Moving Picture, Audio and Data Coding
by Artificial Intelligence
www.mpai.community

MPAI Technical Specification

Multimodal Conversation
MPAI-MMC

V2.2

WARNING

Use of the technologies described in this Technical Specification may infringe patents, copyrights or intellectual property rights of MPAI Members or non-members.

MPAI and its Members accept no responsibility whatsoever for damages or liability, direct or consequential, which may result from the use of this Technical Specification.

Readers are invited to review **Error! Reference source not found.**Notices and Disclaimers.

Technical Specification
Multimodal Conversation (MPAI-MMC) V2.2

1 Foreword

The international, unaffiliated, non-profit ***Moving Picture, Audio, and Data Coding by Artificial Intelligence (MPAI)*** organisation was established in September 2020 in the context of:

1. **Increasing** use of Artificial Intelligence (AI) technologies applied to a broad range of domains affecting millions of people
2. **Marginal** reliance on standards in the development of those AI applications
3. **Unprecedented** impact exerted by standards on the digital media industry affecting billions of people

believing that AI-based data coding standards will have a similar positive impact on the Information and Communication Technology industry.

The design principles of the MPAI organisation as established by the MPAI Statutes are the development of AI-based Data Coding standards in pursuit of the following policies:

1. Publish upfront clear Intellectual Property Rights licensing frameworks.
2. Adhere to a rigorous standard development process.
3. Be friendly to the AI context but, to the extent possible, remain agnostic to the technology thus allowing developers freedom in the selection of the more appropriate – AI or Data Processing – technologies for their needs.

4. Be attractive to different industries, end users, and regulators.
5. Address five standardisation areas:
    1. *Data Type*, a particular type of Data, e.g., Audio, Visual, Object, Scenes, and Descriptors with as clear semantics as possible.
    2. *Qualifier*, specialised Metadata conveying information on Sub-Types, Formats, and Attributes of a Data Type.
    3. *AI Module* (AIM), processing elements with identified functions and input/output Data Types.
    4. *AI Workflow* (AIW), MPAI-specified configurations of AIMs with identified functions and input/output Data Types.
    5. *AI Framework* (AIF), an environment enabling dynamic configuration, initialisation, execution, and control of AIWs.
6. Provide appropriate Governance of the ecosystem created by MPAI Technical Specifications enabling users to:
    1. *Operate* Reference Software Implementations of MPAI Technical Specifications provided together with Reference Software Specifications
    2. *Test* the conformance of an implementation with a Technical Specification using the Conformance Testing Specification.
    3. *Assess* the performance of an implementation of a Technical Specification using the Performance Assessment Specification.
    4. *Obtain* conforming implementations possibly with a performance assessment report from a trusted source through the MPAI Store.

Today, the MPAI organisation rests on four solid pillars:

1. The MPAI Patent Policy specifies the MPAI standard development process and the Framework Licence development guidelines.
2. *Technical Specification: Artificial Intelligence Framework (MPAI-AIF)* specifies an environment enabling initialisation, dynamic configuration, and control of AIWs in the standard AI Framework environment depicted in Figure 1. An AI Framework can execute AI applications called AI Workflows (AIW). An AIW includes interconnected AI Modules (AIM). MPAI-AIF supports small- and large-scale high-performance components and promotes solutions with improved explainability.

