The Roles and Modes of Human Interactions with Automated Machine Learning Systems: A Critical Review and Perspectives

Other titles in Foundations and Trends® in Human-Computer Interaction

Haptics for Human-Computer Interaction: From the Skin to the Brain
Mounia Ziat
ISBN: 978-1-63828-146-7

Human-Computer Interaction in Industry: A Systematic Review on the Applicability and Value-added of Operator Assistance Systems
Mirco Moencks, Elisa Roth, Thomas Bohné and Per Ola Kristensson
ISBN: 978-1-63828-122-1

Modes of Uncertainty in HCI
Robert Soden, Laura Devendorf, Richmond Wong, Yoko Akama and Ann Light
ISBN: 978-1-63828-054-5

A Design Space of Sports Interaction Technology
Dees B.W. Postma, Robby W. van Delden, Jeroen H. Koekoek, Wytse W. Walinga, Ivo M. van Hilvoorde, Bert Jan F. van Beijnum, Fahim A. Salim and Dennis Reidsma
ISBN: 978-1-63828-064-4

Designs on Transcendence: Sketches of a TX machine
Mark Blythe and Elizabeth Buie
ISBN: 978-1-68083-846-6

The Roles and Modes of Human Interactions with Automated Machine Learning Systems: A Critical Review and Perspectives

Thanh Tung Khuat
University of Technology Sydney
thanhtung.khuat@uts.edu.au

David Jacob Kedziora
University of Technology Sydney
david.kedziora@uts.edu.au

Bogdan Gabrys
University of Technology Sydney
bogdan.gabrys@uts.edu.au

now
the essence of knowledge
Boston — Delft

Foundations and Trends® in Human-Computer Interaction

Published, sold and distributed by:
now Publishers Inc.
PO Box 1024
Hanover, MA 02339
United States
Tel. +1-781-985-4510
www.nowpublishers.com
sales@nowpublishers.com

Outside North America:
now Publishers Inc.
PO Box 179
2600 AD Delft
The Netherlands
Tel. +31-6-51115274

The preferred citation for this publication is

T. T. Khuat *et al.*. *The Roles and Modes of Human Interactions with Automated Machine Learning Systems: A Critical Review and Perspectives*. Foundations and Trends® in Human-Computer Interaction, vol. 17, no. 3-4, pp. 195–387, 2023.

ISBN: 978-1-63828-269-3

Foundations and Trends® in Human-Computer Interaction
Volume 17, Issue 3-4, 2023

Editorial Board

Editorial Scope

Topics

Foundations and Trends® in Human-Computer Interaction publishes survey and tutorial articles in the following topics:

- History of the research community
- Theory
- Technology
- Computer Supported Cooperative Work
- Interdisciplinary influence
- Advanced topics and trends

Information for Librarians

Foundations and Trends® in Human-Computer Interaction, 2023, Volume 17, 4 issues. ISSN paper version 1551-3955. ISSN online version 1551-3963. Also available as a combined paper and online subscription.

Contents

The Roles and Modes of Human Interactions with Automated Machine Learning Systems: A Critical Review and Perspectives

Thanh Tung Khuat, David Jacob Kedziora and Bogdan Gabrys

Complex Adaptive Systems Lab, University of Technology Sydney, Australia; thanhtung.khuat@uts.edu.au, david.kedziora@uts.edu.au, bogdan.gabrys@uts.edu.au

ABSTRACT

As automated machine learning (AutoML) systems continue to progress in both sophistication and performance, it becomes important to understand the 'how' and 'why' of human-computer interaction (HCI) within these frameworks, both current and expected. Such a discussion is necessary for optimal system design, leveraging advanced data-processing capabilities to support decision-making involving humans, but it is also key to identifying the opportunities and risks presented by ever-increasing levels of machine autonomy. Within this context, we focus on the following questions: (i) What does HCI currently look like for state-of-the-art AutoML algorithms, especially during the stages of development, deployment, and maintenance? (ii) Do the expectations of HCI within AutoML frameworks vary for different types of users and stakeholders? (iii) How can HCI be managed so that AutoML solutions acquire human trust and

Thanh Tung Khuat, David Jacob Kedziora and Bogdan Gabrys (2023), "The Roles and Modes of Human Interactions with Automated Machine Learning Systems: A Critical Review and Perspectives", Foundations and Trends® in Human-Computer Interaction: Vol. 17, No. 3-4, pp 195–387. DOI: 10.1561/1100000091.

broad acceptance? (iv) As AutoML systems become more autonomous and capable of learning from complex open-ended environments, will the fundamental nature of HCI evolve? To consider these questions, we project existing literature in HCI into the space of AutoML; this connection has, to date, largely been unexplored. In so doing, we review topics including user-interface design, human-bias mitigation, and trust in artificial intelligence (AI). Additionally, to rigorously gauge the future of HCI, we contemplate how AutoML may manifest in effectively open-ended environments. This discussion necessarily reviews projected developmental pathways for AutoML, such as the incorporation of high-level reasoning, although the focus remains on how and why HCI may occur in such a framework rather than on any implementational details. Ultimately, this review serves to identify key research directions aimed at better facilitating the roles and modes of human interactions with both current and future AutoML systems.

1

Introduction

Broad interest in machine learning (ML) has ebbed and flowed ever since the 1950s, but recent years have arguably witnessed a new phase in the history of the field: an unprecedented level of technological uptake and engagement by the mainstream. From deepfakes for memes to recommendation systems for commerce, ML has become a regular fixture in broader society. Unsurprisingly though, this ongoing transition from purely academic confines is not smooth; the general public does not have the extensive expertise in data science required to fully exploit the capabilities of ML.

The ideal solution for democratisation is to make the application of ML optionally independent of human involvement. This is the primary goal of automated/autonomous machine learning (AutoML/AutonoML), an endeavour that, despite a rich multi-faceted history (Kedziora *et al.*, 2020), has only truly taken off within the last decade. Of course, ML itself already involves automation, relying on computers mechanically processing algorithms to build models from sample data. Thus, it is essential to note that the meaning of AutoML has come to encompass operations *around and beyond* fitting a specific model. Even then, the definition of AutoML can be fuzzy in the literature. Some lean

towards user-centric commentary, describing AutoML by its goal of empowering domain experts to effortlessly construct ML applications without relying on data scientists, even with limited expertise in statistics and ML (Yao *et al.*, 2018; Zöller and Huber, 2021). Others focus on the technical operations themselves, e.g. defining AutoML in terms of mechanising ML pipeline construction, with tasks including data preparation, feature engineering, hyperparameter optimisation, and model selection (He *et al.*, 2021). Notably, though, there has yet to be a consensus on the limits of AutoML, with some, for instance, including automated model deployment (Waring *et al.*, 2020). In short, originally popularised by significant optimisation advances applied narrowly to model selection (Thornton *et al.*, 2013; Swearingen *et al.*, 2017; Salvador *et al.*, 2019), the scope of AutoML has since expanded to automating all aspects of an ML application. The AutonoML term was even recently coined to differentiate one-and-done ML solutions from next-generation continuous learners that, in principle, could operate without human intervention *ad infinitum*. Such is the promise of AutoML/AutonoML: as long as there is a will and a way, it seems inevitable that ML systems will move ever closer towards autonomy.

As of the early 2020s, much has been written specifically around mechanisms and integrated systems for automating operations in both general ML (Kedziora *et al.*, 2020) and the fashionable subclass known as deep learning (DL) (Dong *et al.*, 2021); the topic of mechanising the latter is abbreviated as AutoDL. These discussions have mostly taken the notion of 'automation' to heart, wrestling with the challenges of how computers can make high-level decisions on their own. However, one important topic has been left underexplored: how do humans fit into the picture? This is crucial to consider, as, no matter how far its capacity for autonomous function evolves, the purpose of an AutoML system is to support human decision-making. Thus, perhaps counterintuitively, interactions cannot be an afterthought (Amershi *et al.*, 2019).

Even with an academic focus on model accuracy and algorithmic efficiency, systems cannot be considered optimal if they do not welcome and make use of optional human input. Moreover, beyond academia, the concept of 'performant' ML becomes much more complex and user-centred (Scriven *et al.*, 2022); the most promising algorithms and

architectures will likely be ones that flexibly tailor outputs to satisfy a very broad set of requirements. Then there is debate on just how much autonomy ML systems should be given. While the nature of the human-system relationship may eventually become one of collaboration (Wang *et al.*, 2019a), it is unlikely that humans will ever relinquish supervisory oversight (Endsley, 2017). Many researchers have echoed similar opinions, stating that human experience is indispensable and AI cannot be expected to autonomously operate in a socially responsible way (Xin *et al.*, 2021). For this multitude of reasons, a holistic appreciation of AutoML requires an associated study of human-computer interaction (HCI).

This is a rich topic; the nature of human interactions with AutoML, both in terms of roles and modes, has evolved and will continue to evolve alongside developments in the field. Consider, as an analogy to those developments, the history of artificial intelligence (AI) with respect to the game of chess. In the late 1960s, Mac Hack became the first chess program to play in human tournaments and even score a victory in doing so (Greenblatt *et al.*, 1967). Automated but heavily reliant on domain knowledge – it incorporated approximately 50 expertise-based heuristics – and hardly a threat to human dominance in chess, Mac Hack can be likened to proto-AutoML model-recommendation systems that were developed prior to the 2010s (Vanschoren, 2011; Serban *et al.*, 2013): novel and impressive for the time, but severely limited. Eventually though, by 1997, the swell of computational resources and advances in algorithmic techniques enabled a chess-playing computer known as Deep Blue to defeat a reigning world champion (Campbell *et al.*, 2002). As with the new wave of hyperparameter-optimising AutoML systems in the 2010s (Thornton *et al.*, 2013; Swearingen *et al.*, 2017; Salvador *et al.*, 2019), themselves becoming more and more capable of scaling competition leaderboards (Erickson *et al.*, 2020), Deep Blue heralded an era where computers would be far more competent than humans at performing a specific task.

Notably, even during the famed contest of 1997, Deep Blue was far from autonomous, leveraging a human-prescribed database of openings and endgame meta-knowledge, while also being manually adapted by grandmasters between games. Only in 2017, with the initial release of

AlphaZero (Zhang and Yu, 2020), has human input almost completely been removed, with an AI system autonomously learning to dominate humans in chess via self-play. In fact, the newest generation of chess-based AI has started to shift the roles of humans from mentors to students, with, for instance, an AI propensity for 'h-pawn thrusts' giving high-level players pause for thought (Miller *et al.*, 2020). The field of AutoML has not yet reached the same level of autonomy[1], but it is nonetheless worth asking: is this the state of interactions to plan for in the future? Will AutoML eventually produce more insight on how to solve an ML task than it currently receives?

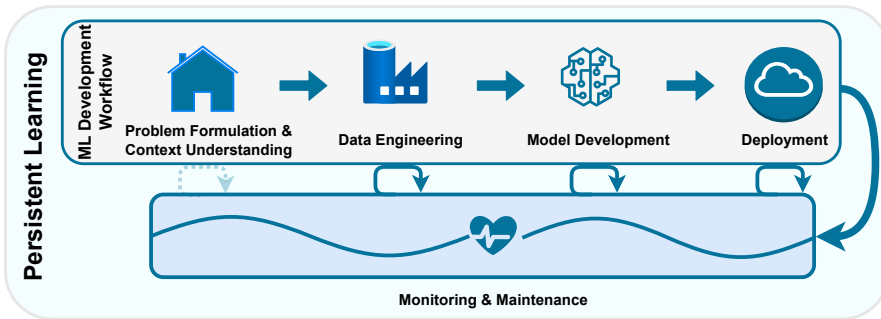

Figure 1.1: General schematic for a machine learning (ML) workflow, i.e. the operations involved in producing and maintaining an ML model for an ML application.

