



Human-centric algebraic machine learning

Evaluation - Initial

Deliverable D4.10

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

uc3m



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: 54 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: **D4.10**
 Title: **Evaluation - Initial**
 Month: 54
 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: 2025-02-27
 Version: v2.0
 Status: Finished

Version history

Version	Date	Responsible	Author/Reviewer	Comments
v1.0	2024-12-10	Inria	Inria	draft
v2.0	2025-02-27	Inria	Inria	final

Executive summary

AML 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 D4.10 - Evaluation - Initial* of the ALMA project. The goal of this deliverable is to run a controlled laboratory experiment to determine the effectiveness of the different interaction techniques.

This deliverable was integrated into D7.3 - Intelligent Tools for supporting creative professionals (Initial), where we described the design of *AMLExplorer* and evaluated it with 12 participants. This deliverable described the design of a study in which participants compared their experiences training the AML algorithm to their experiences training a more traditional neural network, and the preliminary results for the potential of AML for creative expression. We will repeat the information already presented in D7.3 - Intelligent Tools for supporting creative professionals (Initial) for completeness.

Contents

Executive Summary	3
Contents	4
Acronyms	5
1 Introduction	6
2 Demonstration as described in D7.3	6
3 Study Design as described in D7.3	8
3.1 Overview	8
3.2 Participants	8
3.3 Setup	8
3.4 Procedure	8
3.5 Data collection	9
4 Results as described in D7.3	9
5 Discussion as described in D7.3	10
5.1 Limitations	10
5.2 Future Work	10
6 Summary	10
A Appendix: Questionnaire	11

Acronyms

AML Algebraic Machine Learning.

1 Introduction

This report corresponds to *Deliverable D4.10 - Evaluation - Initial* of the ALMA project. The goal of this deliverable is to run a controlled laboratory experiment to determine the effectiveness of the different interaction techniques. This deliverable was integrated into D7.3 - Intelligent Tools for supporting creative professionals (Initial), where we described the design of *AMLExplorer* and evaluated it with 12 participants. This deliverable described the design of a study in which participants compared their experiences training the AML algorithm to their experiences training a more traditional neural network, and the preliminary results for the potential of AML for creative expression. We will repeat the information already presented in D7.3 - Intelligent Tools for supporting creative professionals (Initial) for completeness.

In the following, we will outline the demo of the AML tools and the conventional neural network, the study design, results, and discussions again. More information about this prototype can be found in D7.3 - Intelligent Tools for supporting creative professionals (Initial). We developed this idea further and presented an iteration for more exploratory music generation with AML in D7.4 - Intelligent Tools for supporting creative professionals (Final) as well as for professional users to train and explore AML potentials in D4.9 - Working Prototype and Designs (Final).

2 Demonstration as described in D7.3

We describe a typical walk-through of our system from a user's perspective and explain the different conditions that were evaluated in this deliverable.

First, the user starts by collecting their own gesture data for each of the preferred classes in order to combine them later to musical patterns (see Fig. 1). By selecting a label and pressing down the collect button, a stream of data is collected with the highlight of the finger points most relevant to the model. The user can delete or edit the collected data at any point.

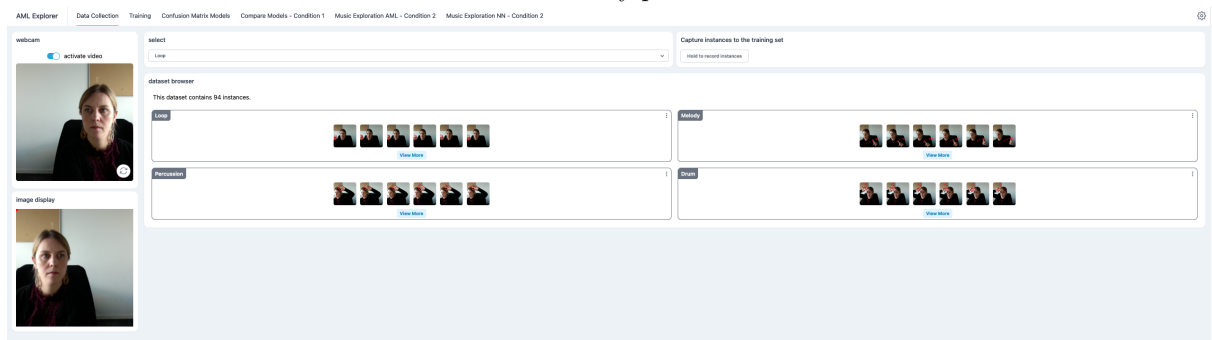


Figure 1: Data Collection

After the user is satisfied with the data, the training tab (Fig. 2a) allows to train the AML model and comparison model. The user can choose to either use the standard training parameter or change them. The accuracy graph and confusion matrix (Fig. 2b) allow the user to better understand the quantitative differences between the two models. This interfaces are described more in detail in D4.8 - Working Prototype and Designs (Revised).

