



Human-centric algebraic machine learning

Human-AML Interaction (Revised)

Deliverable D4.2

WP4 - Human AML Interaction



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 952091.

Consortium



ALGEBRAIC AI



Chantalimaud
Foundation



German
Research Center
for Artificial
Intelligence



TECHNISCHE UNIVERSITÄT
KAISERSLAUTERN



Project

Title: ALMA: Human Centric Algebraic Machine Learning
 Acronym: ALMA
 Coordinator: Proyectos y Sistemas de Mantenimiento SL (eProsima)
 Grant agreement: 952091
 Call: H2020-EIC-FETPROACT-2019
 Program: Horizon 2020
 Start: 01 September 2020
 Duration: 48 months
 Website: <https://alma-ai.eu>
 E-mail: alma@eprosima.com
 Consortium: **Proyectos y Sistemas de Mantenimiento SL (eProsima) (EPROS)**, Spain
Champalimaud Foundation (CF), Portugal
ALGEBRAIC AI (ALG), Spain
German Research Center for Artificial Intelligence (DFKI), Germany
Inria (Inria), France
TU Kaiserslautern (TUK), Germany
FIWARE Foundation e.V. (FF), Germany
Technical Research Centre of Finland (VTT), Finland
Carlos III University (UC3M), Spain

Deliverable

Number: **D4.2**
 Title: **Human-AML Interaction (Revised)**
 Month: 18
 Work Package: WP4 - Human AML Interaction
 Work Package leader: Inria
 Deliverable leader: National Institute for Research in Digital Science and Technology (Inria)
 Deliverable type: R: Document, report
 Dissemination level: Public (PU)
 Date of submission: 2023-09-28
 Version: v1.0
 Status: Finished

Version history

Version	Date	Responsible	Author/Reviewer	Comments
v1.0	2023-09-28	Inria	Inria	final

Executive summary

Algebraic Machine Learning (AML) is a novel machine learning technique based on algebraic representations of data. Unlike statistical learning, AML algorithms are robust regarding the statistical properties of the data and are parameter-free. AML systems can retain what they learn, including sociocultural ethical practices for human–artificial intelligence systems. The aim of the EU-funded ALMA project is to study the mathematical properties of AML and to leverage them to develop a new generation of interactive human-centric machine learning systems. These systems are intended to minimize bias in order to facilitate trust while better enabling distributed collaborative learning.

This report corresponds to *Deliverable D4.2 - Human-AML Interaction (Revised)* of the ALMA project. The goal of this deliverable is to give insight into current research on exploring machine learning model capabilities and making them more accessible to end-users. While there are a number of factors involved in educating and empowering end-users of machine learning applications like AML, we focus in this deliverable mainly on interactive and active machine learning and its potential for increasing machine learning literacy.

We further provide insights in how to apply these paradigms to AML and present an outlook of the used principles applied in D4.8 - Working Prototype and Designs (Revised).

Contents

Executive Summary	3
Contents	4
Acronyms	5
1 Introduction	6
2 Theory of Mind and AML	6
3 Integrating Humans into Intelligent Systems	7
3.1 Interactive Machine Learning	8
3.2 Active learning for machine teaching	9
3.3 Machine Teaching	9
3.4 Machine Learning Literacy	10
4 Discussion	11
5 Conclusion	11
References	12

Acronyms

AML	Algebraic Machine Learning.
HCI	Human-Computer Interaction.
IML	Interactive Machine Learning.
MIAL	Mixed-Initiative Active Learning.
MT	Machine Teaching.

1 Introduction

The goal of this deliverable is to give insight into current research on exploring machine learning model capabilities and making them more accessible to end-users. While there are a number of factors involved in educating and empowering end-users of machine learning applications like AML, we focus in this deliverable D4.2 - Human-AML Interaction (Revised) mainly on interactive and active machine learning and its potential for increasing machine learning literacy.

The previous deliverable D4.1 - Human-AML Interaction (Initial) focused mainly on the interaction between human and machines and discussed the underlying concepts such as *Grounding* and *Human-centered Explainability*. We indicated how to expand on this notion by introducing more cognitive theories, such as theory of mind, which potential we briefly outline in section 2. However, since the previous deliverable we shifted our focus more towards making algebraic machine learning (AML) easier to understand, train and apply in a creative context. The interaction paradigm we applied to make AML more accessible is *Interactive Machine Learning* (Interactive Machine Learning (IML)) allowing users to gradually train a model through interaction on a particular task. This enables users to better understand the potentials and limitations of ML approaches while also making them more accessible to a wider audience. We present insights into the active machine learning and machine teaching paradigms, emphasizing the importance of end-user participation in each teaching step of complex machine learning models in order to make them more understandable and applicable to real-world applications. We finalize with the potential of such interactive approach for *ML literacy*, which has received increased attention in recent years due to the proliferation of ML applications in everyday use.

