|  |  |
| --- | --- |
|  | Moving Picture, Audio and Data Coding by Artificial Intelligencewww.mpai.community |

|  |  |
| --- | --- |
|  | 2020/12/14 |
| Source | Mariangela Rosano, Valeria Lazzaroli, Guido Perboli |
| Title | MPAI-CAE Functional Requirements work programme |
| Target | MPAI Members |

# Introduction

Moving Picture, Audio and Data Coding by Artificial Intelligence (MPAI) is an [international association](http://mpai.community/) with the mission to develop *AI-enabled data coding standards*. Research has shown that data coding with AI-based technologies is *more efficient* than with existing technol­ogies.

The MPAI approach to AI data coding standards is by defining *AI Modules (AIM)* with standard interfaces that are combined and executed within an MPAI-specified AI-Framework. With its standards, MPAI intends to promote the development of *horizontal markets* of *competing* *proprietary* solutions with standard interfaces tapping from and further promoting AI *innovation.*

This paper describes the current MPAI plan to develop “Compression and understanding of industrial data” (MPAI-CUI), an MPAI area of work that uses AI substantially to reduce the amount of data with a controlled loss of information and extract the most relevant information from the industrial data, with the aim of assessing company performance and predicting the risk of bankruptcy long before.

Chapter 2 explains the MPAI-CUI features, Chapter 3 provides summary information on the advanced IT environment that will execute MPAI- CUI applications and Chapter 4 identifies the items that will likely be the object of the MPAI- CUI standard.

# MPAI-CUI features

Most economic organizations, e.g., companies, etc., produce large quantities of data, often because these are required by regulation. Users of these data maybe the company itself or Fintech and Insurtech services who need to access the flow of company data to assess and mon­itor financial and organizational performance, as well as the impact of vertical risks (e.g., cyber, seismic, etc.). For example, nowadays companies heavily rely on the security and dependability of their Information System for all the categories of workers, including the management of Industrial Control Systems. Adding into the risk analysis process cybersecurity-related parameters will help a more precise estimation of the actual risk exposure. Cybersecurity data will help a reassessment of financial parameters based on risk analysis data.

The sheer amount of data that need to be exchanged is an issue. Analysing those data by humans is typically on­erous and may miss vitally important information. Artificial intelligence (AI) may help reduce the amount of data with a controlled loss of information and extract the most relevant information from the data. AI is considered the most promising means to achieve the goal.

Unfortunately, the syntax and semantics of the flow of data is high dependent on who has produced the data. The format of the date is typically a text file with a structure not designed for indexing, search and ex­traction. Therefore, in order to be able to apply AI technologies to meaningfully reduce the data flow, it is necessary to standardize the formats of the components of the data flow and make the data “AI friendly”.

So far, the AIMs required by the following application area have been considered for possible standardisation by MPAI-CUI:

1. Compression and Understanding of Industrial Data (see 4.1)

# AI Framework

Most MPAI applications considered so far can be implemented as a set of AIMs – AI/ML and even traditional data processing based units with standard interfaces assembled in suitable topologies to achieve the specific goal of an application and executed in an MPAI-defined AI Framework. MPAI is making all efforts to iden­tify processing modules that are re-usable and upgradable without necessarily changing the inside logic.

MPAI plans on completing the development of a 1st generation AI Framework called MPAI-AIF in July 2021.

The MPAI-AIF Architecture is given by *Figure 1*



*Figure 1 –The MPAI-AIF Architecture*

Where

1. *Management and Control* manages and controls the AIMs, so that they execute in the correct order and at the time when they are needed.
2. *Execution* is the environment in which combinations of AIMs operate. It receives external inputs and produces the requested outputs both of which are application specific interfacing with Management and Control and with Communication, Storage and Access.
3. *AI Modules* (AIM) are the basic processing elements receiving processing specific inputs and producing processing specific
4. *Communication* is required in several cases and can be implemented, e.g. by means of a service bus and may be used to connect with remote parts of the framework
5. *Storage* encompasses traditional storage and is used to e.g. store the inputs and outputs of the individual AIMs, data from the AIM’s state and intermediary results, shared data among AIMs.
6. *Access* represents the access to static or slowly changing data that are required by the application such as domain knowledge data, data models, etc.

# MPAI-CUI work plan

# In this chapter there is currently one application area with its relevant AI Modules (AIM) identified and described, and their inputs/outputs summarily specified.

