



Human-centric algebraic machine learning

AML based world models

Deliverable D5.4

WP5 - AML based world models - Initial



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: **D5.4**
 Title: **AML based world models**
 Month: 9
 Work Package: WP5 - AML based world models - Initial
 Work Package leader: DFKI
 Deliverable leader: DFKI
 Deliverable type: R
 Dissemination level: PU
 Date of submission: 2021-05-31
 Version: v1.0
 Status: Finished

Version history

Version	Date	Responsible	Author/Reviewer	Comments
v1.0	2021-04-08	DFKI	V. Fortes, S.Fox	First Draft
v1.1	2021-04-25	ALG	F. Martin-Maroto	Embedding
v1.2	2021-05-12	DFKI	V. Fortes, S.Fox	Complete Draft
v1.2	2021-05-12	ALG	R. Sánchez-Mateos	Formatting and cohesion

Executive summary

In this deliverable, *Deliverable D5.4 - AML based world models - Initial*, we make a start on exploring the potential of AML to support ethical world models. As ALMA is a Future & Emerging Technologies (FET) project, we adhere to the FET aims of novel research that lays the foundations for next generation technologies.

In particular, we explore the potential for AML for the development of AI to support ethical behavior. This addresses ethical issues in human interaction with AI in a new human-centric way. It is a new approach that is based on the latest research in behavioral ethics. This reveals that artificial intelligence has potential to directly address mundane everyday factors that can lead to unethical human behavior.

At the same time, we recognize that FET research should have the medium term goal of converting proofs of concept into industrial applications and systems. Accordingly, in this deliverable, we relate the properties of AML as a new ML paradigm to established organization practice. In particular, we relate AML to quality management systems that connect the everyday practices of organizations to relevant laws, regulations, standards, etc., which provide organizations with ethical frameworks for their internal operations and for their interactions with customers, end-users, and other stakeholders.

This deliverable, D5.4. AML based world models - Initial, will be followed by D5.5 AML based world models - Revised, and D5.8 AML based world models - Final. In addition, Work Package 5 will deliver D5.6 Embedding of ethics and cultural concerns into AML based world models - Initial, and D5.7 Embedding of ethics and cultural concerns into AML based world models - Final. In those future deliverables, we shall report our further research into the potential of Algebraic Machine Learning to improve interactive human-centric machine learning systems in complex settings.

Contents

Executive Summary	3
Contents	4
1 Overview of preliminary AML-based world models	6
1.1 Potential for AML to support ethical action in complex situations	6
1.2 Building transparent solutions to support ethical action	7
2 Preliminary ontology	7
2.1 Reduce sources of human self-regulatory resources depletion	7
2.2 Apply relevant ethical work procedure for current situation	8
3 Exploratory Case	9
3.1 Quality Management Systems (QMS) to facilitate human oversight	9
3.2 Scenario Description	10
3.3 Scenario Implementation	11
3.4 Preliminary Ethical Framework	13
3.4.1 AML based AI Ethical Assistance Interaction	14
3.4.2 Implications for human - AI interaction research	16
3.4.3 Implications for human - AI interaction practice	16
3.5 AML Embedding	16
3.5.1 Simple Embedding	17
3.5.2 Embedding temporal dynamics	19
4 Conclusion	21
References	23

1 Overview of preliminary AML-based world models

Within this deliverable, AML-based world models are models of sociocultural ethical behavior that involve human interactions with AML-based AIs. In particular, an AI's actions should be ethical and transparent in supporting human ethical actions during complex work situations. First, let's discuss its two principal characteristics.

1.1 Potential for AML to support ethical action in complex situations

Recent research by others suggests that most people typically adhere to sociocultural ethical practice unless under stress. This line of research can be summarized with the name *behavioural ethics* [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16]. For example, stress can lead to what can be described as *allostatic overload*. This term refers to the regulatory capacity of a control system being overtaxed by environmental perturbations. This is most important where regulatory work increases to the point where energy demand exceeds energy supply. In other words, to the point where a person is depleted of sufficient resources to function well. This can lead to altered activity in brain areas involved in law-abiding and moral behavior. Specifically, severe stress decreases activity within brain areas that support some of the highest forms of contextual integration, leads to top-down collapse of higher goals, and the favoring of short-term aims [1, 2].

In complex situations, ethical stress can come from interactions between ethical temptation and moral motivation, which depend on dynamic combinations of variables that can come together differently in different situations [3, 4, 5]. These include, for example, individual's self-regulatory resource depletion and organizational pressure. Having the moral motivation to resist ethical temptation can be a struggle [6], in which non-ethical impulses can override moral reflection [7]; sometimes with disastrous consequences for many people [8].

Organizational pressure can arise from conflicts between organizational goals, such as quality versus profit, when operations do not go to plan amidst complex situations. Employees who identify strongly with an organization can go along with organizational pressure to take non-ethical action [9, 10]: especially if their self-regulatory resources are depleted [11], for example from overwork for the organization [12]. Accordingly, organizations should have procedures to prevent employees being placed under pressure to take non-ethical actions [13]. This is always important and particularly important when employees' actions could have negative consequences for customers or anybody else who could be affected: for example, residents in care homes and visitors to residents. Opportunities for AML-based algorithms to support human ethical action are revealed by findings from behavioral ethics research [8], and consideration of actions that can be performed with them. For example, time pressure can increase ethical temptation [14, 15], and AML-based AIs could reduce time pressure on human employees by carrying out routine tasks. Also, provision of moral cues can increase moral motivation [16], and AIs using the AML approach could provide moral cues by presenting the relevant ethical work procedure to the human employee.

Clearly, very careful planning of work procedures is required to ensure that they are compatible with sociocultural perspectives of all potential stakeholders including, for example, employees, customers, end-users, and any others who may interact with the organization. On the other hand, although definitions of ethical behaviour in different settings can be challenging, there is a wide range of resources that can be drawn upon when doing so. These include laws, regulations, standards, and emerging best practice as defined by, for example, professional institutes. Indeed, it is common for research projects to have a deliverable focused on project ethical requirements. In this project that is Deliverable 9.2, and such deliverables can be prepared at the outset of projects with much thought but without delay.

1.2 Building transparent solutions to support ethical action

Sociocultural ethical cognitive processing is both top-down, for example from sociocultural cognitive models, and bottom-up from situation-specific sensory inputs [17, 18]. In terms of AI, such cognition processing can be described as hybrid [19]. Hybrid AI is recognized as being appropriate for artificial ethical agents that seek to mimic human moral judgment in domains that are limited in scope and well-defined [20]. Hybrid AI can be considered to be a third phase in the development of AI that aims to counteract previous weaknesses and harness previous strengths. Specifically, a third phase that aims to combine the knowledge-guided top-down basis of its first phase with the data-driven bottom-up basis of its second phase [21, 22].

Hybrid AI is not monolithic. For example, computer vision AI from one developer / vendor can provide inputs for robot operation AI from another developer / vendor. Within hybrid AI using AML, one of its goals is to make possible for both top-down rules and bottom-up sensory inputs to be human-readable. Notably, within AML, top-down rules can be pre-defined as human-readable first-order-logic statements that can be expressed in algebras. Furthermore, it has the potential for bottom-up sensory inputs to be represented in human-readable formats that can be expressed in the same algebras. A mathematical model for the current situation can be built comprising both pre-defined rules and sensory inputs.