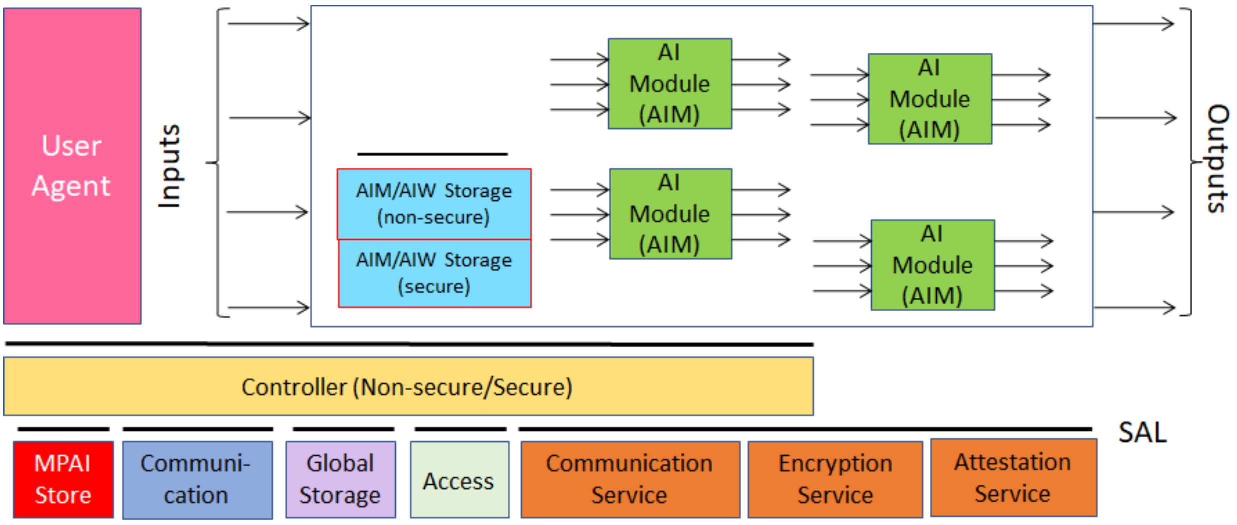


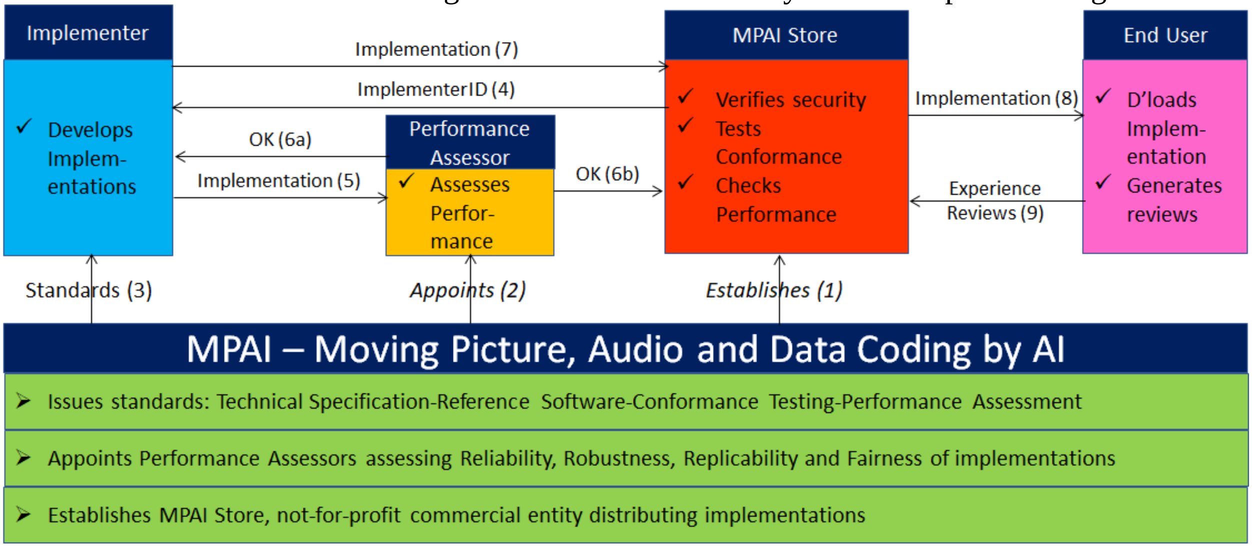
Figure 1 – The AI Framework (MPAI-AIF) V2 Reference Model