In order to allow users to explore the difference more interactive we developed an comparison interface, allowing users to recreate the gesture, observing the models confidences. Through regular adaptation and retraining of the models, the user can explore what aspects of improve the recognition of the individual gesture by each system (Fig. 3).



Figure 3: Understanding qualitative differences - Condition 1

Finally, the user can explore how to combine these self-trained gestures to create new rhythms by moving through the self-created gesture space. To better understand the unique capabilities of the models we separated the exploration condition of the different models but kept the interface the same (see Fig. 4 and 5). Throughout the exploration users could go back and retrain gestures to explore different gesture spaces.



Figure 4: Exploring rhythms using self-trained gestures with AML

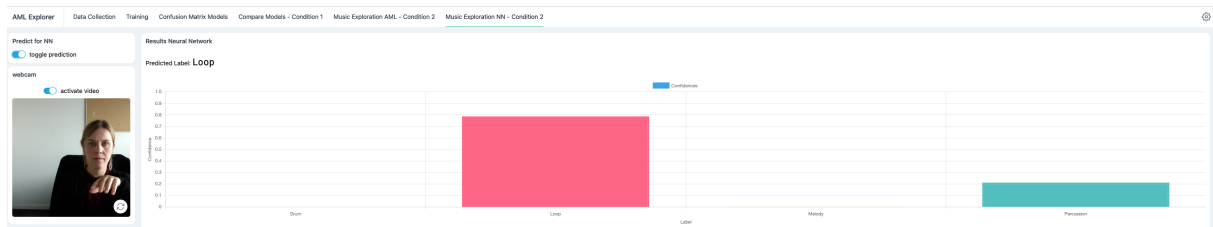


Figure 5: Exploring rhythms using self-trained gestures with NN

All functionalities will be available in the future online version and are creating the baseline for more focused exploration of model explainability in the future.

3 Study Design as described in D7.3

3.1 Overview

We are interested in whether or not interactive exploration of AML-based capabilities improves user's ability to define and control personalized gestures. We designed a simple study with three key objectives:

1. Demonstrate an interactive use of AML;
2. Demonstrate how users can compare the characteristics of AML with another machine learning algorithm; and
3. Determine how well participants can understand the characteristics of each model.

The study involves first asking participants to create their own free-hand mid-air gestures and associate them with parameterized sound variations, and then to train and compare AML with another machine learning model, a Multilayer perceptron which we call 'NN' in the following. We are interested in the process they use to train the models, their confidence in the effectiveness of each model, and whether and how their understanding of the model's characteristics improves through the retraining process.

3.2 Participants

We recruited 12 participants (**5 men, 7 women, ages ranged from 25–36 years**). Nine participants had musical experience and 10 had an engineering background (double naming possible).

3.3 Setup

An Apple Macintosh laptop running *AMLExplorer* (described in section ??) was placed on a table. The experimenter sat next to the participant so that both could easily see the screen.

3.4 Procedure

Each session lasts approximately 30 minutes. The experimenter begins by explaining the study. Participants are informed that their data will be anonymized and handled according to GDPR rules, and that they can withdraw from the study at any time. After obtaining informed consent, the experimenter launches the *AMLExplorer* prototype and begins screen and audio recording. Participants are encouraged to use a talk-aloud protocol in which they describe their intentions and actions as they produce each gesture.

First the users are encouraged to record personal gestures for the four musical attributes loop, melody, percussion and drums in the data collection window and train the AML and NN models based on these data. The following study is then structured with *two conditions* in a within-participants design:

In the *Comparison* condition, participants are first asked to select one of the four different gestures associated with four musical attributes: *loop*, *melody*, *percussion*, and *drum*. They are then asked to recreate this gesture by activating a new recording. Users then compare how well the two models, AML or NN, categorize each gesture using the corresponding confidence percentage score presented for each model. If the participant is not satisfied with how well the models identified the gesture, they can retrain the model by going through the initial steps of data collection and training. Participants are encouraged to explore how both models react to adding and changing input and report on their qualitative understanding of how each model functions. When they have finished training and comparing the models, participants answer the questions associated with the Comparison condition (see Appendix A).

In the *Exploration* condition, participants use each of the two models, AML and NN, to create new rhythms. (The presentation of each model is counter-balanced for order across participants.) When they have finished training and comparing the models, participants answer the questions associated with the Exploration condition (see Appendix A).

3.5 Data collection

We collected audio and screen recordings as well as responses from a questionnaire and hand-written notes.

4 Results as described in D7.3

In the Comparison condition, most participants (9/12 – 75%) either agreed or strongly agreed that they needed to retrain the models. Few participants (3/12 – 25%) never retrained the models and only one retrained more than three times. The rest (8/12 – 67%) retrained 1–3 times. Most participants (10/12 – 83%) felt they could easily perform each gesture again and only one strongly disagreed. However, only two participants (17%) felt they could easily understand how the models' confidence levels improved and the rest were either neutral (4/12 – 33%) or disagreed (5/12 – 42%). One participant said “I really have no idea how these things work, I would probably have to play with it for another hour, as I suspect I'm running into an issue with the hand recognition and recording.” Another participant noted that “More diverse gestures are easier to predict, more variability in the training images also help”.