A comprehensive overview of HCI in AutoML, both current and prospective, needs to be carefully structured. For instance, human involvement in ML applications can be partitioned into two categories: productive and consumptive. While the latter refers to how end-users engage with and benefit from an ML model, the former relates to how such a model comes about. These 'productive' practices can be codified in many ways into an 'ML workflow' (Chapman *et al.*, 2000; Studer *et al.*, 2021), but one particular representation (Kedziora *et al.*, 2020; Dong *et al.*, 2021; Scriven *et al.*, 2022) is schematised in Figure 1.1.

[1]The AlphaZero approach is shared by AutoML systems such as AlphaD3M (Drori *et al.*, 2018), but, for the purpose of the chess analogy, the two are not equivalent; automation in chess is significantly more advanced than automation across an entire ML application.

Within the depiction of an ML workflow, it is clear that there are several phases of operations involved in developing, deploying and maintaining an ML model. Of these, the model development stage receives the most focus in academic AutoML research, especially in the case of DL and neural architecture search (NAS) (Dong *et al.*, 2021), but there have been numerous automation efforts applied to the rest of an archetypal ML workflow. In fact, the capacity to continuously monitor performance and adapt to dynamic changes in data environments has previously been highlighted as a key prerequisite for the transition between AutoML and AutonoML (Kedziora *et al.*, 2020). Relatedly, there exist many theoretical proposals for supporting online learning (Laird and Mohan, 2018), and initial experimental research towards making AutoML systems 'persistent' has recently been undertaken within academia (Bakirov *et al.*, 2021; Celik and Vanschoren, 2021). Meanwhile, in industry, the emerging trend of 'MLOps' reflects the importance of automated deployment to real-world demands (Scriven *et al.*, 2022). In essence, any developers interested in engineering a comprehensive AutoML/AutonoML system must understand the idiosyncrasies of each and every workflow phase, e.g. the associated formats of human inputs/guidance, the minimal requirements for baseline operations, the possible opportunities for additional human-assisted learning, and so on.

Alternatively, rather than partitioning HCI in AutoML by when an interaction occurs, it is sometimes more informative to consider who is undertaking the interaction. This perspective is particularly natural in commerce and industry (Scriven *et al.*, 2022), where, beyond the end-users that consume the outputs of an ML model, there are typically numerous stakeholders linked to the production of a model. These may include technicians in the form of data scientists or software developers, business staff in the form of project managers or domain experts, regulatory groups in the form of third-party auditors or government agencies, etc.

Importantly, the obligations and interests of different stakeholders are usually not mappable to individual stages within the ML workflow depicted in Figure 1.1. Moreover, the modes of their interactions may also vary substantially. Some roles will demand fine control over AutoML

processes, while others will simply require an entry point for inputs. Some roles will desire a window into the mechanisms involved, while others will only want to be alerted if things go wrong. Whatever the case may be, these requirements must be considered at the fundamental level of algorithm and architecture. It is not optimal to focus solely on predictor accuracy and efficiency during system design, only to service any remaining real-world expectations with a hasty patch-up.

Crucially, it is worth emphasising that the concept of a 'user' is intrinsic to the stakeholder view of AutoML. This may seem like an unnecessary distraction to those that are solely interested in improving the statistical theories underlying ML algorithms, but that attitude ignores the greater ecosystem in which ML operates: human decision-making. For instance, humans may be willing to tolerate disliking 40% of AI-recommended music, whereas a 20% false-positive rate for AI-recommended convictions is arguably abysmal. Simply put, human context matters more than any agnostic accuracy metric. It follows then that successfully translating ML model performance into real-world outcomes is contingent on a set of engagement-related requirements (Arrieta *et al.*, 2020; Shin, 2021; Drozdal *et al.*, 2020; Schmidt and Biessmann, 2019; Ehsan *et al.*, 2021), which we will bundle here under the title of 'user experience' (UX). This includes topics that have recently diffused into academic discussions around ML and AutoML, such as accessibility, transparency, fairness, reliability, and so on (Xin *et al.*, 2021).

Accordingly, the concept of a user interface (UI) – the implementation need not be monolithic – becomes particularly important to AutoML within the stakeholder perspective, as this is where UX can most directly be managed. Indeed, designing intelligent UIs is critical for supporting human-guided AutoML (Gil *et al.*, 2019; Lee *et al.*, 2019a), where a technical user might, as an ideal, tweak problem settings, explore data characteristics, limit model search spaces, etc. These interactions may also be constrained or facilitated by the manner in which they occur, e.g. via touch-screen, voice commands, gesture recognition, or even brain signals (Xing-Yu *et al.*, 2013). In short, the field of AutoML would be well served by greater discussions around the concept of interfacing so that, beyond simply enabling the control of

ML operations, users can both inject domain knowledge and extract comprehensible information with ease.

Turning to factors that influence UX, explainability is high up on the list. This is especially a challenge for AutoML, as the core principle of automation is to decouple humans from certain operations. It may seem wasteful then, if not counterproductive, to spend research effort in making those processes transparent, consequently encouraging humans to reinvolve themselves. Sure enough, many current AutoML tools are staunchly black-box systems (Xin *et al.*, 2021), veiling how ML models are built and how predictive/prescriptive outputs are generated. But herein lies the nuance; the aim of AutoML is to remove the *necessity* for human engagement, not the option. Thus, technical obscuration actually impedes ML performance if users cannot understand how to properly insert domain knowledge that would otherwise be beneficial to an ML task (Liu *et al.*, 2017). This is especially a drawback at the current point in time, because human-in-the-loop learning is still often more advantageous than machine-centric ML (Tam *et al.*, 2016).

Regardless, even if AutoML systems were to be completely autonomous, their innards untouched by humans, explainability is also necessary to promote trust (Marcus and Davis, 2019). Surveys indicate data scientists tend to be sceptical of ML models provided by AutoML tools if there are no mechanisms for transparency and understandability (Drozdal *et al.*, 2020). Similarly, end-users only follow ML recommendations if the system can show the reasoning behind them (Van der Waa *et al.*, 2021). This reticence by people to use results they cannot understand or explain can be frustrating for simple business applications, but it is completely warranted in high-stakes contexts (Rudin, 2019), including medical diagnosis, financial investment and criminal justice. To do otherwise could be disastrous. For example, adverse outcomes have been linked to the COMPAS recidivism prediction model (Dieterich *et al.*, 2016), the BreezoMeter real-time air-quality prediction model used by Google during the California wildfires in 2018 (McGough, 2018), and black-box medical diagnosis models in general (Duran and Jongsma, 2021).

Another factor that affects UX, even and especially if stakeholders are not directly aware that they are 'using' the results of ML, is fairness.

This socially conscious requirement has recently been taken up as an important issue by academic research (Zarsky, 2016; Caton and Haas, 2020; Mehrabi *et al.*, 2021), indicating just how far ML has embedded itself into the mainstream, and recognises that predictive/prescriptive accuracy and error may disproportionately affect different people in different ways. Now, granted, there have been efforts towards automating mechanisms for discovering and preventing discrimination in ML models (Hajian *et al.*, 2016), but the challenge is that there are many possible technical definitions for fairness, often orthogonal and sometimes contradictory (Verma and Rubin, 2018; Saxena *et al.*, 2019). Once again, human context matters. So, it is an open question as to how human oversight can best be integrated within an AutoML system, enforcing ethical requirements upon a mechanised process.

Of course, while every ML algorithm applies its own assumptions, many 'unfair' biases are often sourced from biological neurons, i.e. human brains. These can be injected into learning systems via data and knowledge, manifesting in both information content and sampling. As a result, human cognitive biases that are internalised can lead to a deterioration in model reliability, and there are many high-profile examples of this occurring (Hunt, 2016; Zemcik, 2021). The severity of these impacts can also vary depending on context. Healthcare is one example of a high-stakes environment, where cognitive biases in clinical practice can have a strong influence on medical outcomes (Saposnik *et al.*, 2016; Preisz, 2019). Indeed, similar flaws in predictive systems have been shown to hinder social minorities from receiving extra care services (Obermeyer *et al.*, 2019). Accordingly, there is an imperative to more thoroughly consider the nature and implementation of bias mitigation strategies within AutoML.

Fundamentally, the point of all this discussion is that, given a suitable conceptual framework, such as the dual workflow/stakeholder perspective of ML operations, it is possible to engage with many HCI-related issues that, unaddressed, will impede the heretofore surging pace of progress in AutoML. Moreover, this kind of systematic approach does not just clarify the current state of HCI in AutoML; it provides a lens through which the future of this trend in ML can be forecasted. This does not mean speculating on detailed implementations of HCI-

related mechanisms, but rather understanding the projected evolution of human-system interactions, particularly as algorithms and architectures become better at their job. Thus, the aforementioned chess analogy remains useful in illustrating this progression as AutoML systems shift further along the spectrum of autonomy (Simmler and Frischknecht, 2021).

However, it is still valuable to conjecture just a little bit further. AlphaGo (Silver *et al.*, 2016) and AlphaZero (Silver *et al.*, 2018) are extremely competent at their respective games, but they remain constrained within particular environments. An equivalent AutoML system would essentially be autonomous for every phase of the ML workflow in Figure 1.1 except for one holdout: problem formulation and context understanding. Such a constraint is not unexpected, as this phase is likely to be the last bastion of necessary human involvement in ML. Unfortunately, it does stand in the way of numerous AI applications. For instance, there exists plenty of research and development in the field of autonomous vehicles (Hawkins, 2018; Lechner *et al.*, 2020), yet the challenge of operating in unpredictable and effectively boundless driving environments remains, to date, daunting (Harel *et al.*, 2020). Nonetheless, without delving into the domain of artificial general intelligence, these constraints will eventually relax. The novel MuZero system (Schrittwieser *et al.*, 2020) is already emblematic of an emerging reinforcement-learning approach that can be applied agnostically to a variety of games with diverse rules, autonomously building competent models from first principles. In theory, cognitive models may eventually supercharge this process even further, enabling ML systems to efficiently transfer knowledge from one problem to another by actually *understanding* context, rather than by outright ignoring it. So, as AutoML truly becomes AutonoML, and then begins to relax into open-world learning: will human-system interactions change yet again?

As is evident by now, there are many important issues to consider in the overlap between AutoML and HCI. This review addresses these topics, marking the final part in a series of monographs dedicated to a systematic and conceptual overview of AutoML (Kedziora *et al.*, 2020; Dong *et al.*, 2021; Scriven *et al.*, 2022). Specifically, with the series having previously focussed on how computers may perform ML/DL

in the absence of humans (Kedziora *et al.*, 2020; Dong *et al.*, 2021), this work aims to recontextualise AutoML/AutonoML back within the ecosystem of human decision-making. In fact, because ML does not operate in a vacuum within the real world, some organically arising consequences of this interlinkage are already captured by the previous technological survey in the series (Scriven *et al.*, 2022). This review, however, dives much deeper into the fundamentals of HCI in AutoML, driven by the following set of questions:

- What does HCI currently look like for state-of-the-art AutoML algorithms, especially during the stages of development, deployment, and maintenance?
- Do the expectations of HCI within AutoML frameworks vary for different types of users and stakeholders?
- How can HCI be managed so that AutoML solutions acquire human trust and broad acceptance?
- As AutoML systems become more autonomous and capable of learning from complex open-ended environments, will the fundamental nature of HCI evolve?

To best grapple with these questions, the rest of this monograph is structured as follows. Section 2 examines HCI and state-of-the-art AutoML as of the early 2020s. It does so with respect to the workflow/stakeholder perspectives of AutoML, after these are first systematised. Modern approaches to UIs and current concerns around UX, e.g. in terms of explainability and fairness, are also surveyed. Section 3 then extrapolates progress in AutoML to where associated systems are effectively autonomous in all high-level ML operations, excluding problem formulation and context understanding. The evolution of HCI with respect to genuine high-performance AutonoML, albeit restricted to constrained environments, is considered. Section 4 follows by pushing this limit, relaxing restrictions and considering ML within open-ended environments. To anchor such a scenario, modern theories for incorporating high-level 'reasoning' in ML systems are surveyed and debated.