3. ***Technical Specification: Data Types, Formats, and Attributes (MPAI-TFA) V1.0*** specifies Qualifiers, a type of metadata supporting the operation of AIMs receiving data from other

AIMs. Qualifiers convey information on Sub-Types (e.g., the type of colour), Formats (e.g., the type of compression and transport), and Attributes (e.g., semantic information in the Content). Although Qualifiers are human-readable, they are only intended to be used by AIMs. Therefore, Text, Speech, Audio, and Visual Data exchanged by AIWs and AIMs should be interpreted as being composed of Content (Text, Speech, Audio, and Visual as appropriate) and associated Qualifiers. The specifications of most MPAI Data Types reflect this point.

4. ***Technical Specification: Governance of the MPAI Ecosystem (MPAI-GME) V1.1*** defines the following elements:
    1. Standards, i.e., the ensemble of Technical Specifications, Reference Software, Conformance Testing, and Performance Assessment.
    2. Developers of MPAI-specified AIMs and Integrators of MPAI-specified AIWS (Implementers).
    3. MPAI Store in charge of making AIMs and AIWs submitted by Implementers available to Integrators and End Users.
    4. Performance Assessors, independent entities assessing the performance of implementations in terms of Reliability, Replicability, Robustness, and Fairness.
    5. End Users.

The interaction between and among actors of the MPAI Ecosystem are depicted in Figure 2.



Implementer
✓ Develops Implem-entations
Implementation (7)
ImplementerID (4)
OK (6a)
Implementation (5)
Performance Assessor
✓ Assesses Perfor-mance
OK (6b)
MPAI Store
✓ Verifies security
✓ Tests Conformance
✓ Checks Performance
Implementation (8)
Experience Reviews (9)
End User
✓ D'loads Implem-entation
✓ Generates reviews
Standards (3)
Appoints (2)
Establishes (1)
MPAI – Moving Picture, Audio and Data Coding by AI
➤ Issues standards: Technical Specification-Reference Software-Conformance Testing-Performance Assessment
➤ Appoints Performance Assessors assessing Reliability, Robustness, Replicability and Fairness of implementations
➤ Establishes MPAI Store, not-for-profit commercial entity distributing implementations

Figure 2 – The MPAI Ecosystem

2 Introduction (Informative)

From the moment a human built the first machine, there was a need to "communicate" with it. In the past, humans communicated with more primitive machines by touch, later by characters and then with speech and even visual means. Then, more complex machines were built and the need for more sophisticated communication methods arose. Today, as personal devices become more pervasive, and the use of information and other online services become ubiquitous, human-machine communication often becomes more direct and even "personal".

The ability of Artificial Intelligence to learn from interactions with humans gives machines the ability to improve their "conversational" capabilities by better understanding the meaning of what a human types or says and by providing more pertinent responses. If properly trained, machines can also learn to understand additional or hidden meanings of a sentence by analysing a human's text, speech, or gestures. Machines can also be made to develop and rely on "internal statuses" comparable to those driving the attitudes of conversing humans. Thus, they can provide responses – in text, speech, and gestures – that are more human-like and richer in content.

***T**echnical Specification: Multimodal Conversation (MPAI-MMC) V2.2** has been developed by MPAI in pursuit of the following policies:

1. Be friendly to the AI context but, to the extent possible, agnostic to the technology – AI or Data Processing – used in an implementation.
2. Be attractive to different industries, end users, and regulators.
3. Address three levels of standardisation any of which an implementer can freely decide to adopt:
    1. Data types, i.e., the data exchanged by systems.
    2. Components (called AI Modules – AIM).
    3. Connections of components (called AI Workflows – AIW).
4. Specify the data exchanged by components with a semantic that is clear to the extent possible.

The MPAI-MMC V2 Technical Specification will be accompanied by the Reference Software, Conformance Testing, and Performance Assessment Specifications. Conformance Testing specifies methods enabling users to ascertain whether a data type generated by an AIM, an AIM, or an AIW conform with this Technical Specification.