To summarize, we will first introduce the concept of theory of mind and how it relates to human-machine interaction (section 2), before outlining current work on interactive machine learning (section 3.1), active machine learning (section 3.2), machine teaching (section 3.3), and finally ML literacy (section 3.4). We will highlight how each of these concept could be applied to AML and refer further to deliverable D4.8 - Working Prototype and Designs (Revised) as an existing example.

2 Theory of Mind and AML

In the last deliverable we suggested to extend our investigation of Human-AI collaboration towards cognitive concepts like Theory of Mind. Previous work supports that successful cooperation in groups depends on the capacity of every participant to interpret one's own knowledge and understanding as well as the understanding of other collaborators [1].

The study of cognitive development is where the subject of theory of mind first emerged. It is concerned with the capacity to assign mental states to oneself and others, as well as the understanding that these mental states may vary [2]. Understanding mental states—which might be beliefs, intentions, knowledge, desires emotions, or perspectives—lays the groundwork for grounding, which was discussed in the previous deliverable D4.1 - Human-AML Interaction (Initial). Humans can forecast their collaborators' forthcoming actions and deduce the goals and understandings of those collaborators by being able to read their mental states. Although the majority of theory of mind research has concentrated on developmental psychology, a growing body of literature supports the concept's significance for group behavior [3], and group cognition [4], and human-machine collaboration [2]. Human-machine interaction nevertheless is often limited by the level of ability to interpret the 'mind' of machines, on account of their different, sometimes unexpected, behaviour. Due to machines' occasionally surprising behaviours, human-machine interactions are often limited by the lack of understanding the "mind" of machines. Users continue to treat new technological encounters similarly to how they approach other people, attributing them with their own expectations and social structures [5, 2]. Due to systems inability to understand both their own and other people's mental states this is a common source of misunderstanding and failure in human-machine interaction.

In the context of designing collaborative machine, three skills should be considered [6]:

1) *Interpreting one's own mental states*: To enable an intelligent machine to interpret its own mental states, a computational understanding of and access to intentions, desires, beliefs, knowledge, and perspectives are required. At any point during collaboration, a group member's mental state may be influenced by the content of the conversation, the situation, and information about the current goal. In order to decrease the complexity of the problem space through reducing the number of objectives, requirements, and intent involved, the majority of AI applications have been restricted to specific activities. However, this also decreases the machine's capacity to adjust to evolving settings like those observed during a discussion, when it is essential to expand the current belief space. The constraint of single-purpose systems has been addressed in recent collaborative machine learning techniques [7]. To provide cohesive information about a situation, various sources however must be well-integrated with one another [8]. AML allows to share knowledge across users and to integrate different types of human knowledge through embeddings. It further allows to interpret and explain the predicted results by analyzing the trained atoms. The atoms are the closest one can possibly get to the general idea of a machine's own mental state.

2) *Interpreting others' mental states*: Humans are always attempting to assign mental states to other collaborators in order to predict their next behavior [2]. Iterative discussions are made possible by this kind of reasoning. The capacity to follow a common line of reasoning is referred to as incremental conversation, and it serves as the foundation for all argumentation and subsequent debate in human-human interactions. Identifying and predicting human intention [9, 10] and actions [11, 12, 13] is a major focus of machine learning and AI research in the last decade. Other deliverable presented in this project consortium, such as work presented in WP5, provide first insights into behavior predictions with AML as well as highlight the ethical considerations of doing so.

3) *Predicting subsequent behavior*: Prediction of future behavior may be viewed from two perspectives, much like the interpretation of another person's mental state: When interacting with intelligent systems, humans tend to approach them with a certain set of underlying preconceptions, which frequently results in disrupted experiences brought on by unanticipated system behaviour [2, 14]. Researchers are striving to identify the data required for forecasting the behaviour of machines. In the research by Chandrasekaran et al. [2] on how humans perceive AI minds, it was shown that even when given knowledge about the AI's inner mental states, such as certainty and attentiveness, people were unable to anticipate the AI's upcoming action. One driving aspect is explainability, and recent work has shown how explainable AI as well as human-centered explainability can help user to better understand and predict machine behavior, as discussed in more detail in the previous deliverable where we also outline the potential of AML to support explainability overall.