Subsequently, changes to the nature of HCI with respect to these upgraded forms of AutonoML systems are theorised upon. A synthesising discussion is then presented in Section 5, identifying existing issues and potential research directions that may hinder or facilitate the successful interplay of HCI and AutoML, both now and in the future. Finally, Section 6 concludes this review, summarising key findings and perspectives around the roles and modes of human interactions with AutoML/AutonoML systems.

References

Ackerman, E. (2021). "How the U.S. Army is turning robots into team players". *IEEE Spectrum*. URL: https://spectrum.ieee.org/ai-army-robots.

Agathokleous, M. and N. Tsapatsoulis. (2013). "Voting advice applications: missing value estimation using matrix factorization and collaborative filtering". In: *IFIP International Conference on Artificial Intelligence Applications and Innovations*. Springer. 20–29.

Ali, A. R., M. Budka, and B. Gabrys. (2019). "Towards meta-learning of deep architectures for efficient domain adaptation". In: *PRICAI 2019: Trends in Artificial Intelligence*. Cham: Springer International Publishing. 66–79.

Ali, A. R., M. Budka, and B. Gabrys. (2020). "A Review of Meta-level Learning in the Context of Multi-component, Multi-level Evolving Prediction Systems". *CoRR*. abs/2007.10818.

Amershi, S., M. Cakmak, W. B. Knox, and T. Kulesza. (2014). "Power to the people: The role of humans in interactive machine learning". *AI Magazine*. 35(4): 105–120.

Amershi, S., D. Weld, M. Vorvoreanu, A. Fourney, B. Nushi, P. Collisson, J. Suh, S. Iqbal, P. N. Bennett, K. Inkpen, *et al.* (2019). "Guidelines for human-AI interaction". In: *Proceedings of the 2019 chi conference on human factors in computing systems*. 1–13.

Apeh, E., B. Gabrys, and A. Schierz. (2014). "Customer profile classification: To adapt classifiers or to relabel customer profiles?" *Neurocomputing*. 132: 3–13.

Araujo, T., N. Helberger, S. Kruikemeier, and C. H. De Vreese. (2020). "In AI we trust? Perceptions about automated decision-making by artificial intelligence". *AI & SOCIETY*. 35(3): 611–623.

Arazi, O. (2020). "AI Won't Replace Radiologists, But It Will Change Their Work. Here's How". *World Economic Forum*. URL: https://www.weforum.org/agenda/2020/10/how-ai-will-change-how-radiologists-work/.

Arjovsky, M., S. Chintala, and L. Bottou. (2017). "Wasserstein generative adversarial networks". In: *International conference on machine learning*. PMLR. 214–223.

Arnold, T. and M. Scheutz. (2018). "The "big red button" is too late: an alternative model for the ethical evaluation of AI systems". *Ethics and Information Technology*. 20(1): 59–69.

Arrieta, A. B., N. Diaz-Rodriguez, J. Del Ser, A. Bennetot, S. Tabik, A. Barbado, S. Garcia, S. Gil-Lopez, D. Molina, R. Benjamins, *et al.* (2020). "Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI". *Information Fusion*. 58: 82–115.

Azure. (2021). "Interpretability: Model explainability in automated ML". URL: https://docs.microsoft.com/en-us/azure/machine-learning/how-to-machine-learning-interpretability-automl.

Bakirov, R., D. Fay, and B. Gabrys. (2021). "Automated adaptation strategies for stream learning". *Machine Learning*: 1–34.

Bakirov, R., B. Gabrys, and D. Fay. (2017). "Multiple adaptive mechanisms for data-driven soft sensors". *Computers & Chemical Engineering*. 96: 42–54.

Baldassarre, G., V. G. Santucci, E. Cartoni, and D. Caligiore. (2017). "The architecture challenge: Future artificial intelligence systems will require sophisticated architectures, and knowledge of the brain might guide their construction". *Behavioral and Brain Sciences*. 40.

Barbu, A., D. Mayo, J. Alverio, W. Luo, C. Wang, D. Gutfreund, J. Tenenbaum, and B. Katz. (2019). "ObjectNet: A large-scale bias-controlled dataset for pushing the limits of object recognition models". In: *Advances in Neural Information Processing Systems*. Vol. 32.

Bargiela, A. and W. Pedrycz. (2016). "Granular computing". In: *Handbook on computational intelligence: Fuzzy Logic, Systems, Artificial Neural Networks, and Learning Systems*. World Scientific. 43–66.

Baroroh, D. K., C.-H. Chu, and L. Wang. (2021). "Systematic literature review on augmented reality in smart manufacturing: collaboration between human and computational intelligence". *Journal of Manufacturing Systems*. 61: 696–711.

Baylor, D., E. Breck, H.-T. Cheng, N. Fiedel, C. Y. Foo, Z. Haque, S. Haykal, M. Ispir, V. Jain, L. Koc, *et al.* (2017). "Tfx: A tensorflow-based production-scale machine learning platform". In: *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 1387–1395.

Bellamy, R. K., K. Dey, M. Hind, S. C. Hoffman, S. Houde, K. Kannan, P. Lohia, J. Martino, S. Mehta, A. Mojsilovic, *et al.* (2019). "AI Fairness 360: An extensible toolkit for detecting and mitigating algorithmic bias". *IBM Journal of Research and Development*. 63(4/5): 4:1–4:15.

Benderius, O., C. Berger, and V. M. Lundgren. (2017). "The best rated human–machine interface design for autonomous vehicles in the 2016 grand cooperative driving challenge". *IEEE Transactions on intelligent transportation systems*. 19(4): 1302–1307.

Bengio, Y., T. Deleu, N. Rahaman, R. Ke, S. Lachapelle, O. Bilaniuk, A. Goyal, and C. Pal. (2019). "A meta-transfer objective for learning to disentangle causal mechanisms". In: *Proceedings of the 7th International Conference on Learning Representations (ICLR 2019)*.

Berthold, M. R., N. Cebron, F. Dill, T. R. Gabriel, T. Kotter, T. Meinl, P. Ohl, K. Thiel, and B. Wiswedel. (2009). "KNIME-the Konstanz information miner: version 2.0 and beyond". *AcM SIGKDD Explorations Newsletter*. 11(1): 26–31.

Besold, T. R. and K.-U. Kuhnberger. (2015). "Towards integrated neural–symbolic systems for human-level AI: Two research programs helping to bridge the gaps". *Biologically Inspired Cognitive Architectures*. 14: 97–110.

Besold, T. R., K.-U. Kuhnberger, A. d. Garcez, A. Saffiotti, M. H. Fischer, and A. Bundy. (2015). "Anchoring knowledge in interaction: Towards a harmonic subsymbolic/symbolic framework and architecture of computational cognition". In: *International Conference on Artificial General Intelligence*. Springer. 35–45.

Biessmann, F., D. Salinas, S. Schelter, P. Schmidt, and D. Lange. (2018). "Deep Learning for Missing Value Imputation in Tables with Non-Numerical Data". In: *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*. 2017–2025.

Blackhurst, J. L., J. S. Gresham, and M. O. Stone. (2011). "The autonomy paradox". *Armed Forces Journal*: 20–24.

Blawatt, K. R. (2016). "Appendix A: List of Cognitive Biases". In: *Marconomics*. Emerald Group Publishing Limited. 325–336.

Boboc, R. G., F. Girbacia, and E. V. Butila. (2020). "The application of augmented reality in the automotive industry: A systematic literature review". *Applied Sciences*. 10(12): 4259.

Boden, M., J. Bryson, D. Caldwell, K. Dautenhahn, L. Edwards, S. Kember, P. Newman, V. Parry, G. Pegman, T. Rodden, *et al.* (2017). "Principles of robotics: regulating robots in the real world". *Connection Science*. 29(2): 124–129.

Bommasani, R., D. A. Hudson, E. Adeli, R. Altman, S. Arora, S. von Arx, M. S. Bernstein, and et al. (2021). "On the Opportunities and Risks of Foundation Models". *CoRR*. abs/2108.07258. URL: https://arxiv.org/abs/2108.07258.

Booch, G., F. Fabiano, L. Horesh, K. Kate, J. Lenchner, N. Linck, A. Loreggia, K. Murugesan, N. Mattei, F. Rossi, and B. Srivastava. (2021). "Thinking Fast and Slow in AI". In: *Proceedings of the 35th AAAI Conference on Artificial Intelligence*. Vol. 35. No. 17. 15042–15046.

Botvinick, M., D. G. Barrett, P. Battaglia, N. de Freitas, D. Kumaran, J. Z. Leibo, T. Lillicrap, J. Modayil, S. Mohamed, N. C. Rabinowitz, *et al.* (2017). "Building machines that learn and think for themselves". *Behavioral and Brain Sciences*. 40.

Bradshaw, J. M., R. R. Hoffman, D. D. Woods, and M. Johnson. (2013). "The seven deadly myths of autonomous systems". *IEEE Intelligent Systems*. 28(3): 54–61.

Braun, D. A., C. Mehring, and D. M. Wolpert. (2010). "Structure learning in action". *Behavioural brain research*. 206(2): 157–165.

Breque, M., L. De Nul, and A. Petridis. (2021). "Industry 5.0: towards a sustainable, human-centric and resilient European industry". *European Commission, Directorate-General for Research and Innovation*.

Brooks, R. (2019). "A Better Lesson". URL: https://rodneybrooks.com/a-better-lesson/.

Brooks, R. (2020). "An Analogy For The State Of AI". URL: https://rodneybrooks.com/an-analogy-for-the-state-of-ai/.

Brown, T. B., B. Mann, N. Ryder, M. Subbiah, J. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell, S. Agarwal, A. Herbert-Voss, G. Kruger, T. Henighan, R. Child, A. Ramesh, D. M. Ziegler, J. Wu, C. Winter, C. Hesse, M. Chen, E. Sigler, M. Litwin, S. Gray, B. Chess, J. Clark, C. Berner, S. McCandlish, A. Radford, I. Sutskever, and D. Amodei. (2020). "Language Models are Few-Shot Learners". In: *Proceedings of the 34th Conference on Neural Information Processing Systems (NeurIPS)*.

Budka, M. and B. Gabrys. (2010). "Ridge regression ensemble for toxicity prediction". *Procedia Computer Science*. 1(1): 193–201.

Budka, M., B. Gabrys, and E. Ravagnan. (2010). "Robust predictive modelling of water pollution using biomarker data". *Water research*. 44(10): 3294–3308.

Byrne, R. M. (2019). "Counterfactuals in Explainable Artificial Intelligence (XAI): Evidence from Human Reasoning". In: *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence*. 6276–6282.

Calders, T., E. Ntoutsi, M. Pechenizkiy, B. Rosenhahn, and S. Ruggieri. (2021). "Call for Papers: Special issue on Bias and Fairness in AI". *Data Mining and Knowledge Discovery*. URL: https://www.springer.com/journal/10618/updates/19143254.

Campbell, M., A. J. Hoane Jr, and F.-h. Hsu. (2002). "Deep Blue". *Artificial Intelligence*. 134(1-2): 57–83.

Cao, Y., X. Qian, T. Wang, R. Lee, K. Huo, and K. Ramani. (2020). "An Exploratory Study of Augmented Reality Presence for Tutoring Machine Tasks". In: *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–13.

Cao, Y., T. Wang, X. Qian, P. S. Rao, M. Wadhawan, K. Huo, and K. Ramani. (2019). "GhostAR: A time-space editor for embodied authoring of human-robot collaborative task with augmented reality". In: *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology*. 521–534.

Carreto, C., D. Gêgo, and L. Figueiredo. (2018). "An eye-gaze tracking system for teleoperation of a mobile robot". *Journal of Information Systems Engineering & Management*. 3(2): 16.