The **MPAI-MMC V2.2** Technical Specification provides the technologies supporting the implementation of a subset or the totality of the possibilities envisaged by this Introduction:

1. It is organised by Use Cases, such as Conversation with Personal Status, Multimodal Question Answering, and Unidirectional Speech Translation, corresponding to AI Workflows.
2. Each Use Case provides:
    1. The functions.
    2. The Input/Output Data of the AIW implementing it.
    3. The Reference Model specifying the AIM topology.
    4. The AIMs specified in terms of functions performed and Input/Output Data.

In all Chapters and Sections, Terms beginning with a capital letter are defined in *Table 1* if they are specific to this Technical Specification and in *Table 2* if they are common to all MPAI Technical Specifications. All Chapters, Sections, and Annexes are Normative unless they are labelled as Informative.

3 Scope

Technical Specification: Multimodal Conversation (MPAI-MMC) V2.2, in the following also called MPAI-MMC V2.2 or simply MPAI-MMC, specifies:

1. **Data Types** for use by MPAI-MMC V2.2 and other MPAI Technical Specifications.
2. **AI Modules** enabling analysis of text, speech, and other non-verbal components used in human-machine and machine-machine conversation applications.
3. **AI Workflows** implementing Use Cases that use AI Modules and Data Types from MPAI-MMC and other MPAI Technical Specifications to provide recognised applications in the Multimodal Conversation domain.

The Use Cases includes in this Technical Specification are:

1. *Answer to Multimodal Question* (MMC-AMQ) providing a text or speech answer to a text or speech question and an image.
2. *Conversation About a Scene* (MMC-CAS) where a human converses with a machine pointing at the objects scattered in a room and displaying Personal Status in their speech, face, and gestures while the machine responds displaying its Personal Status in speech, face, and gesture.
3. *Conversation with Personal Status* (MMC-CPS), enabling conversation and question answering with a machine able to extract the inner state of the entity it is conversing with and

showing itself as a speaking digital human able to express a Personal Status. By adding or removing minor components to this general Use Case, five Use Cases are spawned:

4. *Conversation with Emotion* (MMC-CWE), enabling audio-visual conversation with a machine impersonated by a synthetic voice and an animated face.
5. *Human-Connected Autonomous Vehicle Interaction* (MMC-HCI) where humans converse with a machine displaying Personal Status after having been properly identified by the machine with their speech and face in outdoor and indoor conditions while the machine responds displaying its Personal Status in speech, face, and gesture.
6. *Multimodal Question Answering* (MQA), enabling request for information about a displayed object.
7. *Text and Speech Translation* (MMC-TST) supporting a variety of text and speech translation applications where users can specify whether speech or text is used as input and, if it is speech, whether their speech features are preserved in the interpreted speech.
8. *Virtual Meeting Secretary* (MMC-VSV) where an avatar not representing a human in a virtual avatar-based video conference extracts Personal Status from Text, Speech, Face, and Gestures, displays a summary of what other avatars say, and receives and act on comments.

The Composite AI Module specified by MPAI-MMC V2.2 is *Personal Status Extraction* (MMC-PSE) that estimates the Personal Status conveyed by Text, Speech, Face, and Gesture – of an Entity, i.e., a real or digital human.

Note that:

1. Each AI Workflow implementing a Use Case normatively defines:
    - The Functions of the AIW implementing it and of the AIMs.
    - The Connections between and among the AIMs
    - The Semantics and the Formats of the input and output data of the AIW and the AIMs.
2. Each Composite AIM normatively defines:
    - The Functions of the Composite AIM implementing it and of the AIMs.
    - The Connections between and among the AIMs
    - The Semantics and the Formats of the input and output data of the AIW and the AIMs.