## Data Compression and Understanding

### Use Case description

# A company may need to access the flow of internal (i.e., financial and governance data) and exter­nal data to assess and mon­itor its financial and organizational performance, as well as the impact of vertical risks (e.g., cyber, seismic, etc.), according to the current regulations (e.g., ISO 31000 on risk assessment and management).

# The company generating the data flow may need to perform compression and understanding for its own needs (e.g., to identify core and non-core data). Indeed, the company itself can analyse its financial performance, identifying possible clues to the crisis or risk of bankruptcy years in advance. It may help the board of directors and decision-makers to make the proper decisions to avoid these situations, conduct what-if analysis, and devise efficient strategies.

# At the same time, a financial institution that receives a request for financial help from a troubled company, can access its financial and organizational data and make an AI-based assessment of that company, as well as a prediction of future performance. This aids the financial institution to take the right decision in funding or not that company, having a broad vision of its situation.

# The AI Framework for this Use Case is given by the following *Figure 2*.

**

*Figure 2 – Compression and understanding of Industrial Data*

### AI modules

The list of the AI modules is given by the *Table 1* below:

*Table 1 – AI Module interaction*

|  |  |  |  |
| --- | --- | --- | --- |
| **AI Module** | **Input** | **Output** | **External data** |
| Financial assessment (FA)  | Financial statement data | Financial features | Standards from knowledgebase |
| Governance assessment (GA) | Governance data | Governance features |  |
| Risk matrix (BMR) | Technical data from BIM, internal assessment on cyber security | Severity  | Socio-economic data from data bases Technical data from KBStandards from KB |
| Decision tree (DT) | Financial features, Governance features | Ranking of features importance  | Data on active and failed companies from back testing |
| Prediction (PRF) | Financial features, Governance features | Probability of company crisisAdequacy of organizational model (indexes) | Data on active and failed companies from back testing |
| Perturbation (PBC) | Probability of company crisis (index); severity from BMR | Index of business continuity | Expert-based mapping rules |

### Workflow

1. Financial assessment

1. Receives financial statement data from input
2. Queries IAS/IFRS KB
3. Computes a set of financial indexes and ratios named financial features (Table 2)
4. Sends the financial features to the Decision tree and Prediction modules

2. Governance assessment

1. Receives text and data on governance structure from input
2. Computes a set of governance features (Table 2)
3. Sends the financial features to the Decision tree module and the Prediction module

3. Risk matrix

1. Receives data related to the internal assessment on vertical risks and other technical non-core data
2. Queries ISO 31000 et al KBs
3. Generates the risk matrix
4. Computes the risk matrix severity, which is a numerical value of the severity associated to the matrix of risks, considering
	1. the values of impact, probability and gravity (from low to high)
	2. the percentage of risk managed by the company.
5. Sends the severity value to the perturbation module

3. Decision tree

1. Receives governance and financial features from Financial assessment and Governance assessment modules
2. Queries Data back testing KB
3. Generates a tree-like graph
4. Sends the ranking of features and their importance to the Prediction module

4. Prediction

1. Receives governance and financial features from Financial assessment and Governance assessment modules and the ranking of features from the Decision tree module
2. Queries Data back testing KB
3. Computes the indexes Probability of company crisis and Adequacy of organizational model, whose values are between 0-1
4. Outputs

i. Probability of company crisis

ii. Adequacy of organizational model

1. Sends the two indexes (b.) to the Perturbation module

5. Perturbation

1. Receives Probability of company crisis index from the Prediction module and the risk matrix severity from the Risk matrix module
2. Queries Expert-based mapping rules
3. Computes the Index of business continuity, whose value is between 0-1
4. Outputs Index of business continuity