Overall the field of theory of mind in the context of AI is constantly growing and work bridging cognitive science and AI research are becoming more frequent. In this context we would like to highlight Langley et al.'s work on "Theory of mind and preference learning at the interface of cognitive science, neuroscience, and AI: A review" [15], which outline the potential of such cognitive models for developing more powerful AI. Based on the above analysis we can see potential for new approaches like AML to play a role in such developments in the future.

3 Integrating Humans into Intelligent Systems

There are different levels of user engagement when developing AI agents. Traditional "*Human-in-the-loop*" [16] approaches treat users as a component of the system, either as a source of data or a source of error. Rafner et al. [17] distinguish between "*Human-In-The-Loop*" (HITL) systems that plan, execute or evaluate data acquisition tasks and "*Human-On-The-Loop*" (HOTL) systems that require a human to check the results [18]. They contrast these with "*Human-Out-Of-The-Loop*" (HOOTL) systems that do not include human users at all [19]. The latter argues that AI will save users time and effort that

they will be able to expend on more interesting cognitive tasks. However, this detachment of the training leads to a number of wrong assumptions about human needs as well as system capabilities.

A number of work on AML have so far used humans as data sources or to validate the results. Putting AML in a more interactive context revealed a lack of understanding about how AML learns and the potentials it provides for creative professionals. As a result, we believe there is a greater need to involve the user in the AML process in order to gain a better understanding of the algorithm's capabilities and limitations.

We therefore turned to interactive machine learning approaches to involve the user more into the training process as well as allowing them to iterate on the system to adapt its predictions to the desired outcome. In order to outline work in this field relevant to AML further, we will start introducing some background on interactive machine learning (IML) and active learning, before going deeper into current insights of machine teaching and its potential to increase machine learning literacy for non-expert users.

3.1 Interactive Machine Learning

According to Fails et al. *Interactive Machine Learning (IML)* is a framework that enables the end-user to engage with a machine learning algorithm by examining its capabilities and correcting its errors [20]. This opens several research directions addressing human and machine relations during the learning and application process. The ability to personalize IML systems enables experts who lack machine learning competence to integrate their knowledge by modifying the data and correcting model predictions. This allows domain experts such as medical doctors to get involved in the development and training of knowledge systems for health [21, 22, 23]. Due to the high degree of control applied during training and the ability to catch errors before deployment, this not only increases the recommender systems' transparency but also their trustworthiness and applicability for the domain expert.

The Crayons [20] application was one of the earliest IML programs that tried to teach an algorithm how to segment objects in an image. Users can annotate images to further define model partitions once the algorithm offers a segmentation. The technique is iterative since the user alternately examines the model forecast and adjusts the partition using paint stroke, which improves the overall model accuracy. However, in order to create systems that may be used as tools, IML requires workflows centered on contextual human working practices which can be challenging for some ML approaches [24].

Nevertheless there are a number of successful examples of applying IML to different professional domains such as music or design. Using user-generated examples as a starting point, Laguna et al. [25] focused on how musicians might construct and explore gesture-sound mapping made with IML. Instead of computing the mapping evaluation from a validation set, expert-user insights were used instead to provide more control the musician. In our deliverable D4.8 - Working Prototype and Designs (Revised), we take a similar approach, where we allow users to record their own data set to train the AML model.

Earlier studies on the creation of musical instruments suggested interactive machine learning as a tool for iterative design, allowing artists to, for example, map gestures and sensor information to a sound synthesis system [26, 27, 28]. Thereby researchers discovered that non-ML experts often utilize a different strategy than ML engineers would [28] by using quantitative metrics and direct evaluations between the data they provided and the objective of the model. In line with earlier research, we use gesture-based interaction with AML as a use case to allow musicians to explore new sound pattern using IML.

One challenge in IML is the evolution of the users perception or intention during the ML processes resulting in potentially inconsistent labeling [29]. To reduce such inconsistencies, Kulesza et al. proposed a tool to provide structure to the labeling process [30]. Researchers have highlighted the possibility of explanations for enhancing labeling consistency in addition to structuring the process [31]. For example, through the design of an email classification system, they demonstrated how describing the classification

to the user may improve both their understanding of the learning system and their ability to correct algorithmic errors quickly.