The word *normatively* implies that an Implementation claiming Conformance to:

1. An *AIW*, shall:
1.
    1. Perform the AIW function specified in the appropriate Section of Chapter 5.
    2. All AIMs, their topology and connections should conform with the AIW Architecture specified in the appropriate Section of Chapter 5.
    3. The AIW and AIM input and output data should have the formats specified in the appropriate Sections of Chapter 7.
2. An *AIM*, shall:
1.
    1. Perform the functions specified by the appropriate Section of Chapter 5 or 6.
    2. Receive and produce the data specified in the appropriate Section of Chapter 7.
    3. A data *Format*, the data shall have the format specified in Chapter 7.

Implementers of this Technical Specification should note that:

1. The Reference Software of this Technical Specification may be to develop Implementations.
2. The Conformance Testing specification may be used to test the conformity of an Implementation to this Standard.
3. The level of Performance of an Implementation may be assessed based on the Performance Assessment specification of this Standard.

All Users should consider Notices and Disclaimers.

MPAI-MMC V2.2 has been developed by the MPAI Multimodal Conversation Development Committee (MM-DC). MPAI expects to produce future MPAI-MMC Versions extending the scope of the Use Cases and/or add new Use Cases supported by existing of new AI Modules and Data Types within the scope of Multimodal Conversation.

4 Definitions

Capitalised Terms have the meaning defined in *Table 1*. Terms applicable to all MPAI Technical Specifications are defined in *Table 2*.

Lower case Terms have the meaning commonly defined for the context in which they are used. For instance, *Table 1* defines *Object* and *Scene* but does not define *object* and *scene*.

A dash "-" preceding a Term in *Table 1* indicates the following readings according to the font:

1. Normal font: the Term in the table without a dash and preceding the one with a dash should be read before that Term. For example, "Avatar" and "- Model" will yield "Avatar Model."
2. *Italic* font: the Term in *Table 1* without a dash and preceding the one with a dash should be read after that Term. For example, "Avatar" and "- Portable" will yield "Portable Avatar."

Table 1 – Table of terms and definitions

Term	Definition
Attitude	
– *Social*	The coded representation of the internal state related to the way a human or avatar intends to position vis-à-vis the Environment or subsets of it, e.g., "Respectful", "Confrontational", "Soothing".
– *Spatial*	Position and Orientation and their velocities and accelerations of an Audio and Visual Object in a Virtual Environment.
Audio	Digital representation of an analogue audio signal sampled at a frequency between 8-192 kHz with a number of bits/sample between 8 and 32, and non-linear and linear quantisation.
– Object	Coded representation of Audio information with its metadata. An Audio Object can be a combination of Audio Objects.
– Scene	The Audio Objects of an Environment with Object location metadata.
Audio-Visual Object	Coded representation of Audio-Visual information with its metadata. An Audio-Visual Object can be a combination of Audio-Visual Objects.
Audio-Visual Scene	(AV Scene) The Audio-Visual Objects of an Environment with Object location metadata.
Avatar	An animated 3D object representing a real or fictitious person in a Virtual Space.
– Model	An inanimate avatar exposing interfaces enabling animation.
Cognitive State	The coded representation of the internal state reflecting the way a human or avatar understands the Environment, such as "Confused", "Dubious", "Convinced".
Colour (of speech)	The timber of an identifiable voice independent of a current Personal Status and language.
Connected Autonomous Vehicle	A vehicle able to autonomously reach an assigned geographical position by: 1. Understanding human utterances. 2. Planning a route. 3. Sensing and interpreting the Environment. 4. Exchanging information with other CAV.