AML-based rules can be transparent through their compliance with work procedures that are defined by humans. These work procedures can be described through inter-related formats. Firstly, in natural language and simple diagrams that are readily understandable to a wide range of humans. These can then be distilled into first-order-logic rules. In particular, within the top-down pre-defined human-readable first-order-logic rules. Here, it is important to note that top-down rules defined with AML do not change opaquely and unpredictably. Rather, revision of top-down rules are subject to human review. Overall, there can be transparency of top-down rules, bottom-up inputs, and the models of current situations using AML. Hence, situation-specific actions of AML algorithms can be transparent because the mathematical models of situations can be analysed by humans in order to understand the logic of AI situation-specific actions [23]. In particular, this can be possible because the eventual mathematical models have their grounding in inter-related formats that are bridged with algebraic representations: from first-order-logic statements to code that handles sensory inputs.

2 Preliminary ontology

Preliminary ontology for world model in which AML-based algorithms can offers ethical assistance is provided in Figure 1. This shows that AI ethical assistance can facilitate human ethical action by minimizing causes of human non-ethical actions. This can be done by reducing sources of human self-regulatory resources depletion in order to reduce situation ethical temptation. At the same time, it can be done by applying the relevant ethical work procedure as a cue in order to increase personal moral motivation, see Figure 4.

2.1 Reduce sources of human self-regulatory resources depletion

Depletion of human self-regulatory resources can undermine resistance to ethical temptation [4, 16]. Potential for depletion can be reduced by maximizing signal-to-noise ratio and by minimizing non-value-adding actions. It is important to maximize signal-to-noise ratio in information communication so that people can receive the intended meaning from information immediately. This is important because the more ways that the same information can be interpreted the more attempts may have to be made for information to be acted upon as intended, and the more attempts that have to be made the more depleted people can become. Accordingly, information should be free from presentational ambiguities such as inconsistencies between numbers, words, and diagrams. Moreover, ethical assistance should be designed to avoid causing human employees anxiety, which can have negative effects on human perception. For

This checking and authorization is done by humans. Furthermore, checking and authorizing is done in accordance with relevant laws, regulations, standards, etc. Moreover, QMS work procedures are subject to periodic review both internally by the organization itself and by external certification authorities. Both internal and external review is carried out by humans who are experts in the relevant type of work. Importantly, reference to relevant work procedures can provide cues for reminding ethical actions and for encouraging ethical actions: both of which can motivate human ethical action [31, 16]. Reminding involves presenting the relevant ethical work procedure for the situation to the human. Encouraging ethical actions can involve informing about the organization's positive track record so far and the importance of maintaining it. The AI ethical assistance should present the same reminding and encouraging information, for example, to the employee and to the employee's supervisor. This is necessary to make sure that both are paying attention to ethical work procedures even if one or both of them is depleted [32].

As summarized in Figure 1, situational ethical temptation and employee moral motivation can be linked to each other. For example, there can be unexpected disturbances in complex situations, which lead to extra actions having to be taken that reduce the amount of time available to complete the work. Accordingly, moral motivation should be increased through reminding and encouraging whenever there is an unexpected disturbance of work, which can lead to employees being more tempted to take non-ethical short-cuts [33, 34]. The temptation to take short-cuts could be reduced by introducing AML-based AI. For example, temptation can be reduced by explaining that immediate short-cuts will lead to more actions in the medium-term, e.g. by having to undo work and do it again in accordance with the ethical work procedure. Moreover, importance will increase if ethical action in complex in work situations is recognized.

Ideally, human-AI interaction needs to be designed so that the system helps mitigate unethical actions. This design being carried out by humans during the development of work procedures, which comply with relevant laws, regulations, standards, etc., that are verified by humans in external QMS certification authorities. Here, it is important to note that QMS is applied throughout the world. For example, according to the website of the International Standards Organization (ISO), there are over one million organizations in more than 170 countries that are certified to ISO 9001, which sets out the criteria for a quality management system. Thus, it is likely that many organizations will need to incorporate AI into their QMS. Accordingly, it is realistic to situate introduction of AI into the practicalities of QMS, and to make the best of the opportunities that doing so can provide.

3 Exploratory Case

3.1 Quality Management Systems (QMS) to facilitate human oversight

In this exploratory scenario, there is a fictional care home. Ethical work procedures are defined through reference to all relevant laws, regulations, standards, etc., by human experts. The process of formulating and reviewing work procedures is carried out within the care home's QMS. Notably, in order to have Quality Assurance (QA) certification, organizations need to have a QMS that is in accordance with relevant laws, regulations, and standards. QA certification bodies are external to the organization applying for the Quality Assurance certification.

Within QMS, there can be at least four levels of human oversight. First, there is general legislation in society about what is and is not ethical. Second, sector regulations, standards formulated by humans have to be consistent with relevant legislation. Third, there are the Quality Assurance certification bodies whose human personnel that have to ensure that any applicant's QMS is in accordance with laws, regulations, and standards. Fourth, there is the human team in the organization that develops the QMS usually with a consultancy firm that specializes in QMS. Then, the QMS is continually reviewed internally by human managers to ensure internal conformance to QMS procedures. In addition, the QMS

is reviewed externally by the Quality Assurance certification body that ensures that the QMS is kept up to date with new laws, regulations, and standards. In daily operations, what is and what is not ethical is determined through conformance or non-conformance with the relevant QMS work procedure, which as explained above has at least four levels of human oversight.

Within Quality Management Systems (QMS), work procedures are developed to be consistently capable of enabling the organization to meet the requirements of customers, end-users and other stakeholders, and to do so within relevant laws, regulations, and sector standards. Importantly, the development process involves designing out ambiguities and other sources of uncertainty from work procedures in order that they can be carried out consistently by diverse work forces. Thus, the design and monitoring of work procedures within the QMS is consistent with human-on-the-loop (HOTL) where there is very active human intervention during the design cycle of the system and monitoring the system's operation. Moreover, there is human-in-command (HIC) as humans have the capability to oversee the overall activity of the system in terms of its relation to laws, regulations, and standards.

This could be very difficult, but many inter-related methods have been developed to address the challenge of designing out uncertainties from work procedures. For example, there can be systematic application of techniques such as process modelling, and failure mode and effects analysis, which are applied in order to define work procedures with specificity and make them fail-safe. From the point-of-view of introducing AML, the reduction of uncertainty about the content of work and allocation of work in QMS work procedures increases potential for algebraic descriptions. By contrast, trying to introduce AML-based algorithms into work environments involving haphazard application of ad hoc procedures may be neither possible nor worthwhile.

3.2 Scenario Description

In the following exploratory case scenario, an AML-based algorithm is introduced into organizational practices that are well developed and defined within a care home that has a QMS system. Nonetheless, as shown in the exploratory case, there can be dynamic complexity within organizations working within a QMS. Thus, the potential of such AML-based systems is explored in order to reveal opportunities and challenges in work settings that have potential for implementation, but which are nonetheless complex.