Cashman, D., S. R. Humayoun, F. Heimerl, K. Park, S. Das, J. Thompson, B. Saket, A. Mosca, J. Stasko, A. Endert, *et al.* (2019). "A User-based Visual Analytics Workflow for Exploratory Model Analysis". *Computer Graphics Forum*. 38(3): 185–199.

Caton, S. and C. Haas. (2020). "Fairness in machine learning: A survey". *arXiv preprint arXiv:2010.04053*.

Celik, B. and J. Vanschoren. (2021). "Adaptation strategies for automated machine learning on evolving data". *IEEE Transactions on Pattern Analysis and Machine Intelligence*.

Chandrashekar, G. and F. Sahin. (2014). "A survey on feature selection methods". *Computers & Electrical Engineering*. 40(1): 16–28.

Chapman, P., J. Clinton, R. Kerber, T. Khabaza, T. Reinartz, C. Shearer, and R. Wirth. (2000). *CRISP-DM 1.0: Step-by-step data mining guide*. SPSS.

Chen, M. and A. Golan. (2015). "What may visualization processes optimize?" *IEEE transactions on visualization and computer graphics*. 22(12): 2619–2632.

Chen, Z. and B. Liu. (2018). "Lifelong machine learning". *Synthesis Lectures on Artificial Intelligence and Machine Learning*. 12(3): 1–207.

Cheng, L., K. R. Varshney, and H. Liu. (2021). "Socially Responsible AI Algorithms: Issues, Purposes, and Challenges". *Journal of Artificial Intelligence Research*. 71: 1137–1181.

Chouldechova, A. (2017). "Fair prediction with disparate impact: A study of bias in recidivism prediction instruments". *Big data*. 5(2): 153–163.

Cloud, G. (2021). "Explaining predictions". URL: https://cloud.google.com/automl-tables/docs/explain.

Crandall, J. W. and M. A. Goodrich. (2002). "Characterizing efficiency of human robot interaction: A case study of shared-control teleoperation". In: *IEEE/RSJ international conference on intelligent robots and systems*. Vol. 2. IEEE. 1290–1295.

Crotty, A., A. Galakatos, E. Zgraggen, C. Binnig, and T. Kraska. (2015). "Vizdom: interactive analytics through pen and touch". *Proceedings of the VLDB Endowment*. 8(12): 2024–2027.

d'Alessandro, B., C. O'Neil, and T. LaGatta. (2017). "Conscientious classification: A data scientist's guide to discrimination-aware classification". *Big data*. 5(2): 120–134.

Dai, W. and M. G. Genton. (2018). "Multivariate functional data visualization and outlier detection". *Journal of Computational and Graphical Statistics*. 27(4): 923–934.

Dale, R. (2021). "GPT-3: What's it good for?" *Natural Language Engineering*. 27(1): 113–118.

DataRobot. (2021). "Automated Machine Learning Platform". URL: https://www.datarobot.com/platform/automated-machine-learning/.

Deuschel, T. and T. Scully. (2016). "On the Importance of Spatial Perception for the Design of Adaptive User Interfaces". In: *Proceedings of the 10th International Conference on Self-Adaptive and Self-Organizing Systems (SASO)*. IEEE. 70–79.

Dieterich, W., C. Mendoza, and T. Brennan. (2016). "COMPAS risk scales: Demonstrating accuracy equity and predictive parity". *Northpointe Inc*.

Dignum, V. (2017). "Responsible Autonomy". In: *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI-17*. 4698–4704.

Dignum, V. (2018). "Ethics in artificial intelligence: introduction to the special issue". *Ethics and Information Technology*. 20(1): 1–3.

Domingos, P. (2012). "A Few Useful Things to Know about Machine Learning". *Communications of the ACM*. 55(10): 78–87.

Dong, X., D. J. Kedziora, K. Musial, and B. Gabrys. (2021). "Automated Deep Learning: Neural Architecture Search Is Not the End". *arXiv preprint arXiv:2112.09245*.

Dreyfus, H., S. E. Dreyfus, and T. Athanasiou. (2000). *Mind over machine*. Simon and Schuster.

Drori, I., Y. Krishnamurthy, R. Rampin, R. Lourenço, J. Ono, K. Cho, C. Silva, and J. Freire. (2018). "AlphaD3M: Machine learning pipeline synthesis". In: *AutoML Workshop at ICML*.

Drozdal, J., J. Weisz, D. Wang, G. Dass, B. Yao, C. Zhao, M. Muller, L. Ju, and H. Su. (2020). "Trust in AutoML: exploring information needs for establishing trust in automated machine learning systems". In: *Proceedings of the 25th International Conference on Intelligent User Interfaces*. 297–307.

Duran, J. M. and K. R. Jongsma. (2021). "Who is afraid of black box algorithms? On the epistemological and ethical basis of trust in medical AI". *Journal of Medical Ethics*. 47(5): 329–335.

Ehsan, U. (2021). "Towards Human-Centered Explainable AI: the journey so far". *The Gradient*.

Ehsan, U., B. Harrison, L. Chan, and M. O. Riedl. (2018). "Rationalization: A neural machine translation approach to generating natural language explanations". In: *Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society*. 81–87.

Ehsan, U., Q. V. Liao, M. Muller, M. O. Riedl, and J. D. Weisz. (2021). "Expanding explainability: Towards social transparency in AI systems". In: *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–19.

Ehsan, U., P. Tambwekar, L. Chan, B. Harrison, and M. O. Riedl. (2019). "Automated rationale generation: a technique for explainable AI and its effects on human perceptions". In: *Proceedings of the 24th International Conference on Intelligent User Interfaces*. 263–274.

Emmanouilidis, C., P. Pistofidis, L. Bertoncelj, V. Katsouros, A. Fournaris, C. Koulamas, and C. Ruiz-Carcel. (2019). "Enabling the human in the loop: Linked data and knowledge in industrial cyber-physical systems". *Annual reviews in control*. 47: 249–265.

Endsley, M. R. (1995). "Toward a theory of situation awareness in dynamic systems". *Human Factors*. 37(1): 32–64.

Endsley, M. R. (2017). "From here to autonomy: lessons learned from human–automation research". *Human factors*. 59(1): 5–27.

Endsley, M. (2011). "Bringing cognitive engineering to the information fusion problem: Creating systems that understand situations". In: *Plenary Presentation to the 14th International Conference on Information Fusion*.

Erickson, N., J. Mueller, A. Shirkov, H. Zhang, P. Larroy, M. Li, and A. Smola. (2020). "Autogluon-tabular: Robust and accurate automl for structured data". *arXiv preprint arXiv:2003.06505*.

Fang, D., H. Xu, X. Yang, and M. Bian. (2020). "An augmented reality-based method for remote collaborative real-time assistance: from a system perspective". *Mobile Networks and Applications*. 25(2): 412–425.

Fedus, W., B. Zoph, and N. Shazeer. (2021). "Switch Transformers: Scaling to Trillion Parameter Models with Simple and Efficient Sparsity". *CoRR*. abs/2101.03961. URL: https://arxiv.org/abs/2101.03961.

Fei, G., S. Wang, and B. Liu. (2016). "Learning cumulatively to become more knowledgeable". In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 1565–1574.

Fekete, J.-D. (2004). "The infovis toolkit". In: *IEEE Symposium on Information Visualization*. 167–174.

Feurer, M., K. Eggensperger, S. Falkner, M. Lindauer, and F. Hutter. (2021). "Auto-Sklearn 2.0: The Next Generation". *CoRR*. abs/2007.04074. URL: https://arxiv.org/abs/2007.04074.

Feurer, M., A. Klein, K. Eggensperger, J. Springenberg, M. Blum, and F. Hutter. (2015a). "Efficient and Robust Automated Machine Learning". In: *Advances in Neural Information Processing Systems*. Vol. 28.

Feurer, M., J. Springenberg, and F. Hutter. (2015b). "Initializing bayesian hyperparameter optimization via meta-learning". In: *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 29. No. 1. 1128–1135.

Fischer, L., L. Ehrlinger, V. Geist, R. Ramler, F. Sobiezky, W. Zellinger, D. Brunner, M. Kumar, and B. Moser. (2021). "AI System Engineering—Key Challenges and Lessons Learned". *Machine Learning and Knowledge Extraction*. 3(1): 56–83.

Fjelland, R. (2020). "Why general artificial intelligence will not be realized". *Humanities and Social Sciences Communications*. 7(1): 1–9.

Furnkranz, J. (2005). "From local to global patterns: Evaluation issues in rule learning algorithms". In: *Local pattern detection*. Springer. 20–38.

Furnkranz, J., T. Kliegr, and H. Paulheim. (2020). "On cognitive preferences and the plausibility of rule-based models". *Machine Learning*. 109(4): 853–898.

Gabrys, B. (2005). "Do Smart Adaptive Systems Exist?—Introduction". In: *Do Smart Adaptive Systems Exist?* Springer. 1–17.

Gabrys, B. and A. Bargiela. (2000). "General fuzzy min-max neural network for clustering and classification". *IEEE transactions on neural networks*. 11(3): 769–783.

Gajos, K. Z., K. Everitt, D. S. Tan, M. Czerwinski, and D. S. Weld. (2008). "Predictability and accuracy in adaptive user interfaces". In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 1271–1274.

Gallistel, C. R. and A. P. King. (2011). *Memory and the computational brain: Why cognitive science will transform neuroscience*. Vol. 6. John Wiley & Sons.

Gama, J., I. Zliobaite, A. Bifet, M. Pechenizkiy, and A. Bouchachia. (2014). "A survey on concept drift adaptation". *ACM computing survey*. 46(4): 1–37.

Garcez, A. d. and L. C. Lamb. (2020). "Neurosymbolic AI: The 3rd Wave". *arXiv preprint arXiv:2012.05876*.

Garciarena, U., R. Santana, and A. Mendiburu. (2018). "Analysis of the complexity of the automatic pipeline generation problem". In: *Proccedings of the IEEE Congress on Evolutionary Computation (CEC)*. IEEE. 1–8.

Gauci, J., N. Cauchi, K. Theuma, D. Zammit-Mangion, and A. Muscat. (2015). "Design and evaluation of a touch screen concept for pilot interaction with avionic systems". In: *2015 IEEE/AIAA 34th Digital Avionics Systems Conference (DASC)*. IEEE. 3C2–1.

Geist, V., M. Moser, J. Pichler, R. Santos, and V. Wieser. (2021). "Leveraging machine learning for software redocumentation—A comprehensive comparison of methods in practice". *Software: Practice and Experience*. 51(4): 798–823.

Gennatas, E. D., J. H. Friedman, L. H. Ungar, R. Pirracchio, E. Eaton, L. G. Reichmann, Y. Interian, J. M. Luna, C. B. Simone, A. Auerbach, *et al.* (2020). "Expert-augmented machine learning". *Proceedings of the National Academy of Sciences*. 117(9): 4571–4577.

Getoor, L. (2019). "Responsible Data Science". In: *2019 IEEE International Conference on Big Data (Big Data)*. IEEE. 1–1.

Gil, Y., J. Honaker, S. Gupta, Y. Ma, V. DÓrazio, D. Garijo, S. Gadewar, Q. Yang, and N. Jahanshad. (2019). "Towards human-guided machine learning". In: *Proceedings of the 24th International Conference on Intelligent User Interfaces*. 614–624.

Golovin, D., B. Solnik, S. Moitra, G. Kochanski, J. Karro, and D. Sculley. (2017). "Google vizier: A service for black-box optimization". In: *Proceedings of the 23rd ACM SIGKDD international conference on knowledge discovery and data mining*. 1487–1495.

Gorecky, D., M. Schmitt, M. Loskyll, and D. Zuhlke. (2014). "Human-machine-interaction in the industry 4.0 era". In: *2014 12th IEEE international conference on industrial informatics (INDIN)*. IEEE. 289–294.

Goyal, A., A. Lamb, J. Hoffmann, S. Sodhani, S. Levine, Y. Bengio, and B. Scholkopf. (2021). "Recurrent independent mechanisms". In: *Proceedings of the 9th International Conference on Learning Representations (ICLR)*.