A review of different application domains of IML showed further that the IML framework can be beneficial for both, users and the learning model, since it allows users to more thoroughly comprehend the "blackbox" model and to form more nuanced critical opinions about the system outputs [32].

3.2 Active learning for machine teaching

Human-Robot Interaction has been the main area where active learning (AL) for machine teaching has been studied thus far. In recent years, mixed-initiative active learning (Mixed-Initiative Active Learning (MIAL)), i.e. the ability of a system to contribute own queries, has drawn increasing interest from the human-computer interaction (Human-Computer Interaction (HCI)) community [33]. In MIAL there are two major distinctions: system-triggered and user-triggered query contributions.

In related studies, participants e.g. used two colored papers with varying form, size, and color to teach systems concepts like "house." They showed that completely supervised learning outperformed random sampling, however limited the users input [33]. In a follow up study they showed that MIAL performed on equal with active learning but was favored in terms of user engagement. AL mode by itself was seen to be obtrusive and to "negatively affect the teacher's mental model" [34]. Since user-triggered MIAL allows the teacher "full control of the interaction," it was preferred over learner-triggered MIAL based on a number of dimensions, including perceived robot intelligence, ease of teaching, and enjoyment. This preference correlated with the ability of active learning to more precisely predict system performance than supervised learning. Overall, they observed that balancing control early in the learning process and removing unnecessary queries made it easier to increase collaboration, and that user-triggered active learning was the one that participants preferred the most because the interaction appeared natural and simple, despite being thought to be less effective than the other two active conditions.

Data acquisition in active incremental learning: A special aspect of active learning is the data acquisition strategies. Commonly, AL selects the optimal sequence of instances from a pool of data for training, which is not appropriate for fully customized use cases because the data required is not known in advance. Alternative AL algorithms have been put forth in the literature, such as stream-based AL, which "makes immediate query decisions at each instance during a single scan of the data stream" [35]. The algorithm determines whether to query or eliminate a sample at each point in a sequence. Because of this, on the one hand, expensive calculations over huge datasets are not necessary, while on the other hand, stream-based learners disregard the underlying data distribution [36]. However, this is less of a concern for the proposed AL with AML. First, in a fully personalized AL approach, where a user teaches a model from the beginning, the distribution of the future examples is often unknown and second, AML is a model based on combinatorial optimisation with algebras that does not base their predictions on the algebraic representation of the problem.

3.3 Machine Teaching

The term "machine teaching" (Machine Teaching (MT)) originally referred to a theoretical mathematics problem in which the objective was to determine the smallest number of examples required to make an ML algorithm achieve a pre-defined target state [37, 38, 39, 40]. The study of machine teaching (MT) is connected to human-centered ML research [41], as well as human-robot interaction [42, 43], in which a robot is taught a set of tasks by a person. It can be viewed as the opposite of the general machine learning (ML) approach, which commonly seeks to identify a model's final state and tries to "make the learner better by improving ML algorithms" [44]. Instead, MT aims to "make the teacher more productive at building machine learning models" [44]. The HCI-centered approach to MT typically uses non-expert users as teachers such as developers, domain experts, or even the general public. It aims to increase the

level of knowledge on and about ML technologies, often referred to as ML literacy, to a larger community. In order to address the growing need for intelligent and predictive systems with qualified machine teachers who wouldn't need to be familiar with the underlying technology, researchers argued that research should make the interactive process of teaching machines “easy, fast, and universally accessible” [44].

However, there is still lack of research exploring how humans can actually “teach” machines. In order to address this lack of knowledge researchers have turned to ML practitioners to identify applicable strategies as well as observed end-user strategies in a number of experiments. For instance, Hohman et al. found that ML practitioners generally improve model performance by iterating on their data (e.g., gathering extra data, adding labels), as opposed to iterating on the model's architecture and hyper-parameters. The authors presented a prototype for managing various data iterations and highlighted common types of data iterations. [45].

Focusing on end-users, Hong and colleagues conducted experiments on how participants would use their own mobile pictures to train and deploy a image recognition software [46]. One key finding was that participants provided a wide range of images of the same item rather than comparable sample images, which the authors ascribe to human recognition methods that are often independent of size, viewpoint, position, and lighting. A recent think-aloud study [47] looked at how adult novices teach an image classifier using sketches created by participants. The results highlighted how users' strategies and knowledge changed as they learned how to use the system, emphasizing how novices' comprehension is impacted by the optimization inertia of NN. The authors propose directing end-users more towards strategies that structure the training examples in order to progressively add complexity to the training set. Along the same line, Cakmak et al. investigated how users spontaneously train a binary classifier using examples from a limited range of combinations and discovered that participants did not spontaneously produce the best teaching sequences to increase model accuracy [42]. They argue that automated guidance can help improve this result.