In particular, the AML-based AI implementation in the exploratory scenario carries out **very specific tasks**, which are consistent with recent advances in medical scanning and analyses by AI. The specific tasks are the scanning of smart care home resident's skin maladies:

- *analyse* scans of skin maladies to determine the current condition
- *identify* trends in each resident's scans to determine whether the skin malady is deteriorating or improving
- *present* these trend information to human carers
- *recommend* to the human carer potential medication options

The overall goal being to improve the residents' skin maladies, while administering neither too little nor too much skin medication. For individual residents, this focus on providing the correct amount of medication is important to facilitate efficient and safe improvement of residents skin maladies. More generally, it is important that medication is applied optimally in order to avoid widespread negative unintended consequences such as medications becoming generally less effective. Hence, it is the care home's ethical responsibility to design and maintain a system that facilitates the correct skin treatment for residents by human carers. As explained above, this system has been designed and is monitored with human-in-command (HIC) and human-on-the-loop (HOTL). Moreover, throughout, there is human-in-the-loop (HITL) as there is capability for human intervention in every decision cycle of the system.

Specially, the exploratory scenario is that in a smart care home, there are C carers and R residents. Our scenario is concerned with the treatment of the skin malady, dermatitis, which may occur in those residents. This smart home is equipped with a "robot" that, through a camera, can evaluate the gravity of the dermatitis through scanning as outlined above. When a resident shows symptoms, they will use the "robot" and report to a carer. The carer then decides on a treatment based on the readings from the robot as well as the complaints expressed by the resident. Each carer has their own (assumed here to be unchanging in the short-term) view on health. Carers with a more optimistic view could opt for too little medication, while overcautious carer's may opt for too much medication. However, either too little medication or too much medication can lead to the deterioration of symptoms. This can be upsetting for residents. Of course, there are several sources of uncertainty in the scenario. The skin readings from the robot may not always be indisputably accurate. Thus, there is a probability rather than an indisputable certainty of reading per skin malady severity. The residents also have their own overall humor; general attitudes towards their own health, which is influenced by their (here, assumed to be unchanging in the short-term) baseline attitude on their health, and how the dermatitis is making them feel. A resident's attitude about own health can influence the severity of their complaints, which in turn can influence the medication the carer selects. Medication options are a non-prescription moisturizing ointment that can be used indefinitely; prescription steroid cream that can be used for only two weeks followed by three week break before being used again; antibiotic tablets to be used for one week only once per year.

Another factor that changes the resident's mood is a visit from a dog, which improves the overall mood of the resident. On the other hand, some residents are allergic to the dog and so its visits can flare their allergies and worsen the state of their dermatitis or even start it. Regarding the dog, the care home has procedures for the registration of guests and any pets that they may bring. This is needed for several reasons, including the need to know who is in the building for security reasons; in the unfortunate event of a fire; etc. Through connection with the routine registration of visitors, the AML system knows which days the dog visited, but not who is allergic.

We can describe this scenario also as a causal graph, as we can see in Figure 2. As depicted there, some information is not directly available to the AML system, but can be derived from its interactions with things that it can observe. For example, the system can access the care home's guest register in order to determine what pets may have visited. The idea is that prior to its deployment the system will learn having access to a historic of data generated from that causal graph, which has some labels (we know at some points the disease severity because some lab tests were done). Some questions we want the AI system to answer can be seen in Figure 3. The main purpose of the AI in this scenario is to guide carers to the required treatment by keeping them up-to-date with the trend in each resident's skin malady condition.

3.3 Scenario Implementation

Let's describe here more details of how we will simulate the scenarios and therefore generate our historic data (as well as test data).

As we explained before, in our scenario there are C carers and R residents. Thus, we will simulate different numbers of carers and residents. In each simulation, every carer C_j will have a one view of health $CarerVOH_j$, which will be an integer value in the range $[-4, 4]$. The same goes for the resident's base view on health, $ResidentBaseVOH_i$, which will be in the range $[-4, 4]$. Each R_i is either allergic or not to the dog, $AllergicR_i$. Those variables are defined in the beginning of the simulation and will not change in the short-term. To generate our diverse worlds, we will create multiple scenarios with a different number of carers, residents and at each simulation vary their views on health.

At the start of the simulating time, t_0 , a subset of R is selected to start having dermatitis. The real disease severity is represented by an integer value in the range $[0, 15]$, and the initial value for will be sampled from a distribution $P_{initialDerm}$. At each time t a resident has a value for disease severity $R_t^{Derm_i}$. Other

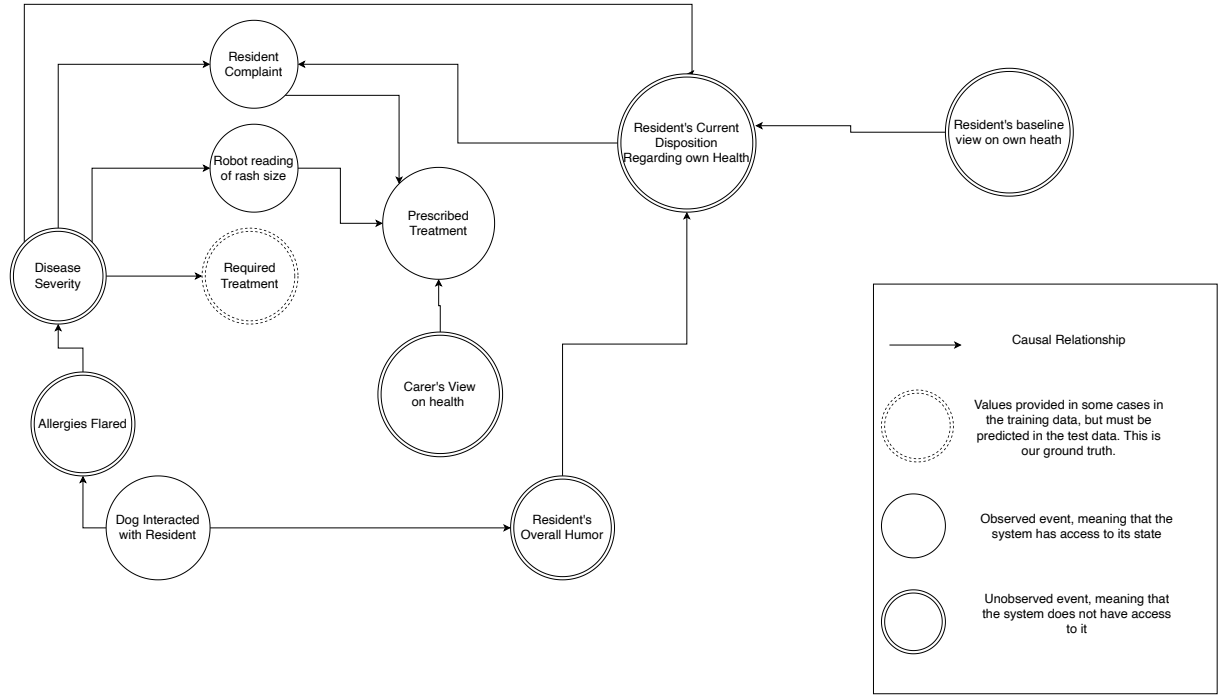


Figure 2: Case overview

factors are also considered. For example, pets may visit. This can include a dog visiting the smart home. The dog at time t is a binary variable Dog_t . If the dog came, it will interact with all residents. For those selected R_i that are $AllergicR_i$, their R_t^{Derm} will increase by some value sampled from another distribution. Otherwise R_t^{Derm} is determined by the natural evolution of the skin malady and the effects of the treatment. The types of treatment are the 3 mentioned earlier. Thus, $PredTreatment_t$ is one of the set and the same goes for $RequiredTreatment_t$. Regarding levels of complaint and robot readings, those can also be in the range $[0, 15]$.