Greenblatt, R. D., D. E. Eastlake III, and S. D. Crocker. (1967). "The Greenblatt chess program". In: *Proceedings of the November 14-16, 1967, fall joint computer conference*. 801–810.

Greene, J., F. Rossi, J. Tasioulas, K. B. Venable, and B. Williams. (2016). "Embedding ethical principles in collective decision support systems". In: *Proceedings of the Thirtieth Conference on Artificial Intelligence (AAAI)*. 4147–4151.

Gronchi, G. and F. Giovannelli. (2018). "Dual Process Theory of Thought and Default Mode Network: A Possible Neural Foundation of Fast Thinking". *Frontiers in Psychology*. 9: 1237.

Guidotti, R., A. Monreale, S. Ruggieri, F. Turini, F. Giannotti, and D. Pedreschi. (2018). "A survey of methods for explaining black box models". *ACM computing surveys (CSUR)*. 51(5): 1–42.

Guillaume, S. (2001). "Designing fuzzy inference systems from data: An interpretability-oriented review". *IEEE Transactions on fuzzy systems*. 9(3): 426–443.

Guo, X., S. Singh, H. Lee, R. L. Lewis, and X. Wang. (2014). "Deep learning for real-time Atari game play using offline Monte-Carlo tree search planning". In: *Advances in neural information processing systems*. 3338–3346.

Gustavsson, P., A. Syberfeldt, R. Brewster, and L. Wang. (2017). "Human-robot collaboration demonstrator combining speech recognition and haptic control". *Procedia CIRP*. 63: 396–401.

Hainc, N., C. Federau, B. Stieltjes, M. Blatow, A. Bink, and C. Stippich. (2017). "The Bright, Artificial Intelligence-Augmented Future of Neuroimaging Reading". *Frontiers in Neurology*. 8: 489.

Hajian, S., F. Bonchi, and C. Castillo. (2016). "Algorithmic bias: From discrimination discovery to fairness-aware data mining". In: *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*. 2125–2126.

Harari, Y. N. (2014). *Sapiens: A brief history of humankind*. Random House.

Hardt, M., X. Chen, X. Cheng, M. Donini, J. Gelman, S. Gollaprolu, J. He, P. Larroy, X. Liu, N. McCarthy, *et al.* (2021). "Amazon SageMaker Clarify: Machine Learning Bias Detection and Explainability in the Cloud". In: *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. 2974–2983.

Harel, D., A. Marron, and J. Sifakis. (2020). "Autonomics: In search of a foundation for next-generation autonomous systems". *Proceedings of the National Academy of Sciences*. 117(30): 17491–17498.

Harris, C. (2020). "Mitigating Cognitive Biases in Machine Learning Algorithms for Decision Making". In: *Proceedings of the Web Conference*. 775–781.

Harvey, H. (2020). "Why AI will not replace radiologists". *Towards Data Science*. URL: https://towardsdatascience.com/why-ai-will-not-replace-radiologists-c7736f2c7d80.

Hawkins, A. J. (2018). "Inside Waymo's Strategy to Grow the Best Brains for Self-Driving Cars". URL: https://www.theverge.com/2018/5/9/17307156/google-waymo-driverless-cars-deep-learning-neural-net-interview.

He, X., K. Zhao, and X. Chu. (2021). "AutoML: A survey of the state-of-the-art". *Knowledge-Based Systems*. 212: 106622.

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

Heyen, F., T. Munz, M. Neumann, D. Ortega, N. T. Vu, D. Weiskopf, and M. Sedlmair. (2020). "ClaVis: An Interactive Visual Comparison System for Classifiers". In: *Proceedings of the International Conference on Advanced Visual Interfaces*. 1–9.

Hofmann, M. and R. Klinkenberg. (2016). *RapidMiner: Data mining use cases and business analytics applications*. CRC Press.

Holzinger, A., A. Carrington, and H. Muller. (2020). "Measuring the quality of explanations: the system causability scale (SCS)". *KI-Kunstliche Intelligenz*: 1–6.

Holzinger, A., G. Langs, H. Denk, K. Zatloukal, and H. Muller. (2019). "Causability and explainability of artificial intelligence in medicine". *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*. 9(4): e1312.

Hook, K. (2000). "Steps to take before intelligent user interfaces become real". *Interacting with computers*. 12(4): 409–426.

Horvitz, E. (1999). "Principles of mixed-initiative user interfaces". In: *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*. 159–166.

Hu, Z., X. Xin, W. Xu, Y. Sun, Z. Jiang, X. Wang, Y. Liu, S. Lu, and M. Zhao. (2019). "A Literature Review of the Research on Interaction Mode of Self-driving Cars". In: *International Conference on Human-Computer Interaction*. Springer. 29–40.

Hunt, E. (2016). "Tay, Microsoft's AI chatbot, gets a crash course in racism from Twitter". *The Guardian*. 24(3).

Hunter, J. D. (2007). "Matplotlib: A 2D graphics environment". *Computing in science & engineering*. 9(03): 90–95.

Iyengar, V. (2019). "New Innovations in Driverless AI". URL: https://www.h2o.ai/blog/new-innovations-in-driverless-ai/.

Jacovi, A., A. Marasovic, T. Miller, and Y. Goldberg. (2021). "Formalizing Trust in Artificial Intelligence: Prerequisites, Causes and Goals of Human Trust in AI". In: *Proceedings of the Fourth Annual ACM FAccT Conference*.

Jameson, A. (2007). "Adaptive interfaces and agents". In: *The human-computer interaction handbook*. CRC Press. 459–484.

Jordan, M. I. (2019). "Artificial intelligence—the revolution hasn't happened yet". *Harvard Data Science Review*. 1(1).

Jumper, J., R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Zidek, A. Potapenko, *et al.* (2021). "Highly accurate protein structure prediction with AlphaFold". *Nature*. 596(7873): 583–589.

Just, J. and S. Ghosal. (2019). "Deep Generative Models Strike Back! Improving Understanding and Evaluation in Light of Unmet Expectations for OoD Data". *arXiv preprint arXiv:1911.04699*.

Kadlec, P. and B. Gabrys. (2009a). "Architecture for development of adaptive on-line prediction models". *Memetic Computing*. 1(4): 241–269.

Kadlec, P. and B. Gabrys. (2009b). "Evolving on-line prediction model dealing with industrial data sets". In: *2009 IEEE Workshop on Evolving and Self-Developing Intelligent Systems*. IEEE. DOI: 10.1109/esdis.2009.4938995.

Kadlec, P. and B. Gabrys. (2009c). "Soft Sensor Based on Adaptive Local Learning". In: *Advances in Neuro-Information Processing*. Springer Berlin Heidelberg. 1172–1179. DOI: 10.1007/978-3-642-02490-0_142.

Kadlec, P. and B. Gabrys. (2009d). "Soft sensors: where are we and what are the current and future challenges?" *IFAC Proceedings Volumes*. 42(19): 572–577. DOI: 10.3182/20090921-3-tr-3005.00098.

Kadlec, P. and B. Gabrys. (2010). "Adaptive on-line prediction soft sensing without historical data". In: *Proceedings of the 2010 international joint conference on neural networks (IJCNN)*. IEEE. 1–8.

Kadlec, P. and B. Gabrys. (2011). "Local learning-based adaptive soft sensor for catalyst activation prediction". *AIChE Journal*. 57(5): 1288–1301.

Kahneman, D. (2011). *Thinking, fast and slow*. Macmillan.

Kamiran, F., S. Mansha, A. Karim, and X. Zhang. (2018). "Exploiting reject option in classification for social discrimination control". *Information Sciences*. 425: 18–33.

Karimi, F., M. Genois, C. Wagner, P. Singer, and M. Strohmaier. (2018). "Homophily influences ranking of minorities in social networks". *Scientific reports*. 8(1): 1–12.

Kedziora, D. J., K. Musial, and B. Gabrys. (2020). "AutonoML: Towards an Integrated Framework for Autonomous Machine Learning". *arXiv preprint arXiv:2012.12600*.

Kemp, C., A. Perfors, and J. B. Tenenbaum. (2007). "Learning overhypotheses with hierarchical Bayesian models". *Developmental science*. 10(3): 307–321.

Khuat, T. T., F. Chen, and B. Gabrys. (2021a). "An effective multiresolution hierarchical granular representation based classifier using general fuzzy min-max neural network". *IEEE Transactions on Fuzzy Systems*. 29(2): 427–441.

Khuat, T. T., D. Ruta, and B. Gabrys. (2021b). "Hyperbox-based machine learning algorithms: a comprehensive survey". *Soft Computing*. 25(2): 1325–1363.

Kingston, J. K. (2016). "Artificial intelligence and legal liability". In: *Proceedings of the International conference on innovative techniques and applications of artificial intelligence*. Springer. 269–279.

Kitchin, R. (2017). "Thinking critically about and researching algorithms". *Information, Communication & Society*. 20(1): 14–29.

Kleinberg, J., S. Mullainathan, and M. Raghavan. (2017). "Inherent Trade-Offs in the Fair Determination of Risk Scores". In: *Proceedings of the 8th Innovations in Theoretical Computer Science Conference*. Vol. 67. 43:1–43:23.

Kliegr, T., S. Bahnik, and J. Furnkranz. (2021). "A review of possible effects of cognitive biases on interpretation of rule-based machine learning models". *Artificial Intelligence*. 295: 103458.

Klien, G., D. D. Woods, J. M. Bradshaw, R. R. Hoffman, and P. J. Feltovich. (2004). "Ten challenges for making automation a "team player" in joint human-agent activity". *IEEE Intelligent Systems*. 19(6): 91–95.

Kokar, M. M. and M. R. Endsley. (2012). "Situation awareness and cognitive modeling". *IEEE Intelligent Systems*. 27(3): 91–96.

Korteling, J., G. Van de Boer-Visschedijk, R. Blankendaal, R. Boonekamp, and A. Eikelboom. (2021). "Human-versus Artificial Intelligence". *Frontiers in Artificial Intelligence*. 4.

Kraska, T. (2018). "Northstar: an interactive data science system". *Proceedings of the VLDB Endowment*. 11(12): 2150–2164.

Kulesza, T., M. Burnett, W.-K. Wong, and S. Stumpf. (2015). "Principles of explanatory debugging to personalize interactive machine learning". In: *Proceedings of the 20th international conference on intelligent user interfaces*. 126–137.

Kusner, M. J., J. Loftus, C. Russell, and R. Silva. (2017). "Counterfactual Fairness". In: *Advances in Neural Information Processing Systems*. Vol. 30.

Laird, J. and S. Mohan. (2018). "Learning fast and slow: Levels of learning in general autonomous intelligent agents". In: *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 32. 7983–7987.

Lake, B. M., R. Salakhutdinov, and J. B. Tenenbaum. (2015). "Human-level concept learning through probabilistic program induction". *Science*. 350(6266): 1332–1338.

Lake, B. M., T. D. Ullman, J. B. Tenenbaum, and S. J. Gershman. (2017a). "Building machines that learn and think like people". *Behavioral and brain sciences*. 40.

Lake, B. M., T. D. Ullman, J. B. Tenenbaum, and S. J. Gershman. (2017b). "Ingredients of intelligence: From classic debates to an engineering roadmap". *Behavioral and Brain Sciences*. 40.

Lane, H. C. (2012). "Cognitive Models of Learning". In: *Encyclopedia of the Sciences of Learning*. Springer. 608–610.

Lechner, M., R. Hasani, A. Amini, T. A. Henzinger, D. Rus, and R. Grosu. (2020). "Neural circuit policies enabling auditable autonomy". *Nature Machine Intelligence*. 2(10): 642–652.

Lee, D. J.-L., S. Macke, D. Xin, A. Lee, S. Huang, and A. Parameswaran. (2019a). "A Human-in-the-loop Perspective on AutoML: Milestones and the Road Ahead". *IEEE Data Engineering Bulletin*. 42(2): 59–70.