Combining both, Wall et al. studied how ML expert comments and teaching approaches might be used to create advice for novices in the process of training a classifier [48]. While the provided advice to novices during the training process did not necessarily improved the classifiers performance, it was perceived as less difficult and more structured by the participants.

In our proposed interactive AML pipeline in D4.8 - Working Prototype and Designs (Revised), we apply this knowledge by providing a guiding step-by-step progression divided into four separate interfaces. The pipeline employs an AL approach to teach ML models user-defined gestures. We accomplish this by giving the end-user the ability to define, record, and delete self-specified training data. We also provide users with the ability to add more training data at any stage, as recommended by existing literature. In later stages, we visualize the trained models' accuracy and allow the user to adapt and compare different model results. We apply these models directly to an application, in this case sound exploration, and allow the user to retrain and adapt any stage of the process using the tools provided.

3.4 Machine Learning Literacy

There is an increasing need and amount of work on how to educate end-users to better understand and reflect on ML technology, which includes data literacy (see review [49]), AI literacy (see review [50, 51]) and overall digital literacy (see review [52]). IML has been utilized as one approach to obtain deeper understanding of ML mechanisms among novices due to its interactive and engaging nature. As a result, a number of IML research focuses on establishing IML systems for ML teaching, or on how participation with an IML system might lead to people learning about ML [53].

Some studies, for instance, used IML to help young children develop more ML literacy skills. In system called Scratch, a visual programming language made specifically for kids [54], Agassi et al. [55] created a gesture detection IML component to encourage kids to include gesture recognition into their Scratch

project so they could independently gather gesture data and learn how to train the model. Research showed that introducing kids early to ML processes, e.g. to train their own gesture-based recognition system, can aid to better understand, reflect and critically think about ML applications in the real world [56]. Sanchez et al. have shown that these results can be transferred to adults using IML pipelines like Marcelle [53, 57].

Following this premise, we created an active teaching application in D4.8 - Working Prototype and Designs (Revised) to make AML more accessible to end users. Based on our previous experience with AML, we believe there is a strong need for end-users such as creative professionals to be able to explore and better understand new ML approaches such as AML. With the increased proliferation of ML applications and the increased diversity of ML approaches, we should think more thoroughly about how to make more complex models, such as AML, that are not intuitively understood, more accessible to a wider audience.

4 Discussion

This deliverable aims to provide background on recent studies investigating machine learning model capabilities and increasing their usability for end users through active learning.

Continuing our last deliverable D4.1 - Human-AML Interaction (Initial), we provide insights into more recent work on theory of mind in human-machine collaboration. We outline the three main concepts involved in TOM and describe their relation to ML. While AML is still in its infancy, we see potential for AML approaches to address these concepts in the future.

Given that AML is a relatively new ML approach with distinct strengths and weaknesses from other ML approaches, we moved on to make these distinctions more understandable and explorable for users. We introduced current work on interactive machine learning and active learning in this deliverable as a way to involve end-users more in the ML process while also assisting them in exploring and making sense of AML potentials. We then discussed active machine learning and how the ability to create and iterate on teaching material for ML applications has great potential for increasing machine learning literacy.

We use this knowledge in our D4.8 - Working Prototype and Designs (Revised) prototype and plan to expand its functionalities in the future to raise more awareness of these differences. Overall, we see a need for more tools that assist users in engaging with and comprehending complex machine learning approaches such as AML. The value added by domain experts to the capacity of ML models is invaluable, as described in existing literature, and the increasing application of ML models in complex human tasks, such as health or art, necessitates greater involvement of these experts. There is a great opportunity to apply ML approaches like AML, which allow for direct decoding of human knowledge as part of their input, in a more interactive way with end-users. While our prototype is an initial attempt to make AML more explorable and interactive, we believe that more work is required to help end-users fully control and interact with complex AML models.