In order to generate data we will randomly start the disease from time to time in some residents even without dog interaction. This is in accordance with skin maladies being more likely at some times: for example, during cold dry winter months. We begin designing the AML-based algorithm with all variables being discrete with ranges similar to those described before. The relationships are as one would expect in Bayesian networks, where some common sense relationships hold for this scenario, such as:

- Dermatitis severity increases size of the rash as detected by the robot, but its readings have a margin of error.
- Resident's complaints are higher the higher the symptoms and the worse own view on his health.
- Level of prescribed medication is more severe the worse the robot readings and complaints and is decreased or increased based on the view of health of the carer involved.
- An improvement in overall mood improves the overall health disposition.

and so on. The idea is that the relationships are not deterministic (dog increases allergy by 3), but stochastic and interact roughly as one would expect in this scenario. This would involve scans, analyses of scans, identification of patterns in relation to medication already given and pet visits.

Questions for the AI:

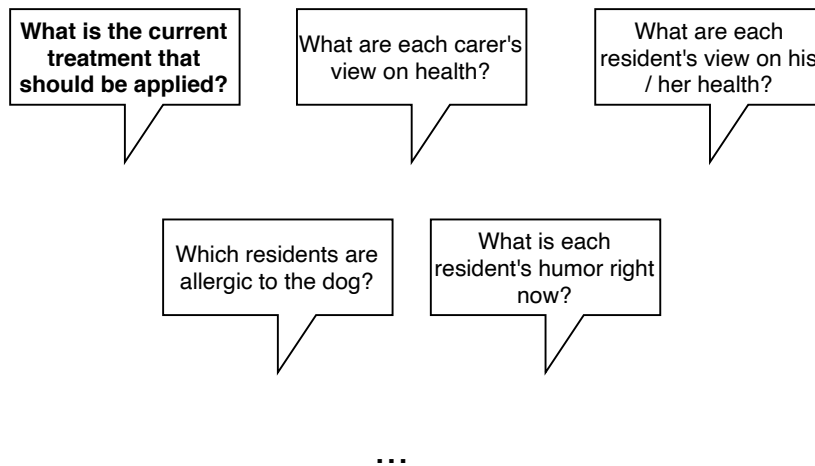


Figure 3: Example of questions we want the AI to answer in this scenario. In bold the main thing we want to predict in it, as it determines the actions of the AI following the care rules we will describe in Section 2.

3.4 Preliminary Ethical Framework

In the case, an AML robot could provide ethical support to human carers who treat the skin maladies of residents in a care home.

Also, the AML-based system applies relevant ethical work procedures by presenting scan analyses and consequent skin care medication advice to the human carer. Here, unethical human action would be to provide too little medication or too much medication. The preliminary ontology we described in Section 2 is related to the case in the subsequent Figure 4.

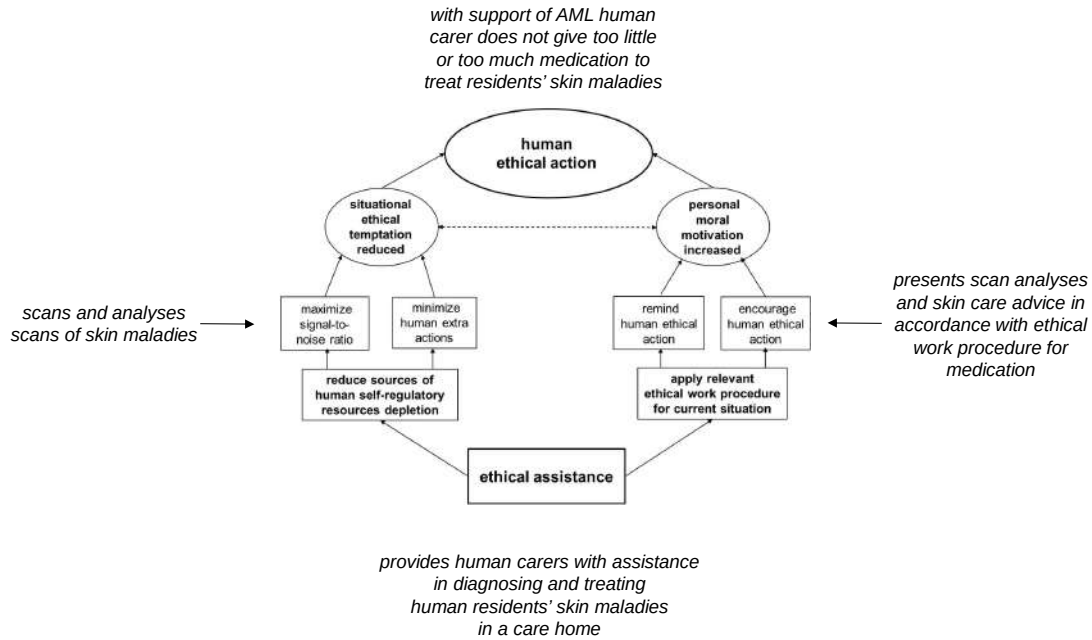


Figure 4: Case related to our preliminary Ontology

3.4.1 AML based AI Ethical Assistance Interaction

A summary of potential interactions between resident's attitude and carers' view is shown in Figure 5 below. In this case, ethical action is action of providing the correct amount of medication for the skin malady of dermatitis.

Conversely, unethical action would be to provide too little medication or too much medication. In particular, the provision of too little medication or too much medication to many residents over a long period of time. Here, too little or too much medication being medication that is inconsistent with symptom patterns of dermatitis. This need not be deliberate under treatment or over treatment. Rather, as is found in behavioral ethics research, the human carer would not have any deliberate intention to take unethical action. Instead, failings could arise from the difficulty of human carers maintaining in-depth up-to-date knowledge of each resident's symptom trajectory. Also, action that can be considered to be unethical can arise from depletion. For example, when human carers have a wide variety of different types of tasks to carry out. Accordingly, it is the care home's ethical responsibility to design and maintain a system within which human carers are not lacking in up-to-date information of each resident's symptom trajectory, and where human carers are not depleted from having too many tasks to carry out. Potential for unintended under treatment and over treatment can be increased if an organization has many residential care homes at many locations. Potential can be further increased if erratic treatment becomes the norm in many organizations' residential care homes. Research indicates that increasing the ratio of skilled caring staff can reduce the potential for unethical action [35, 36], and this is a role

that AML-based algorithms have the potential to support with by undertaking for example the scanning and analyses of dermatitis rashes. Here, it is important to note that introduction of objective functions such as scanning and analyses by AI assistants should not prevent attention being paid to residents' and carers' subjective concerns. Not least because residents and carers can have intuitions that go beyond what can be discovered from automated analyses of scans, etc. Overall, it is the organization's ethical responsibility to design and maintain a system that facilitates the correct use of medication. In particular, a system that minimizes potential for use of too little medication or too much medication. Here, this is done through introduction of AML to support human carers.