Lee, J., C. Lee, and G. J. Kim. (2017). "Vouch: multimodal touch-and-voice input for smart watches under difficult operating conditions". *Journal on Multimodal User Interfaces*. 11(3): 289–299.

Lee, M. K. (2018). "Understanding perception of algorithmic decisions: Fairness, trust, and emotion in response to algorithmic management". *Big Data & Society*. 5(1): 1–16.

Lee, M. K., D. Kusbit, A. Kahng, J. T. Kim, X. Yuan, A. Chan, D. See, R. Noothigattu, S. Lee, A. Psomas, and A. D. Procaccia. (2019b). "WeBuildAI: Participatory Framework for Algorithmic Governance". *Proceedings of the ACM on Human-Computer Interaction*. 3(CSCW): 1–35.

Lemke, C., M. Budka, and B. Gabrys. (2015). "Metalearning: a survey of trends and technologies". *Artificial Intelligence Review*. 44(1): 117–130. DOI: 10.1007/s10462-013-9406-y.

Lemke, C. and B. Gabrys. (2010). "Meta-learning for time series forecasting and forecast combination". *Neurocomputing*. 73(10-12): 2006–2016. DOI: 10.1016/j.neucom.2009.09.020.

Li, J., K. Cheng, S. Wang, F. Morstatter, R. P. Trevino, J. Tang, and H. Liu. (2017). "Feature Selection: A Data Perspective". *ACM Computing Surveys*. 50(6): 1–45.

Lin, C.-T. and T.-T. N. Do. (2021). "Direct-sense brain–computer interfaces and wearable computers". *IEEE Transactions on Systems, Man, and Cybernetics: Systems*. 51(1): 298–312.

Linardatos, P., V. Papastefanopoulos, and S. Kotsiantis. (2021). "Explainable AI: A review of machine learning interpretability methods". *Entropy*. 23(1): 18.

Lippmann, R., W. Campbell, and J. Campbell. (2016). "An Overview of the DARPA Data Driven Discovery of Models (D3M) Program". In: *NIPS Workshop on Artificial Intelligence for Data Science*.

Lipton, P. (1990). "Contrastive explanation". *Royal Institute of Philosophy Supplements*. 27: 247–266.

Littman, M. L., I. Ajunwa, G. Berger, C. Boutilier, M. Currie, F. Doshi-Velez, G. Hadfield, M. C. Horowitz, C. Isbell, H. Kitano, K. Levy, T. Lyons, M. Mitchell, J. Shah, S. Sloman, S. Vallor, and T. Walsh. (2021). "Gathering Strength, Gathering Storms: The One Hundred Year Study on Artificial Intelligence (AI100) 2021 Study Panel Report". *Tech. rep*. Stanford, CA: Stanford University. 1–82. URL: http://ai100.stanford.edu/2021-report.

Liu, S., X. Wang, M. Liu, and J. Zhu. (2017). "Towards better analysis of machine learning models: A visual analytics perspective". *Visual Informatics*. 1(1): 48–56.

Liu, Y., W. Tian, and S. Li. (2021). "Meta-data Augmentation Based Search Strategy Through Generative Adversarial Network for AutoML Model Selection". In: *Proceedings of the Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD)*. Springer. 312–324.

Longo, L. (2016). "Argumentation for knowledge representation, conflict resolution, defeasible inference and its integration with machine learning". In: *Machine Learning for Health Informatics*. Springer. 183–208.

Lu, Y., J. S. Adrados, S. S. Chand, and L. Wang. (2021). "Humans Are Not Machines—Anthropocentric Human–Machine Symbiosis for Ultra-Flexible Smart Manufacturing". *Engineering*. 7(6): 734–737.

Mahmud, S., X. Lin, and J.-H. Kim. (2020). "Interface for Human Machine Interaction for assistant devices: a review". In: *Proccedings of the 10th Annual Computing and Communication Workshop and Conference*. IEEE. 768–773.

Mao, Y., D. Wang, M. Muller, K. R. Varshney, I. Baldini, C. Dugan, and A. Mojsilovic. (2019). "How data scientists work together with domain experts in scientific collaborations: To find the right answer or to ask the right question?" *Proceedings of the ACM on Human-Computer Interaction*. 3: 1–23.

Marcus, G. (2020). "The next decade in AI: four steps towards robust artificial intelligence". *arXiv preprint arXiv:2002.06177*.

Marcus, G. and E. Davis. (2019). *Rebooting AI: Building artificial intelligence we can trust*. Vintage.

Marcus, G. and E. Davis. (2021). "Has AI found a new Foundation?" *The Gradient*. URL: https://thegradient.pub/has-ai-found-a-new-foundation.

Marcus, G. F. (2019). *The algebraic mind: Integrating connectionism and cognitive science*. MIT press.

Margetis, G., S. Ntoa, M. Antona, and C. Stephanidis. (2021). "Human-Centered Design of Artificial Intelligence". In: *Handbook of Human Factors and Ergonomics*. Wiley Online Library. 1085–1106.

McCarthy, J. (2007). "From here to human-level AI". *Artificial Intelligence*. 171(18): 1174–1182.

McGough, M. (2018). "How bad is sacramento's air, exactly? Google results appear at odds with reality, some say". *Sacramento Bee*. URL: https://www.sacbee.com/news/state/california/fires/article216227775.html.

Mehrabi, N., F. Morstatter, N. Saxena, K. Lerman, and A. Galstyan. (2021). "A Survey on Bias and Fairness in Machine Learning". *ACM Computing Surveys*. 54(6): 1–35.

Mendoza, M., M. Mendoza, E. Mendoza, and E. Gonzalez. (2015). "Augmented reality as a tool of training for data collection on torque auditing". *Procedia Computer Science*. 75: 5–11.

Mertens, M. and H. J. Damveld. (2012). "An avionics touch screen-based control display concept". In: *Head-and Helmet-Mounted Displays XVII; and Display Technologies and Applications for Defense, Security, and Avionics VI*. Vol. 8383. International Society for Optics and Photonics. 83830L.

Mesko, B. (2019). "The real era of the art of medicine begins with artificial intelligence". *Journal of medical Internet research*. 21(11): e16295.

Mikolov, T., A. Joulin, and M. Baroni. (2016). "A roadmap towards machine intelligence". In: *International Conference on Intelligent Text Processing and Computational Linguistics*. Springer. 29–61.

Miller, J. D., R. Yampolskiy, O. Haggstrom, and S. Armstrong. (2020). "Chess as a Testing Grounds for the Oracle Approach to AI Safety". *arXiv preprint arXiv:2010.02911*.

Miller, T. (2019). "Explanation in artificial intelligence: Insights from the social sciences". *Artificial intelligence*. 267: 1–38.

Molnar, C. (2020). *Interpretable machine learning*. Lulu. com.

Moore, M. (2018). "What is Industry 4.0? Everything you need to know". *World of Tech*.

Muelling, K., A. Venkatraman, J.-S. Valois, J. E. Downey, J. Weiss, S. Javdani, M. Hebert, A. B. Schwartz, J. L. Collinger, and J. A. Bagnell. (2017). "Autonomy infused teleoperation with application to brain computer interface controlled manipulation". *Autonomous Robots*. 41(6): 1401–1422.

Müller, J. (2020). "Enabling technologies for industry 5.0, results of a workshop with europe's technology leaders". *Directorate-General for Research and Innovation*.

Murray, J. S. (2018). "Multiple imputation: a review of practical and theoretical findings". *Statistical Science*. 33(2): 142–159.

Nardi, L., D. Koeplinger, and K. Olukotun. (2019). "Practical Design Space Exploration". In: *Proceedings of the 27th International Symposium on Modeling, Analysis, and Simulation of Computer and Telecommunication Systems (MASCOTS)*. 347–358. DOI: 10.1109/MASCOTS.2019.00045.

Newell, A. (1980). "Physical symbol systems". *Cognitive science*. 4(2): 135–183.

Newell, A. and H. A. Simon. (1972). *Human problem solving*. Vol. 104. No. 9. Prentice-hall Englewood Cliffs, NJ.

Nguyen, T.-D., D. J. Kedziora, K. Musial, and B. Gabrys. (2021a). "Exploring Opportunistic Meta-knowledge to Reduce Search Spaces for Automated Machine Learning". In: *Proceedings of the International Joint Conference on Neural Networks (IJCNN)*. 1–8. DOI: 10.1109/IJCNN52387.2021.9533431.

Nguyen, T.-D., K. Musial, and B. Gabrys. (2021b). "AutoWeka4MCPS-AVATAR: Accelerating automated machine learning pipeline composition and optimisation". *Expert Systems with Applications*. 185: 115643. DOI: https://doi.org/10.1016/j.eswa.2021.115643.

Nicas, J., N. Kitroeff, D. Gelles, and J. Glanz. (2019). "Boeing built deadly assumptions into 737 Max, blind to a late design change". *The New York Times*. URL: https://www.nytimes.com/2019/06/01/business/boeing-737-max-crash.html.

North, C. (2006). "Toward measuring visualization insight". *IEEE computer graphics and applications*. 26(3): 6–9.

Ntoutsi, E., P. Fafalios, U. Gadiraju, V. Iosifidis, W. Nejdl, M.-E. Vidal, S. Ruggieri, F. Turini, S. Papadopoulos, E. Krasanakis, I. Kompatsiaris, K. Kinder-Kurlanda, C. Wagner, F. Karimi, M. Fernandez, H. Alani, B. Berendt, T. Kruegel, C. Heinze, K. Broelemann, G. Kasneci, T. Tiropanis, and S. Staab. (2020). "Bias in data-driven artificial intelligence systems—An introductory survey". *WIREs Data Mining and Knowledge Discovery*. 10(3): e1356.

Nunez, H., C. Angulo, and A. Catala. (2002). "Rule extraction from support vector machines". In: *Proceedings of the European Symposium on Artificial Neural Networks*. 107–112.

Obermeyer, Z., B. Powers, C. Vogeli, and S. Mullainathan. (2019). "Dissecting racial bias in an algorithm used to manage the health of populations". *Science*. 366(6464): 447–453.

Ono, J. P., S. Castelo, R. Lopez, E. Bertini, J. Freire, and C. Silva. (2021). "PipelineProfiler: A Visual Analytics Tool for the Exploration of AutoML Pipelines". *IEEE Transactions on Visualization and Computer Graphics*. 27(2): 390–400.

Orseau, L. and S. Armstrong. (2016). "Safely Interruptible Agents": 557–566.

Palade, V., D.-C. Neagu, and R. J. Patton. (2001). "Interpretation of trained neural networks by rule extraction". In: *International Conference on Computational Intelligence*. Springer. 152–161.

Pan, S. J. and Q. Yang. (2009). "A survey on transfer learning". *IEEE Transactions on knowledge and data engineering*. 22(10): 1345–1359.

Parasuraman, R., T. B. Sheridan, and C. D. Wickens. (2000). "A model for types and levels of human interaction with automation". *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*. 30(3): 286–297.

Parisi, G. I., R. Kemker, J. L. Part, C. Kanan, and S. Wermter. (2019). "Continual lifelong learning with neural networks: A review". *Neural Networks*. 113: 54–71.

Park, H., J. Kim, M. Kim, J.-H. Kim, J. Choo, J.-W. Ha, and N. Sung. (2019). "VisualHyperTuner: Visual analytics for user-driven hyperparameter tuning of deep neural networks". In: *Proceedings of the 2nd SysML Conference*.

Pearl, J. (2019). "The seven tools of causal inference, with reflections on machine learning". *Communications of the ACM*. 62(3): 54–60.

Pearl, J. and D. Mackenzie. (2018). *The book of why: the new science of cause and effect*. Basic books.

Peled, I. and O. Zaslavsky. (1997). "Counter-Examples That (Only) Prove and Counter-Examples That (Also) Explain". *FOCUS on Learning Problems in mathematics*. 19(3): 49–61.