5 Conclusion

The goal of this deliverable D4.2 - Human-AML Interaction (Revised) is to provide background on recent studies investigating machine learning model capabilities and increasing end user usability. We presented insights from both a human-centric and cognitive-science-based perspective on developing a theory of mind between humans and machines. We describe how interactive machine learning can aid in such interactions in order to better understand ML models. We introduce the active learning and machine teaching paradigms and discuss how such approaches could help make AML more accessible to a wider audience. Finally, we emphasise the importance of machine learning literacy as a concept, particularly for new and more complex ML approaches such as AML.

References

- [1] David Engel et al. “Reading the mind in the eyes or reading between the lines? Theory of mind predicts collective intelligence equally well online and face-to-face”. In: *PloS one* 9.12 (2014), e115212.
- [2] Arjun Chandrasekaran et al. “It Takes Two to Tango: Towards Theory of AI’s Mind”. In: *arXiv* (2017).
- [3] Dominic Abrams et al. “The role of cognitive abilities in children’s inferences about social atypicality and peer exclusion and inclusion in intergroup contexts”. In: *British Journal of Developmental Psychology* 32.3 (2014), pp. 233–247.
- [4] Gerry Stahl. “From intersubjectivity to group cognition”. In: *Computer Supported Cooperative Work (CSCW)* 25.4-5 (2016), pp. 355–384.
- [5] Susan R Fussell et al. “How people anthropomorphize robots”. In: *Human-Robot Interaction (HRI), 2008 3rd ACM/IEEE International Conference on*. IEEE. 2008, pp. 145–152.
- [6] Janin Koch and Antti Oulasvirta. “Group cognition and Collaborative AI”. In: *Human and Machine Learning: Visible, Explainable, Trustworthy and Transparent* (2018), pp. 293–312.
- [7] Fabian Lang and Andreas Fink. “Collaborative machine scheduling: Challenges of individually optimizing behavior”. In: *Concurrency and Computation: Practice and Experience* 27.11 (2015), pp. 2869–2888.
- [8] Robert L Goldstone and Georg Theiner. “The multiple, interacting levels of cognitive systems (MILCS) perspective on group cognition”. In: *Philosophical Psychology* 30.3 (2017), pp. 334–368.
- [9] Nuria M Oliver, Barbara Rosario, and Alex P Pentland. “A Bayesian computer vision system for modeling human interactions”. In: *IEEE transactions on pattern analysis and machine intelligence* 22.8 (2000), pp. 831–843.
- [10] Rana El Kaliouby and Peter Robinson. “Mind reading machines: Automated inference of cognitive mental states from video”. In: *Systems, Man and Cybernetics, 2004 IEEE International Conference on*. Vol. 1. IEEE. 2004, pp. 682–688.
- [11] Carl Vondrick, Hamed Pirsiavash, and Antonio Torralba. “Anticipating visual representations from unlabeled video”. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2016, pp. 98–106.
- [12] Carl Vondrick et al. “Predicting motivations of actions by leveraging text”. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2016, pp. 2997–3005.
- [13] Rana El Kaliouby and Peter Robinson. “Real-time inference of complex mental states from facial expressions and head gestures”. In: *Real-time vision for human-computer interaction*. Springer, 2005, pp. 181–200.
- [14] Thomas B Sheridan. “Human-robot interaction: status and challenges”. In: *Human factors* 58.4 (2016), pp. 525–532.
- [15] Christelle Langley et al. “Theory of mind and preference learning at the interface of cognitive science, neuroscience, and AI: A review”. In: *Frontiers in Artificial Intelligence* 5 (2022), p. 62.
- [16] Robert Munro Monarch. *Human-in-the-Loop Machine Learning: Active learning and annotation for human-centered AI*. New York, NY: Manning, June 2021, p. 424.
- [17] Janet Rafner et al. “Deskilling, Upskilling, and Reskilling: a Case for Hybrid Intelligence”. In: *Morals & Machines* 1 (Jan. 2021), pp. 24–39. DOI: 10.5771/2747-5174-2021-2-24.
- [18] Saeid Nahavandi. “Trusted Autonomy Between Humans and Robots: Toward Human-on-the-Loop in Robotics and Autonomous Systems”. In: *IEEE Systems, Man, and Cybernetics Magazine* 3.1 (2017), pp. 10–17. DOI: 10.1109/MSMC.2016.2623867.
- [19] Chad Steelberg. *The path to an AI-connected government*. Tech. rep. Data Center Dynamics, Apr. 2019. DOI: <https://www.datacenterdynamics.com/en/opinions/path-ai-connected-government>.
- [20] Jerry Alan Fails and Dan R Olsen Jr. “Interactive machine learning”. In: *Proceedings of the 8th international conference on Intelligent user interfaces*. ACM. New York, New York, USA: ACM