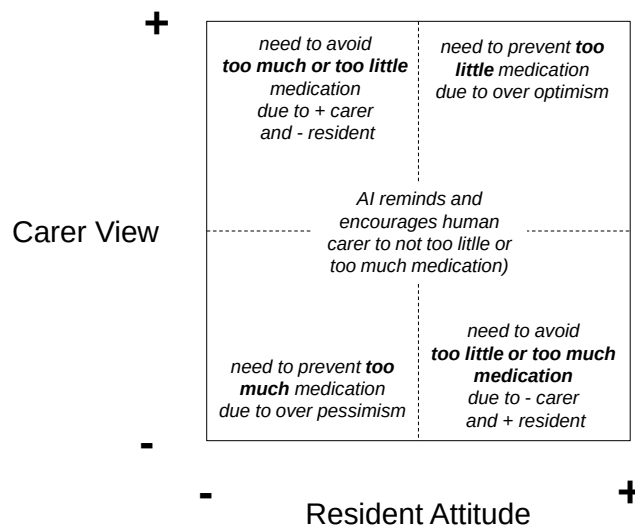


Figure 5: Interactions between Carers and Residents

This leads to the desired interaction options of a preliminary algorithm, which is consistent with the work procedures developed through four levels of human oversight within QMS, see Algorithm 1, where both the resident's attitude and carers' view is taken into account and ethical action is cued so that the correct level of medication is applied.

Algorithm 1: Interaction options

```

if resident is (-) then
  if carer is (-) then
    | cue always both for less medication, strongly if images do not match resident's descriptions
  else if carer is (+) then
    | cue for less medication when images do not match resident's descriptions.
else if resident is (+) then
  if carer is (-) then
    | cue for more medication when images do not match resident's descriptions.
  else if carer is (+) then
    | cue always for more medication when images do not match resident's descriptions.

```

3.4.2 Implications for human - AI interaction research

In ethics literature, artificial intelligence (AI) has been a source of anxiety for decades [37]. For example, there are concerns that AI may apply the biases of its developers and/or the biases of the data that it has been trained on [38]. A more general concern is that AI cannot be transparent because it is too complex to be understood [39]. Current concerns take place amidst fears about a future in which AI will become unaccountable to humans. In particular, it is argued that the capacity of AI to learn will lead to situations where the humans who designed and/or deployed AI may not be able to control or even predict what the AI will do [40]. However, it is explained in the preceding sections of this deliverable that AI has the potential to support human ethical action, and AML can be instrumental in ensuring that it is transparent in doing so.

This is a new direction for applied research that is aligned with basic research that relates the potential of stress overload to undermine ethical behavior in both humans and robots. In particular, the potential for the regulatory capacity of a control system to be overtaxed by environmental perturbations to the extent that regulatory work increases to the point where energy demand exceeds energy supply. As pointed out by [1], the prospect of ‘stressed robots’ that weigh selfish and selfless goals may not seem very appealing, but may ultimately prove to be of significant value. For instance, AML robots should be programmed to never abandon higher level ethical goals over lower level expediency goals. Clearly, this will be a formidable challenge needing extensive further research.

3.4.3 Implications for human - AI interaction practice

As explained above, AML-based AI could support human ethical action through referring to and by presenting work procedures. To ensure human oversight of work procedures, they can be incorporated into organizations quality management systems (QMS). These provide an established basis for the formalization of work procedures, including those with ethical implications [41].

QMS has been successful in improving the performance of many organizations [42]. Importantly, QMS procedures and adherence to QMS procedures are verified by external bodies in accordance with regulatory standards [43]. QMS are recognized in the ethics literature as a means of incorporating ethical considerations into company procedures. For example, [44] proposed procedures for incorporating corporate social responsibility into QMS. Subsequent research indicates that QMS practices may facilitate the development of social responsibility [45, 46]. Crucially, as summarized in Figure 6 below, QMS work procedures are updated by human QMS management review. In particular, existing work procedures should be revised and/or new work procedures should be added to deal with complex work situations.

Management reviews make reference to recent patterns of non-conformance. This is done to update the QMS by better aligning it with the organization’s current situation, and so provide the basis for preventing future non-conformance between ethical practice, organization’s goals and **human needs**. Thus, incorporation of AML-based algorithms into QMS could provide human oversight and further increase the transparency.

3.5 AML Embedding

Let’s then start describing our scenario using AML. The objective of the AI system is to predict what is the required medication treatment and what is the one that a specific carer will provide. By detecting mismatches between both the carer can be reminded / encouraged to take ethical action avoiding under or over medication. We will consider progressively more complex embeddings applied to this scenario.

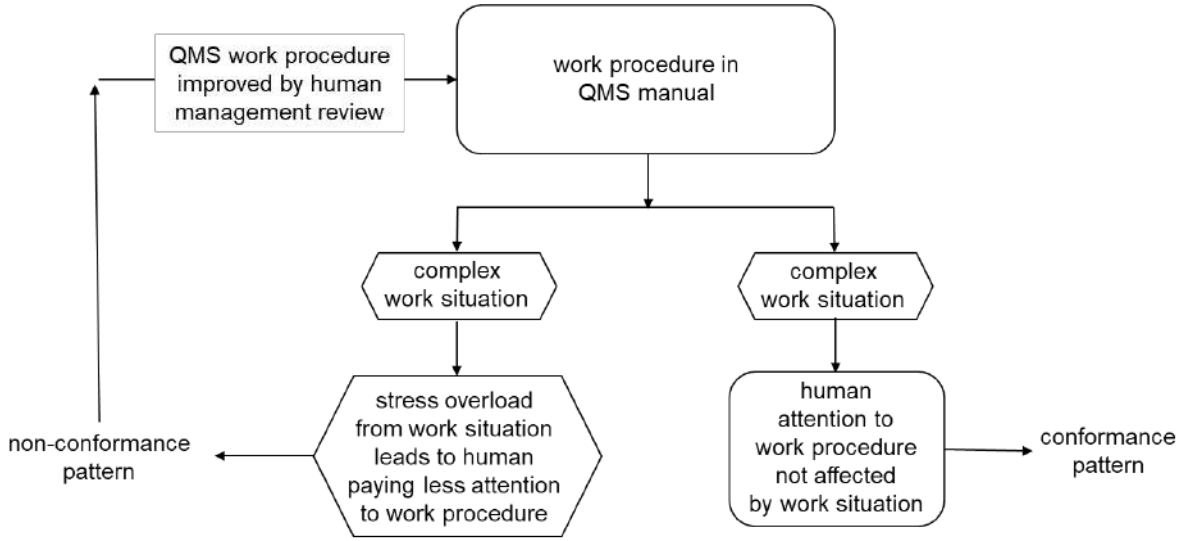


Figure 6: Human oversight of work procedures through QMS management review

3.5.1 Simple Embedding

A simple embedding of the scenario is to represent each medication decision independently using as constants

- C_j representing one of the C carers.
- m representing the required type of medication out of the 3 different types available.
- r representing the robot's estimation of the skin rash size.
- c representing the complaint severity of the resident.

As a carer C_j decides m based on r and c , we have

$$m \leq r \odot c \odot C_j \quad (1)$$

Using past information on how the carers provided medication, the algebra will thus encode how each carer responds to the evidence available. This may or may not correspond to the ideal medication. In order to encode that we can create one constant C_{ideal} to represent the ideal medication level. This means representing in the algebra the examples of correct medication we have as

$$m \leq r \odot c \odot C_{ideal} \quad (2)$$

This does not include all information from this scenario. For example, we also have visits from the dog and other variables related to residents. Those include whether or not they are allergic to the dog and also the variables related to his attitude towards health. Those can be included using

- R constants with R_i representing each one of the R residents.
- Two constants, d_0 and d_1 , representing whether or not the dog visited that day, respectively.

Thus we update our embedding for any carer C_j (ideal or not) making treatment decisions for a resident R_i also including the dog status d as

$$m \leq r \odot c \odot C_j \odot R_i \odot d \quad (3)$$

In this example the allergies and nature of carers and residents are not explicitly modeled as constants. Instead, the algebra will *learn those patterns* as it learns the relationships between the variables as the atoms are constructed.