	5. Acting on the CAV’s motion actuation subsystem.
Context	Additional information about a communication emitted by an Entity, such as language, culture etc..
Data	Information in digital form.
– Format	The standard digital representation of Data.
– Type	An instance of Data with a specific Data Format.
Descriptor	Coded representation of text, audio, speech, or visual feature.
Digital Representation	Data corresponding to and representing a real entity.
Emotion	The coded representation of the internal state resulting from the interaction of a human or avatar with the Environment or subsets of it, such as “Angry”, “Sad”, “Determined”.
Entity	A real or Digital Human
Environment	A Virtual Space containing a Scene.
Face	The portion of a 2D or 3D digital representation corresponding to the face of a human.
Factor	One of Emotion, Cognitive State and Attitude.
Gesture	A movement of the body or part of it, such as the head, arm, hand, and finger, often a complement to a vocal utterance.
Grade	The intensity of a Factor.
Human	A human being in a real space.
– *Digital*	A Digitised or a Virtual Human in a Virtual Space.
– *Digitised*	An Object in a Virtual Space that has the appearance of a specific human when rendered.
– *Virtual*	An Object in a Virtual Space created by a computer that has a human appearance when rendered but is not a Digitised Human.
Identifier	The label uniquely associated with a human or an avatar or an object.
Instance	An element of a set of entities – Objects, users etc. – belonging to some levels in a hierarchical classification (taxonomy).
Intention	The result of analysis of the goal of an input question.
Manifestation	The manner of showing the Personal Status, or a subset of it, in any one of Speech, Face, and Gesture.
Meaning	Information extracted from Text such as syntactic and semantic information, Personal Status, and other information, such as an Object Identifier.
Modality	One of Text, Speech, Face, or Gesture.
Object Descriptors	Attribute of the coded representation of an object in a Scene, including its Spatial Attitude.
Orientation	The set of the 3 roll, pitch, yaw angles indicating the rotation around the principal axis (x) of an Object, its y axis having an angle of 90˚ counterclockwise (right-to-left) with the x axis and its z axis pointing up toward the viewer.
Personal Status	The ensemble of information internal to a person, including Emotion, Cognitive State, and Attitude.
Portable Avatar	A Data Type representing an Avatar and its Context.
Pitch	The fundamental frequency of Speech. Pitch is the attribute that makes it possible to judge sounds as “higher” and “lower.”
Point of View	The Spatial Attitude of a human or avatar looking at an Environment.

Position	The 3 coordinates (x,y,z) of a representative point of an object in the Real and Virtual Space.
Refined Text	The Text resulting from the analysis of the Text produced by Automatic Speech Recognition made by Natural Language Understanding.
Scene	A structured composition of Objects.
Speech	Digital representation of analogue speech sampled at a frequency between 8 kHz and 96 kHz with a number of bits/sample of 8, 16 and 24, and non-linear and linear quantisation.
– Features	Aspects of a speech segment that enable its description and reproduction, e.g., degree of vocal tension, Pitch, etc., and that can be automatically recognised and extracted for speech synthesis or other related purposes.
– Rate	The number of Speech Units per second.
– Unit	Phoneme, syllable, or word as a segment of Speech.
Summary	An abridged outline of the content of the utterance(s) of one or more Users possibly including their Personal Statuses.
Text	A sequence of characters drawn from a finite alphabet.
Visual Object	Coded representation of Visual information with its metadata. A Video Object can be a combination of Video Objects.
Vocal Gesture	Utterance, such as cough, laugh, hesitation, etc. Lexical elements are excluded.