- Press, 2003, pp. 39–45. ISBN: 1581135866. DOI: 10.1145/604045.604056. URL: <http://dl.acm.org/citation.cfm?doid=604045.604056>.
- [21] Andreas Holzinger. “Interactive machine learning for health informatics: when do we need the human-in-the-loop?” In: *Brain Informatics 3.2* (2016), pp. 119–131.
 - [22] Andreas Holzinger et al. “Towards interactive Machine Learning (iML): applying ant colony algorithms to solve the traveling salesman problem with the human-in-the-loop approach”. In: *International Conference on Availability, Reliability, and Security*. Springer, 2016, pp. 81–95.
 - [23] Dominic Girardi et al. “Interactive knowledge discovery with the doctor-in-the-loop: a practical example of cerebral aneurysms research”. In: *Brain informatics 3.3* (2016), pp. 133–143.
 - [24] Marco Gillies et al. “Human-centred machine learning”. In: *Proceedings of the 2016 CHI conference extended abstracts on human factors in computing systems*. 2016, pp. 3558–3565.
 - [25] Christopher Patrick Laguna and Rebecca Fiebrink. “Improving data-driven design and exploration of digital musical instruments”. In: *CHI’14 Extended Abstracts on Human Factors in Computing Systems*. 2014, pp. 2575–2580.
 - [26] Jules Françoise et al. “Gesture-Sound Mapping by Demonstration in Interactive Music Systems To cite this version : HAL Id : hal-01061221 Gesture – Sound Mapping by Demonstration in Interactive Music Systems”. In: *Proceedings of the 21st ACM international conference on Multimedia* (2014), pp. 1051–1054. DOI: 10.1145/2502081.2502214.
 - [27] Rebecca Fiebrink and Baptiste Caramiaux. *The machine learning algorithm as creative musical tool*. Oxford University Press, 2018, pp. 181–208. ISBN: 9780190226992. DOI: 10.1093/oxfordhb/9780190226992.013.23.
 - [28] Rebecca Fiebrink, Perry R. Cook, and Dan Trueman. “Human Model Evaluation in Interactive Supervised Learning”. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI ’11. ACM, Vancouver, BC, Canada: Association for Computing Machinery, 2011, pp. 147–156. ISBN: 9781450302289. DOI: 10.1145/1978942.1978965. URL: <https://doi.org/10.1145/1978942.1978965>.
 - [29] Qian Yang et al. “Grounding interactive machine learning tool design in how non-experts actually build models”. In: *Proceedings of the 2018 Designing Interactive Systems Conference*. 2018, pp. 573–584.
 - [30] Todd Kulesza et al. “Structured labeling for facilitating concept evolution in machine learning”. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (2014), pp. 3075–3084. DOI: 10.1145/2556288.2557238.
 - [31] Todd Kulesza et al. “Principles of explanatory debugging to personalize interactive machine learning”. In: *Proceedings of the 20th International Conference on Intelligent User Interfaces*. ACM, 2015, pp. 126–137. DOI: 10.1145/2678025.2701399.
 - [32] Saleema Amershi et al. “Power to the people: The role of humans in interactive machine learning”. In: *AI Magazine* 35.4 (2014), pp. 105–120.
 - [33] Maya Cakmak and Andrea L Thomaz. “Mixed-Initiative Active Learning”. In: (2008).
 - [34] Maya Cakmak, Crystal Chao, and Andrea L Thomaz. “Designing Interactions for Robot Active Learners”. In: *IEEE Transactions on Autonomous Mental Development* 2.2 (2010), pp. 108–118. ISSN: 1943-0604. DOI: 10.1109/TAMD.2010.2051030.
 - [35] Chen Change Loy et al. “Stream-based joint exploration-exploitation active learning”. In: *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*. 2012, pp. 1560–1567.
 - [36] Shen Shyang Ho and Harry Wechsler. “Query by Transduction”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 30.9 (Sept. 2008), pp. 1557–1571. ISSN: 01628828. DOI: 10.1109/TPAMI.2007.70811. URL: <http://ieeexplore.ieee.org/document/4384495/>.
 - [37] Ayumi Shinohara and Satoru Miyano. “Teachability in computational learning”. In: *New Generation Computing* 8 (1991), pp. 337–347.
 - [38] Weiyang Liu et al. “Iterative Machine Teaching”. In: *arXiv preprint arXiv:1705.10470* (2017). URL: <http://arxiv.org/abs/1705.10470>.