We can translate this embedding into AML-DL. For the case with 7 residents and 3 carers, the variables are:

```
1 treatment = CV("treatment", 3)
2 robot_readings = CV("robot_readings", 16)
3 complaint_levels = CV("complaint_levels", 16)
4 carers = CV("carers", 3)
5 ideal_carer = C("ideal_carer")
6 residents = CV("residents", 7)
7 has_dog = CV("dog", 2)
```

and the embedding is realised by adding the inclusions.

```
1 for treatment_i, robot_i, compl_i, car_i, res_i, d_i in all_past_interactions:
2     treatment_given = F("treatment", treatment_i)
3     day_info = M(F("robot_readings", robot_i), F("complaint_levels", compl_i),
4                 F("carers", car_i), F("residents", res_i), F("has_dog", d_i))
5     INCL(treatment_given, day_info)
6     for t_o in range(3):
7         if t_o != treatment_i:
8             EXCL(F("treatment", t_o), day_info)
```

For those days where we know what the recommended medication would have been, we add inclusions in the form

```
1 for ideal_treatment_i, robot_i, compl_i, res_i, d_i in correct_interactions:
2     treatment_ideal = F("treatment", ideal_treatment_i)
3     day_info = M(F("robot_readings", robot_i), F("complaint_levels", compl_i),
4                 F("ideal_carer", ideal_carer), F("residents", res_i), F("has_dog", d_i))
5     INCL(treatment_ideal, day_info)
6     for t_o in range(3):
7         if t_o != ideal_treatment_i:
8             EXCL(F("treatment", t_o), day_info)
```

and let the algebra learn this embedding. After it has learned it, we will have for any new interaction, we can gather the relevant constants and build terms in the form

```
1 v_now = M(robot_reading_now, complaint_level_now, current_carer, current_resident,
2           has_dog_now)
```

to estimate what will be the medication a specific carer would provide in this situation. This can be done by for each medication treatment analysing how many of its atoms are included in `v_now`. The same can be done for the ideal carer using

```
1 v_ideal = M(robot_reading_now, complaint_level_now, ideal_carer, current_resident,
2            has_dog_now)
```

giving us the level of medication that is recommended in the current case. With this information, the AI can remind / encourage the carer to provide the appropriate level of care if it detects over or under treatment.

We can also include domain information about the problem in our embedding. Let's show how by constructing an example of how we could extend our simple embedding to model the fact that the complaint level of a resident at a particular time is at one and only one level. In other words, a specific

case cannot contain both the highest and lowest complain level. More formally, if the current complaint level is represented by the constant c_1 it cannot be also c_2 . More formally

$$c_i \not\leq \odot_{i \neq j} c_j \quad (4)$$

As we can see this knowledge did not add any constant to our embedding, just exclusions. This can be represented in AML-DL by adding the exclusivity information as

```

1 for i in range(16):
2     ci = F("complaint_levels", ci)
3     aux = None
4     for j in range(16):
5         if i != j:
6             cj = F("complaint_levels", j)
7             if aux is None:
8                 aux = cj
9             else:
10                aux = M(aux, cj)
11 EXCL(ci, aux)

```

The same idea can be applied to the robot's reading levels, but for brevity we will omit the math and code, as it follows the same pattern of this one. With this example we have shown how one can include simple domain knowledge in the embedding. Another way we could include domain information is to add the notion of ordering to the complaints and robot readings, in a similar way as done in deliverable 5.1, where we have an embedding for features are represented by numbers. Those ideas will be explored in further versions, but it suffices to say that even when creating an embedding for individual situations without time dynamics, we can include rich relationships that, together with the information available in the training data, will be learned by the algebra.

3.5.2 Embedding temporal dynamics

Up to now each instant is a single decision and thus our embedding does not directly encode the temporal relationship between the variables. This is very important in this scenario, as the skin malady evolves through time according to its natural schedule influenced by other variables (treatment, dog visits, etc). Thus, we will incorporate time in our simple embedding, again in increasing levels of complexity. The simplest way to do it is to include in each medication treatment decision information regarding the past context of the variables. For example, instead of only using one set of constants for the current complaints of the resident, we can include in the embedding also previous values $c_0, c_{-1}, c_{-2} \dots c_{-w}$ where w is the size of our look back window. The embedding would then become

$$m \leq r_0 \odot r_{-1} \odot \dots r_{-w} \odot c_0 \odot c_{-1} \odot \dots c_{-w} \odot C_j \odot R_i \odot d_0 \odot d_{-1} \odot \dots d_{-w} \quad (5)$$

Using the shorthand $\odot_{0..-w}^i v_i$ for $v_a \odot \dots v_b$ we can simplify the embedding as

$$m \leq (\odot_{0..-w}^i r_i) \odot (\odot_{0..-w}^k c_k) \odot C_j \odot R_i \odot (\odot_{0..-w}^l d_l) \quad (6)$$

The constant definitions from our previous example can then be represented by the AML-DL code

```

1 treatment = CV("treatment", 3)
2 carers = CV("carers", 3)
3 residents = CV("residents", 7)
4 ideal_carer = C("ideal_carer")
5
6 for time_i in range(w):
7     CV("robot_readings_back" + str(time_i), 16)
8     CV("complaint_levels_back" + str(time_i), 16)
9     CV("has_dog_back" + str(time_i), 2)

```

and the embedding itself can be done by creating those inclusions of Equation 3.5.2 by iterating over the temporal events in temporal order from the resident's perspective using

```

1  for res_i in range(7):
2      past_interactions = get_ordered_interactions_for_resident(res_i)
3      for t_now in range(len(past_interactions)):
4          look_back_window = past_interactions[max(0, t_now - w):t_now+1]
5          ti = w
6          aux = F("residents",res_i)
7          for treatment_i, robot_i, compl_i, car_i, res_i, d_i in look_back_window:
8              aux = M(aux,temporal_info,F("robot_readings_back" + str(ti) robot_i), F("
9              complaint_levels_back" + str(ti), compl_i), F("has_dog_back" + str(ti), d_i))
10
11             # If we are reading the end of the window.
12             if ti == 0:
13                 # Get information about current time.
14                 treatment_given_i = treatment_i
15                 current_carer_i = car_i
16                 ti -= 1
17             temporal_info = M(aux, F("carers",current_carer_i))
18             INCL(F("treatment", treatment_given_i), temporal_info)
19             for t_o in range(3):
20                 if t_o != treatment_given_i:
21                     EXCL(F("treatment", t_o), temporal_info)

```

On the other hand, this simple embedding includes many constants. For each temporal variable with p possible states we need to generate $w * p$ constants. Moreover, we have not yet added explicit temporal relationships between the variables. One way we can do it is to also enforce

$$r_0 \odot c_0 \leq (\odot_{-1..-w}^i r_i) \odot (\odot_{-1..-w}^k c_k) \odot C_j \odot R_i \odot (\odot_{-1..-w}^l d_l) \odot (\odot_{-1..-w}^t m_t) \quad (7)$$

which forces the creation of an internal model for r and c . This means that the present state should be predictable from previous history, that is, the algebra needs to understand that a particular sequence of facts is equivalent (or what has led) to the current state of affairs. This would cost us constants for the previous treatments applied, that is $w * 3$ constants more and also building the inclusion of Equation 3.5.2, which we will leave out for brevity.