*Table 2 – Financial and governance attributes*

|  |  |  |
| --- | --- | --- |
| **Feature** | **Feature value** | **Feature type** |
| **1** | Absolute value | Revenue/Profit |
| **2** | Index/Percentage (%) | Revenue/Profit |
| **3** | Absolute value | Revenue/Profit |
| **4** | Absolute value | Revenue/Profit |
| **5** | Index/Percentage (%) | Revenue/Profit |
| **6** | Index/Percentage (%) | Cost/Debt |
| **7** | Absolute value | Cost/Debt |
| **8** | Index/Percentage (%) | Cost/Debt |
| **9** | Absolute value | Cost/Debt |
| **10** | Index/Percentage (%) | Cost/Debt |
| **11** | Absolute value | Production |
| **12** | Absolute value | Production |
| **13** | Index/Percentage (%) | Revenue/Profit |
| **14** | Absolute value | Production |
| **15** | Index/Percentage (%) | Cost/Debt |

|  |  |  |
| --- | --- | --- |
| **Feature** | **Feature value** | **Feature type** |
| **1** | Absolute value | Decision maker data |
| **2** | Index/Percentage (%) | Shareholder data |
| **3** | Absolute value | Shareholder data |
| **4** | Absolute value | Decision maker data |
| **5** | Absolute value | Decision maker data |

In the following subsections each AIM is analysed in detail, providing the lists of inputs and outputs of the AIMs required by the Use Case.

### Financial assessment

|  |  |
| --- | --- |
| **Function** | To analyse the data generated by the companies (i.e., financial statements) to assess the preliminary financial performances in the form of indexes. To build and extract the financial features for the machine learner |
| **Inputs** | 1. Financial statement data |
| **Outputs** | 1. Financial attributes (financial features) for the machine learner |

### Governance assessment

|  |  |
| --- | --- |
| **Function** | To build and extract the features related to the adequacy of the governance asset for the machine learner |
| **Inputs** | 1. Governance data (semantics of governance elements) on the company structures and boards |
| **Outputs** | 1. Governance attributes (governance features) for the machine learner |

### Risk matrix

|  |  |
| --- | --- |
| **Function** | To build the risk matrix to assess the impact of vertical risks (i.e., in this use case cyber and seismic) |
| **Inputs** | 1. Technical data from BIM2. internal assessment on cyber security  |
| **Outputs** | 1. Level (number) of the severity of a specific risk |
| **External data** | Socio-economic data from data bases |

### Decision tree

|  |  |
| --- | --- |
| **Function** | To create the decision trees for making decisions according to the Random Forest algorithm  |
| **Inputs** | 1. Financial features2. Governance features |
| **Outputs** | 1. Ranking of the features for the specific company with importance of related features |
| **External data** | Data on active and failed companies from back testing  |

### Prediction

|  |  |
| --- | --- |
| **Function** | To predict values of the likelihood of company default in a time horizon of 36 months and of the adequacy of the organizational model |
| **Inputs** | 1. Financial features2. Governance features |
| **Outputs** | 1. Probability of company crisis (index)2. Adequacy of the organizational model (index) |
| **External data** | Data on active and failed companies from back testing |

### Perturbation

|  |  |
| --- | --- |
| **Function** | To perturb the value of the probability of company crisis computed before, considering the impact of vertical risks on company performances  |
| **Inputs** | 1. Probability of company crisis (index)2. Severity of a specific risk |
| **Outputs** | 1. Index of business continuity |
| **External data** | Expert-based mapping rules |

# Conclusions

The document in its current form is work in progress. MPAI intends to add more details to the existing to enable MPAI to issue a Call for Technologies. MPAI may also add more usage exam­ples.

When the document will be considered sufficiently mature, MPAI will issue a Call for Technol­ogies requesting MPAI members and the industry members to submit proposals for:

1. Data formats suitable as inputs and outputs of the identified AIMs
2. Additions or removal of input/output signals to the identified AIMs with identification of data formats required by the new input/output signals
3. Possible alternative partitioning of the AIMs implementing the example cases providing
	1. Arguments in support of the proposed partitioning
	2. Detailed specifications of the inputs and outputs of the proposed AIMs
4. New Use Cases fully described as in the final version of this document.

Respondents will be asked to state in their submissions their intention to adhere to the Framework Licence developed for MPAI-CUI when licencing their technologies if included in the MPAI-CUI standard. Please note that “a Framework Licence is the set of conditions of use of a licence without the values, e.g. currency, percent, dates etc.”. The *Framework Licence* willgive the MPAI-CUI standard a *clear IPR licensing* framework.

The MPAI-CUI Framework Licence will be developed, as for all other MPAI Framework Licences, in compliance with the gener­ally accepted principles of competition law.

# References

[1] Perboli G., Arabnezhad E., A Machine Learning-based DSS for Mid and Long-Term Company Crisis Prediction. CIRRELT-2020-29. July 2020.