Table 2 – MPAI-wide Terms

Term	**Definition**
Access	Static or slowly changing data that are required by an application such as domain knowledge data, data models, etc.
AI Framework (AIF)	The environment where AIWs are executed.
AI Model (AIM)	A data processing element receiving AIM-specific Inputs and producing AIM-specific Outputs according to according to its Function. An AIM may be an aggregation of AIMs.
AI Workflow (AIW)	A structured aggregation of AIMs implementing a Use Case receiving AIW-specific inputs and producing AIW-specific outputs according to the AIW Function.
Application Standard	An MPAI Standard designed to enable a particular application domain.
Channel	A connection between an output port of an AIM and an input port of an AIM. The term "connection" is also used as synonymous.
Communication	The infrastructure that implements message passing between AIMs.
Component	One of the 7 AIF elements: Access, Communication, Controller, Internal Storage, Global Storage, Store, and User Agent
Composite AIM	An AIM aggregating more than one AIM.
Component	One of the 7 AIF elements: Access, Communication, Controller, Internal Storage, Global Storage, Store, and User Agent
Conformance	The attribute of an Implementation of being a correct technical Implementation of a Technical Specification.
– Testing	The normative document specifying the Means to Test the Conformance of an Implementation.
– Testing Means	Procedures, tools, data sets and/or data set characteristics to Test the Conformance of an Implementation.
Connection	A channel connecting an output port of an AIM and an input port of an AIM.

Controller	A Component that manages and controls the AIMs in the AIF, so that they execute in the correct order and at the time when they are needed
Data	Information in digital form.
– Format	The standard digital representation of Data.
– Type	An instance of Data with a specific Data Format.
– Semantics	The meaning of Data.
Descriptor	Coded representation of a text, audio, speech, or visual feature.
Digital Representation	Data corresponding to and representing a physical entity.
Ecosystem	The ensemble of actors making it possible for a User to execute an application composed of an AIF, one or more AIWs, each with one or more AIMs potentially sourced from independent implementers.
Explainability	The ability to trace the output of an Implementation back to the inputs that have produced it.
Fairness	The attribute of an Implementation whose extent of applicability can be assessed by making the training set and/or network open to testing for bias and unanticipated results.
Function	The operations effected by an AIW or an AIM on input data.
Global Storage	A Component to store data shared by AIMs.
AIM/AIW Storage	A Component to store data of the individual AIMs.
Identifier	A name that uniquely identifies an Implementation.
Implementation	1. An embodiment of the MPAI-AIF Technical Specification, or 2. An AIW or AIM of a particular Level (1-2-3) conforming with a Use Case of an MPAI Application Standard.
Implementer	A legal entity implementing MPAI Technical Specifications.
ImplementerID (IID)	A unique name assigned by the ImplementerID Registration Authority to an Implementer.
ImplementerID Registration Authority (IIDRA)	The entity appointed by MPAI to assign ImplementerID's to Implementers.
Instance ID	Instance of a class of Objects and the Group of Objects the Instance belongs to.
Interoperability	The ability to functionally replace an AIM with another AIW having the same Interoperability Level
– Level	The attribute of an AIW and its AIMs to be executable in an AIF Implementation and to: 1. Be proprietary (Level 1) 2. Pass the Conformance Testing (Level 2) of an Application Standard 3. Pass the Performance Testing (Level 3) of an Application Standard.
Knowledge Base	Structured and/or unstructured information made accessible to AIMs via MPAI-specified interfaces
Message	A sequence of Records transported by Communication through Channels.
Normativity	The set of attributes of a technology or a set of technologies specified by the applicable parts of an MPAI standard.
Performance	The attribute of an Implementation of being Reliable, Robust, Fair and Replicable.