We are now modeling implicitly modeling state. This state can also be defined explicitly so that we have a direct representation of the underlying (unknown to the system) ground truth state of the disease. While we do not have direct access to it, we can learn about it by its causes and consequences, as depicted in Figure 2. As of yet the algebra has not modeled directly this disease state. One way to do it is to add N constants s_n representing each of the possible states of the dermatitis. At the current time, the skin malady has to be at one of those states, as they represent values from no skin malady to its highest severity. If $\hat{s}$ contains the current state of the skin malady, then

$$\hat{s} \leq s_a \odot s_b \odot s_n \quad (8)$$

as it is one of the possible ones. Moreover, it is one and only one and therefore

$$s_i \odot s_j \not\leq \hat{s} \quad (9)$$

As we want $\hat{s}$ to represent the current state of affairs, it must be compatible with the latest rash reading and resident's complaint level, as well as the current dog situation. Formally,

$$r_0 \odot c_0 \odot d_0 \leq \hat{s} \quad (10)$$

meaning that we can change Equation 3.5.2 to

$$\hat{s} \leq (\odot_{-1..-w}^i r_i) \odot (\odot_{-1..-w}^k c_k) \odot C_j \odot R_i \odot (\odot_{-1..-w}^l d_l) \odot (\odot_{-1..-w}^t m_t) \quad (11)$$

Of course, we would still use Equation 3.5.2. This means that our explicit representation of skin malady state needs only N constants, as we can destroy the previous $\hat{s}$ when we move to a new example.

But we can go further. Up to now, from the point of view of the algebra, there is no explicit link between values at time t (let's call that $r(t)$) and at time $t - 1$ ($r(t - 1)$) inside the same context window. As we predict the new time-step, what was $r(t)$ is now $r(t - 1)$. Thus we could reduce the number of constants and directly embed this idea of a sliding window. Instead of creating for a variable that has p possible states in a w window $p * w$ constants, we can create w constants $v_0, v_{-1}, \dots, v_{-w}$ representing a time and p constants $p_0 \dots p_p$ representing a value and link time with value through equations

$$\begin{cases} \delta \odot v_0 = r(t) \odot \delta \\ \delta \odot v_{-1} = r(t - 1) \odot \delta \\ \dots \\ \delta \odot v_{-w} = r(t - w) \odot \delta \end{cases} \quad (12)$$

where $r(t)$ is replaced by the p constant that holds at time t for the variable v . This constant δ links all variables to their current values. To link every value to its previous one, and thus explicitly tell the algebra that one follows the other, we create further equations linking each v_t with its value at the previous instant using a new constant ϵ and equations

$$\begin{cases} \epsilon \odot v_0 = r(t - 1) \odot \epsilon \\ \epsilon \odot v_{-1} = r(t - 2) \odot \epsilon \\ \dots \\ \epsilon \odot v_{-w-1} = r(-w) \odot \epsilon \end{cases} \quad (13)$$

If we continue this process we will create w such linking constants. Since those can be shared between variables this leaves us with a smaller number of constants as we had before $w * p_v * N_v$ constants in total for N_v variables and now each variable uses $p + w$ constants and therefore we have $w + N_v * (p + w)$ constants in total. This comes at the cost of a bigger set of equations in the embedding, $\frac{w+w^2}{2}$ to be exact. On the other hand those equations represent for the algebra the idea of a sliding window and, even more important, they are only present at training time since for our classification task the inclusion looks very similar to before, as we are using

$$\hat{s} \leq (\odot_{-1..-w}^i r_i) \odot (\odot_{-1..-w}^k c_k) \odot C_j \odot R_i \odot (\odot_{-1..-w}^l d_l) \odot (\odot_{-1..-w}^t m_t) \odot \delta \quad (14)$$

since we are working with results for the current state (and therefore only dealing with the δ variable).

4 Conclusion

In this deliverable, *Deliverable D5.4 - AML based world models* Initial, we have made a start on exploring the potential of AML to support ethical world models. As ALMA is a Future & Emerging Technologies (FET) project, we have adhered to the FET aims of novel multidisciplinary frontier research that lays the foundations for next generation technologies. In particular, we have explored the potential for AML-based algorithms to support human ethical behavior. This addresses ethical issues in human interaction with AI in a new human-centric way. It is a new approach that is based on the latest research in behavioral ethics. At the same time, we have recognized that FET research should have the potential to convert proofs of concept into industrial practice. Accordingly, in this deliverable, we have related AML-algorithms to established practice. We have done this by relating AML-based features to quality management systems (QMS), that connect the everyday practices of organizations to the relevant laws, regulations, standards, etc., which provide organizations with ethical frameworks.

Specifically, based on recent research into behavioural ethics, in this deliverable we have defined a preliminary framework for supporting human ethical action using AI. We have also described an exploratory case scenario and showed how the case scenario can be represented both in our preliminary ontology and then directly as Algebraic Machine Learning (AML) code. In the exploratory case scenario, an AML-based algorithm provides support for human ethical action based upon work procedures, which have multiple levels of human oversight. Overall, this deliverable provides a start for exploring how everyday ethical practices can be improved through reference to behavioural ethics research and implementations of Algebraic Machine Learning. In future Work Package 5 deliverables, we shall report our work in which we further explore the extent to which one type of mathematics can successfully formalize the potential complexity of ethical interactions involving different types of humans and robots. We anticipate that through the course of the ALMA project, we shall find where AML has most potential, and where combination with other types of models are most beneficial. This could involve exploring directions other than that reported in this first deliverable.