Perer, A. and B. Shneiderman. (2008). "Integrating statistics and visualization: case studies of gaining clarity during exploratory data analysis". In: *Proceedings of the SIGCHI conference on Human Factors in computing systems*. 265–274.

Polanyi, M. (2012). *Personal knowledge*. Routledge.

Powers, B. (2018). "Once more with feeling: Teaching empathy to machines". *The Wall Street Journal*. URL: https://www.wsj.com/articles/once-more-with-feeling-teaching-empathy-to-machines-11544713141.

Preisz, A. (2019). "Fast and slow thinking; and the problem of conflating clinical reasoning and ethical deliberation in acute decision-making". *Journal of paediatrics and child health*. 55(6): 621–624.

Pretz, K. (2021). "Stop Calling Everything AI, Machine-Learning Pioneer Says". URL: https://spectrum.ieee.org/stop-calling-everything-ai-machinelearning-pioneer-says.

Proctor, C. and W.-K. Ahn. (2007). "The effect of causal knowledge on judgments of the likelihood of unknown features". *Psychonomic Bulletin & Review*. 14(4): 635–639.

Quintero, C. P., R. T. Fomena, A. Shademan, O. Ramirez, and M. Jagersand. (2014). "Interactive teleoperation interface for semi-autonomous control of robot arms". In: *Proceedings of the Canadian Conference on Computer and Robot Vision*. IEEE. 357–363.

Rader, E., K. Cotter, and J. Cho. (2018). "Explanations as mechanisms for supporting algorithmic transparency". In: *Proceedings of the 2018 CHI conference on human factors in computing systems*. 1–13.

Raheja, J. L., R. Shyam, U. Kumar, and P. B. Prasad. (2010). "Real-time robotic hand control using hand gestures". In: *Proceedings of the Second International Conference on Machine Learning and Computing*. IEEE. 12–16.

Rahwan, I. (2018). "Society-in-the-loop: programming the algorithmic social contract". *Ethics and Information Technology*. 20(1): 5–14.

Rawal, K. and H. Lakkaraju. (2020). "Beyond Individualized Recourse: Interpretable and Interactive Summaries of Actionable Recourses". In: *Proceedings of the 34th Conference on Neural Information Processing Systems*.

Rehder, B. (2003). "A causal-model theory of conceptual representation and categorization". *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 29(6): 1141–1159.

Rhodes, M. (2016). "So. algorithms are designing chairs now". *Wired*. URL: https://www.wired.com/2016/10/elbo-chair-autodesk-algorithm/.

Ribeiro, M. T., S. Singh, and C. Guestrin. (2016). ""Why should i trust you?" Explaining the predictions of any classifier". In: *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*. 1135–1144.

Riedl, M. (2016). "Big red button". URL: https://markriedl.github.io/big-red-button/.

Rocha, T., D. Carvalho, M. Bessa, S. Reis, and L. Magalhaes. (2017). "Usability evaluation of navigation tasks by people with intellectual disabilities: a Google and SAPO comparative study regarding different interaction modalities". *Universal Access in the Information Society*. 16(3): 581–592.

Rossi, F. and A. Loreggia. (2019). "Preferences and Ethical Priorities: Thinking Fast and Slow in AI". In: *Proceedings of the 18th International Conference on Autonomous Agents and MultiAgent Systems*. 3–4.

Rossi, F. and N. Mattei. (2019). "Building ethically bounded AI". In: *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 33. No. 1. 9785–9789.

Rouwhorst, W., R. Verhoeven, M. Suijkerbuijk, T. Bos, A. Maij, M. Vermaat, and R. Arents. (2017). "Use of touch screen display applications for aircraft flight control". In: *2017 IEEE/AIAA 36th Digital Avionics Systems Conference (DASC)*. IEEE. 1–10.

Rudin, C. (2019). "Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead". *Nature Machine Intelligence*. 1(5): 206–215. DOI: 10.1038/s42256-019-0048-x.

Ruhl, C. (2021). "Cognitive bias examples". *Simply Psychology*. URL: www.simplypsychology.org/cognitive-bias.html.

Ruta, D. and B. Gabrys. (2005). "Classifier selection for majority voting". *Information Fusion*. 6(1): 63–81. DOI: 10.1016/j.inffus.2004.04.008.

Sager, C., C. Janiesch, and P. Zschech. (2021). "A survey of image labelling for computer vision applications". *Journal of Business Analytics*: 1–20.

Salvador, M. M., M. Budka, and B. Gabrys. (2016a). "Adapting multicomponent predictive systems using hybrid adaptation strategies with auto-weka in process industry". In: *Workshop on Automatic Machine Learning*. PMLR. 48–57.

Salvador, M. M., M. Budka, and B. Gabrys. (2016b). "Towards automatic composition of multicomponent predictive systems". In: *International conference on hybrid artificial intelligence systems*. Springer. 27–39.

Salvador, M. M., M. Budka, and B. Gabrys. (2017). "Modelling Multi-Component Predictive Systems as Petri Nets". In: *Proceedings of the 15th Annual Industrial Simulation Conference*. 17–23.

Salvador, M. M., M. Budka, and B. Gabrys. (2019). "Automatic composition and optimization of multicomponent predictive systems with an extended auto-WEKA". *IEEE Transactions on Automation Science and Engineering*. 16(2): 946–959.

Samek, W., G. Montavon, A. Vedaldi, L. K. Hansen, and K.-R. Muller. (2019). *Explainable AI: interpreting, explaining and visualizing deep learning*. Vol. 11700. Springer Nature.

Saposnik, G., D. Redelmeier, C. C. Ruff, and P. N. Tobler. (2016). "Cognitive biases associated with medical decisions: a systematic review". *BMC medical informatics and decision making*. 16(1): 1–14.

Saxena, N. A., K. Huang, E. DeFilippis, G. Radanovic, D. C. Parkes, and Y. Liu. (2019). "How Do Fairness Definitions Fare? Examining Public Attitudes Towards Algorithmic Definitions of Fairness". In: *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society*. 99–106.

Schelter, S., F. Biessmann, D. Lange, T. Rukat, P. Schmidt, S. Seufert, P. Brunelle, and A. Taptunov. (2019). "Unit testing data with deequ". In: *Proceedings of the 2019 International Conference on Management of Data*. 1993–1996.

Schmidt, P. and F. Biessmann. (2019). "Quantifying interpretability and trust in machine learning systems". *arXiv preprint arXiv:1901.08558*.

Schrittwieser, J., I. Antonoglou, T. Hubert, K. Simonyan, L. Sifre, S. Schmitt, A. Guez, E. Lockhart, D. Hassabis, T. Graepel, T. Lillicrap, and D. Silver. (2020). "Mastering Atari, Go, chess and shogi by planning with a learned model". *Nature*. 588(7839): 604–609.

Scriven, A., D. J. Kedziora, K. Musial, and B. Gabrys. (2022). "The Technological Emergence of AutoML: A Survey of Performant Software and Applications in the Context of Industry". *arXiv preprint arXiv:2211.04148*.

Serban, F., J. Vanschoren, J.-U. Kietz, and A. Bernstein. (2013). "A survey of intelligent assistants for data analysis". *ACM Computing Surveys*. 45(3): 1–35.

Sevastjanova, R., F. Beck, B. Ell, C. Turkay, R. Henkin, M. Butt, D. A. Keim, and M. El-Assady. (2018). "Going beyond visualization: Verbalization as complementary medium to explain machine learning models". In: *Workshop on Visualization for AI Explainability at IEEE VIS*.

Shang, Z., E. Zgraggen, B. Buratti, F. Kossmann, P. Eichmann, Y. Chung, C. Binnig, E. Upfal, and T. Kraska. (2019). "Democratizing data science through interactive curation of ML pipelines". In: *Proceedings of the 2019 International Conference on Management of Data*. 1171–1188.

Sheridan, T. B. and W. L. Verplank. (1978). "Human and computer control of undersea teleoperators". *Tech. rep.* Massachusetts Inst of Tech Cambridge Man-Machine Systems Lab.

Shin, D. (2020). "How do users interact with algorithm recommender systems? The interaction of users, algorithms, and performance". *Computers in Human Behavior*. 109: 106344.

Shin, D. (2021). "The effects of explainability and causability on perception, trust, and acceptance: Implications for explainable AI". *International Journal of Human-Computer Studies*. 146: 102551.

Shneiderman, B. (2020a). "Design lessons from AI's two grand goals: Human emulation and useful applications". *IEEE Transactions on Technology and Society*. 1(2): 73–82.

Shneiderman, B. (2020b). "Human-centered artificial intelligence: Reliable, safe & trustworthy". *International Journal of Human–Computer Interaction*. 36(6): 495–504.

Silver, D., A. Huang, C. J. Maddison, A. Guez, L. Sifre, G. Van Den Driessche, J. Schrittwieser, I. Antonoglou, V. Panneershelvam, M. Lanctot, *et al.* (2016). "Mastering the game of Go with deep neural networks and tree search". *Nature*. 529(7587): 484–489.

Silver, D., T. Hubert, J. Schrittwieser, I. Antonoglou, M. Lai, A. Guez, M. Lanctot, L. Sifre, D. Kumaran, T. Graepel, *et al.* (2018). "A general reinforcement learning algorithm that masters chess, shogi, and Go through self-play". *Science*. 362(6419): 1140–1144.

Silver, D., J. Schrittwieser, K. Simonyan, I. Antonoglou, A. Huang, A. Guez, T. Hubert, L. Baker, M. Lai, A. Bolton, *et al.* (2017). "Mastering the game of go without human knowledge". *Nature*. 550(7676): 354–359.

Simmler, M. and R. Frischknecht. (2021). "A taxonomy of human-machine collaboration: capturing automation and technical autonomy". *AI & SOCIETY*. 36(1): 239–250.

Sipsas, K., K. Alexopoulos, V. Xanthakis, and G. Chryssolouris. (2016). "Collaborative maintenance in flow-line manufacturing environments: An Industry 4.0 approach". *Procedia Cirp*. 55: 236–241.

Sloman, S. (2014). "Two systems of reasoning: An update". In: *Dual-process theories of the social mind*. Ed. by J. W. Sherman, B. Gawronski, and Y. Trope. The Guilford Press. Chap. 5. 69–79.

Solaki, A., F. Berto, and S. Smets. (2019). "The logic of fast and slow thinking". *Erkenntnis*: 1–30.

Spinner, T., U. Schlegel, H. Schäfer, and M. El-Assady. (2020). "explAIner: A visual analytics framework for interactive and explainable machine learning". *IEEE transactions on visualization and computer graphics*. 26(1): 1064–1074.

Stahl, F., B. Gabrys, M. M. Gaber, and M. Berendsen. (2013). "An overview of interactive visual data mining techniques for knowledge discovery". *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*. 3(4): 239–256.

Stephenson, S. (2020). "Deepgram Pioneers Novel Training Approach Setting New Standard for AI Companies". URL: https://deepgram.com/blog/deepgram-pioneers-novel-training-approach-setting-new-standard-for-ai-companies-2.

Strauch, B. (2017). "Ironies of automation: Still unresolved after all these years". *IEEE Transactions on Human-Machine Systems*. 48(5): 419–433.

Strickland, E. (2021). "The Turbulent Past and Uncertain Future of Artificial Intelligence". *IEEE Spectrum*. URL: https://spectrum.ieee.org/history-of-ai.

Strobelt, H., S. Gehrmann, H. Pfister, and A. M. Rush. (2017). "Lstmvis: A tool for visual analysis of hidden state dynamics in recurrent neural networks". *IEEE transactions on visualization and computer graphics*. 24(1): 667–676.

Studer, S., T. B. Bui, C. Drescher, A. Hanuschkin, L. Winkler, S. Peters, and K.-R. Müller. (2021). "Towards CRISP-ML (Q): a machine learning process model with quality assurance methodology". *Machine Learning and Knowledge Extraction*. 3(2): 392–413.