Half of the participants were satisfied or very satisfied (6/12 – 50%) with the AML model's prediction, while only one participant was very satisfied (1/12 – 12%) with the NN model's predictions. Interestingly, of the participants who had a preference, most preferred AML (6/8 – 75%) and only two (2/8 – 25%) preferred NN. A few participants (3/12 – 25%) felt they better understood how to improve model predictions after their experiences in Condition 1. However, many were neutral (5/12 – 42%) or disagreed (4/12 – 33%).

In Condition 2, only one third of the participants (4/12 – 33%) felt they could easily create rhythms. When asked to rate their experience with the AML model, only three (3/12 – 25%) gave it a positive rating and half gave it a neutral rating. However, only two (2/12 – 17%) gave the NN model a positive rating and only four (4/12 – 33%) gave it a neutral rating. The rest were negative.

When asked to assess the predictability of each model, only two (2/12 – 17%) found the AML model predictable and the rest were either neutral (4/12 – 33%) or negative (6/12 – 50%). Somewhat surprisingly, four participants (4/12 – 33%) rated the NN model as predictable, although half (6/12 – 50%) were negative.

Participants also reported on how they thought the *AMLExplorer* tool might change their interaction with music. Several commented that it helped them think more about rhythms and gestures. One participant liked seeing the relationship between rhythms and visuals and another commented that they would “use their body more”. Another commented that “if it worked it would have been amazing and easier to make rhythms with body interactions more natural”

One participant said: “this could have a positive interaction with music personally. Because, for me, it is hard to play instruments, and with this tool I could create my own instruments.” and another felt it was “an interesting experience, it gives me the idea that I could create a spatial instrument.” Finally, one participant appreciated how “it can create surprises in how the musical clips are transitioned to one another or blended into each other.”

5 Discussion as described in D7.3

The above results are interesting, in that we expected the much more mature NN model to out-perform this initial test of the AML model in an interactive setting. However, the two models performed similarly, and participants expressed a slight preference for AML. We also observed that some participants better understand model capabilities after retraining them, which supports our intention of using interactive ML approaches for AML to increase ML literacy. To be clear, these results are not statistically significant and we would need to run a more thorough study, but they are certainly promising for AML.

5.1 Limitations

This prototype and the associated study are clearly the first step in a much longer research process. Although we have demonstrated the feasibility of creating an interactive AML pipeline that can provide an interactive experience for users, much remains to be done with respect to the design of the interaction and further studies, both quantitative and qualitative, are required to fully assess the benefits and challenges associated with the AML approach to machine learning.

5.2 Future Work

We will extend the existing version of *AMLExplorer* to address further questions regarding explainability. The presented study will serve as a reference point for comparing future iterations of this prototype. The compare tab now enables us to isolate prediction functionalities and analyze them in depth, which will be particularly crucial when dealing with more specific aspects of model understanding.

6 Summary

The Deliverable D7.3 - Intelligent Tools for supporting creative professionals (Initial) described the first evaluation of the *AMLExplorer* working prototype of an interactive AML pipeline. Our goal was to demonstrate a Use Case in which musicians can create their own personally meaningful free-hand gestures and associate them with labels associated with parameterized sounds.

This deliverable was followed by D7.4 - Intelligent Tools for supporting creative professionals (Final), in which we show how AML can be easily applied to the diverse needs of professional musicians, exploring the potential for AML end-user applications. We further continued to investigate new approaches for how to increase ML literacy with new approaches like AML in D4.9 - Working Prototype and Designs (Final).

A Appendix: Questionnaire

The questionnaire first asks for the participant's age and gender, and then whether or not they musical experience and if they have an engineering background.

Condition 1: Comparison

Participants estimated how often they retrained the models.

Four 5-point Likert-type questions (scale: "strongly disagree" to "strongly agree") asked participants to assess their behavior and understanding of the models:

1. I needed to retrain the models.
2. I could easily perform the gesture again.
3. I could easily understand how the model's confidence improved.
4. I better understand how to improve model predictions now.

Participants also rated their satisfaction with each of the two models (scale: "not satisfied" to "highly satisfied") and whether they preferred one model or another.

Condition 2: Exploration

Two 5-point Likert-type questions (scale: "strongly disagree" to "strongly agree") asked participants to assess their ability to create specified musical effects and rhythms:

1. I can easily make a specific gesture for each musical effect.
2. I can easily create rhythms.

Four questions asked participants to rate their experiences (scale: "poor" to "good") with different aspects of the two models:

1. Rate your rhythm creation experience with the AML model.
2. How predictable was the AML model?
3. Rate your rhythm creation experience with the NN model.
4. How predictable was the NN model?

Finally, we asked an open-ended question:

7. How would this tool change your interaction with music?