References

- [1] Rutger Goekoop and Roy de Kleijn. “How higher goals are constructed and collapse under stress: A hierarchical Bayesian control systems perspective”. In: *Neuroscience & Biobehavioral Reviews* 123 (2021), pp. 257–285. ISSN: 0149-7634. DOI: <https://doi.org/10.1016/j.neubiorev.2020.12.021>. URL: <https://www.sciencedirect.com/science/article/pii/S0149763420306965>.
- [2] Farid F. Youssef et al. “Stress alters personal moral decision making”. In: *Psychoneuroendocrinology* 37.4 (2012), pp. 491–498. ISSN: 0306-4530. DOI: <https://doi.org/10.1016/j.psyneuen.2011.07.017>. URL: <https://www.sciencedirect.com/science/article/pii/S0306453011002320>.
- [3] David De Cremer and Wim Vandekerckhove. “Managing unethical behavior in organizations: The need for a behavioral business ethics approach”. In: *Journal of Management & Organization* 23.3 (2017), pp. 437–455.
- [4] Francesca Gino et al. “Unable to resist temptation: How self-control depletion promotes unethical behavior”. In: *Organizational behavior and human decision processes* 115.2 (2011), pp. 191–203.
- [5] Linda Klebe Trevino. “Ethical decision making in organizations: A person-situation interactionist model”. In: *Academy of management Review* 11.3 (1986), pp. 601–617.
- [6] Muel Kaptein. “The battle for business ethics: A struggle theory”. In: *Journal of Business Ethics* 144.2 (2017), pp. 343–361.
- [7] Gary R Weaver and Cynthia E Clark. “Behavioral ethics, behavioral governance, and corruption in and by organizations”. In: *Debates of corruption and integrity*. Springer, 2015, pp. 135–158.
- [8] David De Cremer et al. “Understanding ethical behavior and decision making in management: A behavioural business ethics approach”. In: *British Journal of Management* 22 (2011), S1–S4.
- [9] Mo Chen, Chao C Chen, and Oliver J Sheldon. “Relaxing moral reasoning to win: How organizational identification relates to unethical pro-organizational behavior.” In: *Journal of Applied Psychology* 101.8 (2016), p. 1082.
- [10] Elizabeth E Umphress, John B Bingham, and Marie S Mitchell. “Unethical behavior in the name of the company: the moderating effect of organizational identification and positive reciprocity beliefs on unethical pro-organizational behavior.” In: *Journal of applied psychology* 95.4 (2010), p. 769.
- [11] Carolin Baur et al. “Unable to resist the temptation to tell the truth or to lie for the organization? Identification makes the difference”. In: *Journal of Business Ethics* (2019), pp. 1–20.
- [12] Lorenzo Avanzi et al. “The downside of organizational identification: Relations between identification, workaholism and well-being”. In: *Work & Stress* 26.3 (2012), pp. 289–307.
- [13] Blake E Ashforth et al. “Re-viewing organizational corruption”. In: *Academy of Management review* 33.3 (2008), pp. 670–684.
- [14] Eun-Ju Lee and Jin Ho Yun. “Moral incompetency under time constraint”. In: *Journal of Business Research* 99 (2019), pp. 438–445.
- [15] Shaul Shalvi, Ori Eldar, and Yoella Bereby-Meyer. “Honesty requires time (and lack of justifications)”. In: *Psychological science* 23.10 (2012), pp. 1264–1270.
- [16] Yan Wang et al. “Depletion, moral identity, and unethical behavior: Why people behave unethically after self-control exertion”. In: *Consciousness and cognition* 56 (2017), pp. 188–198.
- [17] Sahil Sholla, Roohie Naaz Mir, and Mohammad Ahsan Chishti. “Towards the design of ethics aware systems for the Internet of Things”. In: *China Communications* 17.2 (2020), pp. 239–252.
- [18] Samuel PL Veissière et al. “Thinking through other minds: A variational approach to cognition and culture”. In: *Behavioral and Brain Sciences* 43 (2020).
- [19] Colin Allen, Iva Smit, and Wendell Wallach. “Artificial morality: Top-down, bottom-up, and hybrid approaches”. In: *Ethics and information technology* 7.3 (2005), pp. 149–155.
- [20] José-Antonio Cervantes et al. “Artificial moral agents: A survey of the current status”. In: *Science and engineering ethics* 26.2 (2020), pp. 501–532.
- [21] Alexander Button et al. “Automated de novo molecular design by hybrid machine intelligence and rule-driven chemical synthesis”. In: *Nature machine intelligence* 1.7 (2019), pp. 307–315.

- [22] Junjie Wang et al. "Lane change decision-making through deep reinforcement learning with rule-based constraints". In: *2019 International Joint Conference on Neural Networks (IJCNN)*. IEEE. 2019, pp. 1–6.
- [23] Fernando Martin-Maroto and Gonzalo G de Polavieja. "Algebraic Machine Learning". In: *arXiv preprint arXiv:1803.05252* (2018).
- [24] Jana Jost et al. "Examining the effects of height, velocity and emotional representation of a social transport robot and human factors in human-robot collaboration". In: *IFIP Conference on Human-Computer Interaction*. Springer. 2019, pp. 517–526.
- [25] Kathleen D Vohs et al. "Making choices impairs subsequent self-control: A limited-resource account of decision making, self-regulation, and active initiative". In: *Self-regulation and self-control*. Routledge, 2018, pp. 45–77.
- [26] Evan Polman and Kathleen D Vohs. "Decision fatigue, choosing for others, and self-construal". In: *Social Psychological and Personality Science* 7.5 (2016), pp. 471–478.
- [27] Marcus Yung et al. "Examining the fatigue-quality relationship in manufacturing". In: *Applied ergonomics* 82 (2020), p. 102919.
- [28] Guido Alessandri et al. "A multilevel moderated mediational model of the daily relationships between hassles, exhaustion, ego-resiliency and resulting emotional inertia". In: *Journal of Research in Personality* 85 (2020), p. 103913.
- [29] Gary M Fleischman et al. "Ethics versus outcomes: Managerial responses to incentive-driven and goal-induced employee behavior". In: *Journal of Business Ethics* 158.4 (2019), pp. 951–967.
- [30] Steven L Grover and Chun Hui. "How job pressures and extrinsic rewards affect lying behavior". In: *International Journal of Conflict Management* (2005).
- [31] Jacob B Hirsh, Jackson G Lu, and Adam D Galinsky. "Moral utility theory: Understanding the motivation to behave (un) ethically". In: *Research in Organizational Behavior* 38 (2018), pp. 43–59.
- [32] Katie E Garrison, Anna J Finley, and Brandon J Schmeichel. "Ego depletion reduces attention control: Evidence from two high-powered preregistered experiments". In: *Personality and Social Psychology Bulletin* 45.5 (2019), pp. 728–739.
- [33] Rogelio Oliva and John D Sterman. "Cutting corners and working overtime: Quality erosion in the service industry". In: *Management Science* 47.7 (2001), pp. 894–914.
- [34] Robert Allan Cooke. "Danger signs of unethical behavior: How to determine if your firm is at ethical risk". In: *Journal of Business Ethics* 10.4 (1991), pp. 249–253.
- [35] R Tamara Konetzka et al. "Malpractice litigation and nursing home quality of care". In: *Health services research* 48.6pt1 (2013), pp. 1920–1938.
- [36] Mei Zhao and D Rob Haley. "Nursing home quality, staffing, and malpractice paid-losses". In: *Journal of Health Care Finance* 38.1 (2011), p. 10.
- [37] Omar EM Khalil. "Artificial decision-making and artificial ethics: A management concern". In: *Journal of Business Ethics* 12.4 (1993), pp. 313–321.
- [38] Ulrich Leicht-Deobald et al. "The challenges of algorithm-based HR decision-making for personal integrity". In: *Journal of Business Ethics* 160.2 (2019), pp. 377–392.
- [39] Ivy Munoko, Helen L Brown-Liburd, and Miklos Vasarhelyi. "The ethical implications of using artificial intelligence in auditing". In: *Journal of Business Ethics* (2020), pp. 1–26.
- [40] Deborah G Johnson. "Technology with no human responsibility?" In: *Journal of Business Ethics* 127.4 (2015), pp. 707–715.
- [41] Pervaiz K Ahmed and Silke Machold. "The quality and ethics connection: toward virtuous organizations". In: *Total Quality Management & Business Excellence* 15.4 (2004), pp. 527–545.
- [42] Michael Bell and Vincent Omachonu. "Quality system implementation process for business success". In: *International Journal of Quality & Reliability Management* 28.7 (2011), pp. 723–734.
- [43] Jay J Schlickman. *ISO 9001: 2000 quality management system design*. Artech House, 2003.
- [44] Ton Van der Wiele et al. "A corporate social responsibility audit within a quality management framework". In: *Journal of Business Ethics* 31.4 (2001), pp. 285–297.

- [45] Muhammad Adnan Khurshid, Muslim Amin, and Wan Khairuzzaman Wan Ismail. “Total quality and socially responsible management (TQSR-M): an integrated conceptual framework”. In: *Benchmarking: An International Journal* (2018).
- [46] Juan José Tarí. “Research into quality management and social responsibility”. In: *Journal of Business Ethics* 102.4 (2011), pp. 623–638.