Sun, S. (2013). "A survey of multi-view machine learning". *Neural computing and applications*. 23(7): 2031–2038.

Sun, S., H. Shi, and Y. Wu. (2015). "A survey of multi-source domain adaptation". *Information Fusion*. 24: 84–92.

Sutton, R. (2019). "The bitter lesson". URL: http://www.incompleteideas.net/IncIdeas/BitterLesson.html.

Swearingen, T., W. Drevo, B. Cyphers, A. Cuesta-Infante, A. Ross, and K. Veeramachaneni. (2017). "ATM: A distributed, collaborative, scalable system for automated machine learning". In: *Proceedings of the IEEE International Conference on Big Data (Big Data)*. IEEE. 151–162.

Tam, G. K., V. Kothari, and M. Chen. (2016). "An analysis of machine- and human-analytics in classification". *IEEE transactions on visualization and computer graphics*. 23(1): 71–80.

Taniguchi, T., E. Ugur, M. Hoffmann, L. Jamone, T. Nagai, B. Rosman, T. Matsuka, N. Iwahashi, E. Oztop, J. Piater, *et al.* (2018). "Symbol emergence in cognitive developmental systems: a survey". *IEEE transactions on Cognitive and Developmental Systems*. 11(4): 494–516.

Tatic, D. and B. Tesic. (2017). "The application of augmented reality technologies for the improvement of occupational safety in an industrial environment". *Computers in Industry*. 85: 1–10.

Taylor, R. M. and J. Reising. (1995). "The Human-Electronic Crew: Can We Trust The Team?" In:

Tenenbaum, J. B., C. Kemp, T. L. Griffiths, and N. D. Goodman. (2011). "How to grow a mind: Statistics, structure, and abstraction". *science*. 331(6022): 1279–1285.

Thompson, N. C., K. Greenewald, K. Lee, and G. F. Manso. (2021). "Deep Learning's Diminishing Returns". *IEEE Spectrum*. URL: https://spectrum.ieee.org/deep-learning-computational-cost.

Thornton, C., F. Hutter, H. H. Hoos, and K. Leyton-Brown. (2013). "Auto-WEKA: Combined selection and hyperparameter optimization of classification algorithms". In: *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 847–855.

Tomlinson, R. (2020). "Introducing MLOps Champion/Challenger Models". URL: https://www.datarobot.com/blog/introducing-mlops-champion-challenger-models/.

Tran, S. N. and A. S. d. Garcez. (2016). "Deep logic networks: Inserting and extracting knowledge from deep belief networks". *IEEE transactions on neural networks and learning systems*. 29(2): 246–258.

Tsividis, P., J. B. Tenenbaum, and L. Schulz. (2015). "Constraints on hypothesis selection in causal learning". In: *Proceedings of the 37th Annual Cognitive Science Society*. 2434–439.

Tversky, A. and D. Kahneman. (1974). "Judgment under Uncertainty: Heuristics and Biases: Biases in judgments reveal some heuristics of thinking under uncertainty". *Science*. 185(4157): 1124–1131.

Ulicny, B. E., J. J. Moskal, M. M. Kokar, K. Abe, and J. K. Smith. (2014). "Inference and ontologies". In: *Cyber Defense and Situational Awareness*. Springer. 167–199.

Vaidya, S., P. Ambad, and S. Bhosle. (2018). "Industry 4.0–a glimpse". *Procedia Manufacturing*. 20: 233–238.

Valiant, L. G. (2003). "Three Problems in Computer Science". *Journal of the ACM*. 50(1): 96–99.

Vallverdu, J. (2020). "Approximate and Situated Causality in Deep Learning". *Philosophies*. 5(1): 2:1–2:12.

Vamplew, P., R. Dazeley, C. Foale, S. Firmin, and J. Mummery. (2018). "Human-aligned artificial intelligence is a multiobjective problem". *Ethics and Information Technology*. 20(1): 27–40.

Van der Waa, J., E. Nieuwburg, A. Cremers, and M. Neerincx. (2021). "Evaluating XAI: A comparison of rule-based and example-based explanations". *Artificial Intelligence*. 291: 103404.

VanLehn, K. (1999). "Rule-learning events in the acquisition of a complex skill: An evaluation of CASCADE". *The Journal of the Learning Sciences*. 8(1): 71–125.

Vanschoren, J. (2011). "Meta-Learning Architectures: Collecting, Organizing and Exploiting Meta-Knowledge". In: *Studies in Computational Intelligence*. Springer Berlin Heidelberg. 117–155.

Vaswani, A., N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin. (2017). "Attention is All you Need". In: *Advances in Neural Information Processing Systems*. Vol. 30.

Vaughan, N., B. Gabrys, and V. N. Dubey. (2016). "An overview of self-adaptive technologies within virtual reality training". *Computer Science Review*. 22: 65–87.

Verma, S. and J. Rubin. (2018). "Fairness definitions explained". In: *IEEE/ACM international workshop on software fairness (fairware)*. IEEE. 1–7.

Vilone, G. and L. Longo. (2020). "Explainable artificial intelligence: a systematic review". *arXiv preprint arXiv:2006.00093*.

Walch, M., L. Jaksche, P. Hock, M. Baumann, and M. Weber. (2017). "Touch screen maneuver approval mechanisms for highly automated vehicles: A first evaluation". In: *Proceedings of the 9th International Conference on Automotive User Interfaces and Interactive Vehicular Applications Adjunct*. 206–211.

Wang, D., E. Churchill, P. Maes, X. Fan, B. Shneiderman, Y. Shi, and Q. Wang. (2020). "From Human-Human Collaboration to Human-AI Collaboration: Designing AI Systems That Can Work Together with People". In: *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–6.

Wang, D., J. D. Weisz, M. Muller, P. Ram, W. Geyer, C. Dugan, Y. Tausczik, H. Samulowitz, and A. Gray. (2019a). "Human-AI Collaboration in Data Science: Exploring Data Scientists' Perceptions of Automated AI". *Proceedings of the ACM on Human-Computer Interaction*. 3: 1–24.

Wang, D., Q. Yang, A. Abdul, and B. Y. Lim. (2019b). "Designing theory-driven user-centric explainable AI". In: *Proceedings of the 2019 CHI conference on human factors in computing systems*. 1–15.

Wang, Q., Y. Ming, Z. Jin, Q. Shen, D. Liu, M. J. Smith, K. Veeramachaneni, and H. Qu. (2019c). "ATMSeer: Increasing Transparency and Controllability in Automated Machine Learning". In: *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–12.

Wanluk, N., S. Visitsattapongse, A. Juhong, and C. Pintavirooj. (2016). "Smart wheelchair based on eye tracking". In: *Proceedings of the 9th Biomedical Engineering International Conference*. IEEE. 1–4.

Waring, J., C. Lindvall, and R. Umeton. (2020). "Automated machine learning: Review of the state-of-the-art and opportunities for healthcare". *Artificial intelligence in medicine*. 104: 101822.

Weidele, D. K. I., J. D. Weisz, E. Oduor, M. Muller, J. Andres, A. Gray, and D. Wang. (2020). "AutoAIViz: opening the blackbox of automated artificial intelligence with conditional parallel coordinates". In: *Proceedings of the 25th International Conference on Intelligent User Interfaces*. 308–312.

Wilson, H. J. and P. R. Daugherty. (2018). "Collaborative intelligence: humans and AI are joining forces". *Harvard Business Review*. 96(4): 114–123.

Wolker, A. and T. E. Powell. (2018). "Algorithms in the newsroom? News readers' perceived credibility and selection of automated journalism". *Journalism*: 1–18.

Wong, C., N. Houlsby, Y. Lu, and A. Gesmundo. (2018). "Transfer Learning with Neural AutoML". In: *Proceedings of the 32nd International Conference on Neural Information Processing Systems*. 8366–8375.

Xin, D., E. Y. Wu, D. J.-L. Lee, N. Salehi, and A. Parameswaran. (2021). "Whither AutoML? Understanding the Role of Automation in Machine Learning Workflows". In: *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–16.

Xing-Yu, W., J. Jing, Y. Zhang, and W. Bei. (2013). "Brain control: human-computer integration control based on brain-computer interface approach". *Acta Automatica Sinica*. 39(3): 208–221.

Yang, Q., A. Steinfeld, C. Rose, and J. Zimmerman. (2020). "Re-examining whether, why, and how human-AI interaction is uniquely difficult to design". In: *Proceedings of the 2020 chi conference on human factors in computing systems*. 1–13.

Yang, Q., J. Suh, N.-C. Chen, and G. Ramos. (2018). "Grounding interactive machine learning tool design in how non-experts actually build models". In: *Proceedings of the 2018 Designing Interactive Systems Conference*. 573–584.

Yao, Q., M. Wang, Y. Chen, W. Dai, Y.-F. Li, W.-W. Tu, Q. Yang, and Y. Yu. (2018). "Taking human out of learning applications: A survey on automated machine learning". *arXiv preprint arXiv:1810.13306*.

Yi, J. S., Y. Ah-Kang, J. Stasko, and J. A. Jacko. (2007). "Toward a deeper understanding of the role of interaction in information visualization". *IEEE transactions on visualization and computer graphics*. 13(6): 1224–1231.

You, K., M. Long, Z. Cao, J. Wang, and M. I. Jordan. (2019). "Universal domain adaptation". In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2720–2729.

Zarsky, T. (2016). "The trouble with algorithmic decisions: An analytic road map to examine efficiency and fairness in automated and opaque decision making". *Science, Technology, & Human Values*. 41(1): 118–132.

Zemcik, T. (2021). "Failure of chatbot Tay was evil, ugliness and uselessness in its nature or do we judge it through cognitive shortcuts and biases?" *AI & SOCIETY*. 36(1): 361–367.

Zhang, C. and A. Eskandarian. (2021). "A Survey and Tutorial of EEG-Based Brain Monitoring for Driver State Analysis". *IEEE/CAA Journal of Automatica Sinica*. 8(7): 1222–1242.

Zhang, H. and T. Yu. (2020). "AlphaZero". In: *Deep Reinforcement Learning*. Springer. 391–415.

Zhang, R., J. Wu, C. Zhang, W. T. Freeman, and J. B. Tenenbaum. (2016). "A Comparative Evaluation of Approximate Probabilistic Simulation and Deep Neural Networks as Accounts of Human Physical Scene Understanding". In: *Proceedings of the 38th Annual Meeting of the Cognitive Science Society, Recogbizing and Representing Events*. 1781–1786.

Zheng, X., Y. Zhang, S. Hong, H. Li, L. Tang, Y. Xiong, J. Zhou, Y. Wang, X. Sun, P. Zhu, *et al.* (2021). "Evolving Fully Automated Machine Learning via Life-Long Knowledge Anchors". *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 43(9): 3091–3107.

Zhu, J., A. Liapis, S. Risi, R. Bidarra, and G. M. Youngblood. (2018). "Explainable AI for designers: A human-centered perspective on mixed-initiative co-creation". In: *2018 IEEE Conference on Computational Intelligence and Games (CIG)*. 1–8.

Zhuang, F., Z. Qi, K. Duan, D. Xi, Y. Zhu, H. Zhu, H. Xiong, and Q. He. (2020). "A comprehensive survey on transfer learning". *Proceedings of the IEEE*. 109(1): 43–76.

Zliobaite, I., A. Bifet, M. Gaber, B. Gabrys, J. Gama, L. Minku, and K. Musial. (2012). "Next challenges for adaptive learning systems". *ACM SIGKDD Explorations Newsletter*. 14(1): 48–55.

Zöller, M.-A. and M. F. Huber. (2021). "Benchmark and survey of automated machine learning frameworks". *Journal of artificial intelligence research*. 70: 409–472.

Zucker, J.-D. (2003). "A grounded theory of abstraction in artificial intelligence". *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*. 358(1435): 1293–1309.