- [39] Weiyang Liu et al. “Towards Black-box Iterative Machine Teaching”. In: *arXiv* (2017). URL: <http://arxiv.org/abs/1710.07742>.
- [40] Xiaojin Zhu. “Machine Teaching: An Inverse Problem to Machine Learning and an Approach Toward Optimal Education”. In: *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence*. 2015, pp. 4083–4087.
- [41] Marco Gillies et al. “Human-centred machine learning”. In: *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems*. 2016, pp. 3558–3565.
- [42] Maya Cakmak and Andrea L Thomaz. “Eliciting good teaching from humans for machine learners”. In: *Artificial Intelligence* 217 (2014), pp. 198–215.
- [43] Andrea L. Thomaz and Maya Cakmak. “Learning about objects with human teachers”. In: *Proceedings of the 4th ACM/IEEE international conference on Human robot interaction* (2009), p. 15. DOI: 10.1145/1514095.1514101.
- [44] Patrice Y Simard et al. “Machine teaching: A new paradigm for building machine learning systems”. In: *arXiv preprint arXiv:1707.06742* (2017).
- [45] Fred Hohman et al. “Understanding and visualizing data iteration in machine learning”. In: *Proceedings of the 2020 CHI conference on human factors in computing systems*. 2020, pp. 1–13.
- [46] Jonggi Hong et al. “Crowdsourcing the perception of machine teaching”. In: *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 2020, pp. 1–14.
- [47] Téó Sanchez et al. “How do People Train a Machine? Strategies and (Mis)Understandings”. In: 5 (CSCW Oct. 2021). DOI: 10.1145/3449236. URL: <https://hal.inria.fr/hal-03182950>.
- [48] Emily Wall, Soroush Ghorashi, and Gonzalo Ramos. “Using expert patterns in assisted interactive machine learning: A study in machine teaching”. In: *Human-Computer Interaction–INTERACT 2019: 17th IFIP TC 13 International Conference, Paphos, Cyprus, September 2–6, 2019, Proceedings, Part III* 17. Springer. 2019, pp. 578–599.
- [49] Ying Cui et al. “Data literacy assessments: a systematic literature review”. In: *Assessment in Education: Principles, Policy & Practice* 30.1 (2023). Publisher: Taylor & Francis, pp. 76–96.
- [50] Matthias Carl Laupichler et al. “Artificial intelligence literacy in higher and adult education: A scoping literature review”. In: *Computers and Education: Artificial Intelligence* (2022). Publisher: Elsevier, p. 100101.
- [51] Davy Tsz Kit Ng et al. “Conceptualizing AI literacy: An exploratory review”. In: *Computers and Education: Artificial Intelligence* 2 (2021). Publisher: Elsevier, p. 100041.
- [52] Sarah Soyeon Oh et al. “Measurement of digital literacy among older adults: systematic review”. In: *Journal of medical Internet research* 23.2 (2021). Publisher: JMIR Publications Toronto, Canada, e26145.
- [53] Téó Sanchez. “Interactive Machine Teaching with and for Novices”. PhD thesis. Université Paris-Saclay, 2022.
- [54] Mitchel Resnick et al. “Scratch: programming for all”. In: *Communications of the ACM* 52.11 (2009), pp. 60–67.
- [55] Adam Agassi et al. “Scratch Nodes ML: A Playful System for Children to Create Gesture Recognition Classifiers”. In: *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*. CHI EA ’19. Glasgow, Scotland Uk: Association for Computing Machinery, 2019, pp. 1–6. ISBN: 9781450359719. DOI: 10.1145/3290607.3312894. URL: <https://doi.org/10.1145/3290607.3312894>.
- [56] Tom Hitron et al. “Can Children Understand Machine Learning Concepts? The Effect of Uncovering Black Boxes”. In: *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. CHI ’19. Glasgow, Scotland Uk: Association for Computing Machinery, 2019, pp. 1–11. ISBN: 9781450359702. DOI: 10.1145/3290605.3300645. URL: <https://doi.org/10.1145/3290605.3300645>.
- [57] Jules François, Baptiste Caramiaux, and Téó Sanchez. “Marcelle: composing interactive machine learning workflows and interfaces”. In: *The 34th Annual ACM Symposium on User Interface Software and Technology*. 2021, pp. 39–53.