediction Tasks. *Applied Sciences*, *10*(1), 90. https://doi.org/10.3390/app10010090
Vuković, M., & Thalmann, S. (2022). Causal Discovery in Manufacturing: A Structured Literature Review. *Journal of Manufacturing and Materials Processing*, *6*(1), 10. https://doi.org/10.3390/jmmp6010010
Walcott, T. H., & Ali, M. (2021). Machine Learning for Smaller Firms: Challenges and Opportunities. In *2021 International Conference on Computing, Electronics & Communications Engineering (iCCECE)* (pp. 82–86). IEEE. https://doi.org/10.1109/iCCECE52344.2021.9534852
Wan, K. W., Wong, C. H., Ip, H. F., Fan, D., Yuen, P. L., Fong, H. Y., & Ying, M. (2021). 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*(4), 1381–1393. https://doi.org/10.21037/qims-20-922
Wan, Z., Xia, X., Lo, D., & Murphy, G. C. (2020). How does Machine Learning Change Software Development Practices? *IEEE Transactions on Software Engineering*, 1. https://doi.org/10.1109/TSE.2019.2937083
Wang, D., Weisz, J. D., Muller, M., Ram, P., Geyer, W., Dugan, C., Tausczik, Y., Samulowitz, H., & Gray, A. (2019). Human-AI Collaboration in Data Science. *Proceedings of the ACM on Human-Computer Interaction*, *3*(CSCW), 1–24. https://doi.org/10.1145/3359313
Wang, S [Shuo], Niu, S., Qu, E., Forsberg, F., Wilkes, A., Sevrukov, A., Nam, K., Mattrey, R. F., Ojeda-Fournier, H., & Eisenbrey, J. R. (2020). Characterization of indeterminate breast lesions on B-mode ultrasound using automated machine learning models. *Journal of Medical Imaging*, *7*(05). https://doi.org/10.1117/1.JMI.7.5.057002

Webster, J., & Watson, R. T. (2002). Analyzing the past to prepare for the future: Writing a literature review. *MIS Quarterly*, *26*(2), 13–23. https://www.jstor.org/stable/4132319
Westenberger, J., Schuler, K., & Schlegel, D. (2022). Failure of AI projects: Understanding the critical factors. *Procedia Computer Science*, *196*, 69–76. https://doi.org/10.1016/j.procs.2021.11.074
Wirth, R., & Hipp, J. (2000). CRISP-DM: Towards a standard process model for data mining. In *Proceedings of the 4th international conference on the practical applications of knowledge discovery and data mining* (Vol. 1, pp. 29–39).
Xanthopoulos, I., Tsamardinos, I., Christophides, V., Simon, E., & Salinger, A. (2020). Putting the Human Back in the AutoML Loop. *EDBT/ICDT Workshops*.
Xin, D., Wu, E. Y., Lee, D. J.-L., Salehi, N., & Parameswaran, A. (2021). Whither AutoML? Understanding the Role of Automation in Machine Learning Workflows. In Y. Kitamura, A. Quigley, K. Isbister, T. Igarashi, P. Bjørn, & S. Drucker (Eds.), *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (pp. 1–16). ACM. https://doi.org/10.1145/3411764.3445306
Yang, H.-S., Kim, K.-R., Kim, S., & Park, J.-Y. (2021). Deep Learning Application in Spinal Implant Identification. *Spine*, *46*(5), E318-E324. https://doi.org/10.1097/BRS.0000000000003844
Yang, P., Zhang, H., Lai, X., Wang, K., Yang, Q [Qingyuan], & Yu, D. (2021). Accelerating the Selection of Covalent Organic Frameworks with Automated Machine Learning. *ACS Omega*, *6*(27), 17149–17161. https://doi.org/10.1021/acsomega.0c05990
Yang, Q [Qian], Suh, J., Chen, N.-C., & Ramos, G. (2018). Grounding Interactive Machine Learning Tool Design in How Non-Experts Actually Build Models. In I. Koskinen, Y. Lim, T. Cerratto-Pargman, K. Chow, & W. Odom (Eds.), *Proceedings of the 2018 Designing Interactive Systems Conference* (pp. 573–584). ACM. https://doi.org/10.1145/3196709.3196729
Zhang, C., & Ye, Z. (2021). Water pipe failure prediction using AutoML. *Facilities*, *39*(1/2), 36–49. https://doi.org/10.1108/F-08-2019-0084
Zhang, Q., Hu, W., Liu, Z., & Tan, J. (2020). TBM performance prediction with Bayesian optimization and automated machine learning. *Tunnelling and Underground Space Technology*, *103*, 103493. https://doi.org/10.1016/j.tust.2020.103493
Zhu, R., Hu, X., Hou, J., & Li, X. (2021). Application of machine learning techniques for predicting the consequences of construction accidents in China. *Process Safety and Environmental Protection*, *145*, 293–302. https://doi.org/10.1016/j.psep.2020.08.006
Zöller, M.-A., & Huber, M. F. (2021). Benchmark and Survey of Automated Machine Learning Frameworks. *Journal of Artificial Intelligence Research*, *70*, 409–472. https://doi.org/10.1613/jair.1.11854