– Assessment	The normative document specifying the Means to Assess the Grade of Performance of an Implementation.
– Assessment Means	Procedures, tools, data sets and/or data set characteristics to Assess the Performance of an Implementation.
– Assessor	An entity Assessing the Performance of an Implementation.
Profile	A particular subset of the technologies used in MPAI-AIF or an AIW of an Application Standard and, where applicable, the classes, other subsets, options and parameters relevant to that subset.
Record	A data structure with a specified structure
Reference Model	The AIMs and theirs Connections in an AIW.
Reference Software	A technically correct software implementation of a Technical Specification containing source code, or source and compiled code.
Reliability	The attribute of an Implementation that performs as specified by the Application Standard, profile, and version the Implementation refers to, e.g., within the application scope, stated limitations, and for the period of time specified by the Implementer.
Replicability	The attribute of an Implementation whose Performance, as Assessed by a Performance Assessor, can be replicated, within an agreed level, by another Performance Assessor.
Robustness	The attribute of an Implementation that copes with data outside of the stated application scope with an estimated degree of confidence.
Scope	The domain of applicability of an MPAI Application Standard
Service Provider	An entrepreneur who offers an Implementation as a service (e.g., a recommendation service) to Users.
Standard	A set of Technical Specification, Reference Software, Conformance Testing, Performance Assessment, and Technical Report of an MPAI application Standard.
Technical Specification	(Framework) the normative specification of the AIF. (Application) the normative specification of the set of AIWs belonging to an application domain along with the AIMs required to Implement the AIWs that includes: 1. The formats of the Input/Output data of the AIWs implementing the AIWs. 2. The Connections of the AIMs of the AIW. 3. The formats of the Input/Output data of the AIMs belonging to the AIW.
Testing Laboratory	A laboratory accredited to Assess the Grade of Performance of Implementations.
Time Base	The protocol specifying how Components can access timing information
Topology	The set of AIM Connections of an AIW.
Use Case	A particular instance of the Application domain target of an Application Standard.
User	A user of an Implementation.
User Agent	The Component interfacing the user with an AIF through the Controller
Version	A revision or extension of a Standard or of one of its elements.
Zero Trust	A cybersecurity model primarily focused on data and service protection that assumes no implicit trust.

5 References

5.1 Normative References

This standard normatively references the following documents, both from MPAI and other standards organisations. MPAI standards are publicly available at https://mpai.community/standards/resources/.

1. MPAI; Technical Specification: Governance of the MPAI Ecosystem (MPAI-GME) V1.1.
2. MPAI; Technical Specification: Artificial Intelligence Framework (MPAI-AIF) V2.0.
3. MPAI; Technical Specification: Context-based Audio Enhancement (MPAI-CAE) V2.2.
4. MPAI; Technical Specification: MPAI Metaverse Model (MPAI-MMM) – Architecture (MMM-ARC) V1.1.
5. MPAI; Technical Specification: Object and Scene Description (MPAI-OSD) V1.1.
6. MPAI; Technical Specification: Portable Avatar Format (MPAI-PAF) V1.2.
7. MPAI; Technical Specifications: AI Module Profiles (MPAI-PRF) V1.0.
8. ITU-R; Long-form file format for the international exchange of audio programme materials with metadata; BS.2088 (10/2019) .
9. ISO 639; Codes for the Representation of Names of Languages – Part 1: Alpha-2 Code.
10. ISO/IEC 10646; Information technology – Universal Coded Character Set.
11. ISO/IEC 14496-10; Information technology – Coding of audio-visual objects – Part 10: Advanced Video Coding.
12. ISO/IEC 14496-12; Information technology – Coding of audio-visual objects – Part 12: ISO base media file format.
13. ISO/IEC 23008-2; Information technology – High efficiency coding and media delivery in heterogeneous environments – Part 2: High Efficiency Video Coding.
14. ISO/IEC 23094-1; Information technology – General video coding – Part 1: Essential Video Coding.

5.2 Informative References

The references provided here are for information purpose.

15. MPAI; The MPAI Statutes.
16. MPAI; The MPAI Patent Policy.
17. MPAI; Framework Licence of the Multimodal Conversation Technical Specification (MPAI-MMC) V1; https://mpai.community/standards/mpai-mmc/framework-licence/mpai-mmc-v1-framework-licence/.
18. MPAI; Framework Licence of the Multimodal Conversation Technical Specification (MPAI-MMC) V2; https://mpai.community/standards/mpai-mmc/call-for-technologies/mpai-mmc-v2-call-for-technologies/.
19. Ekman, Paul (1999), “Basic Emotions”, in Dalgleish, T; Power, M (eds.), Handbook of Cognition and Emotion (PDF), Sussex, UK: John Wiley & Sons.
20. Emotion Markup Language (EmotionML) 1.0; https://www.w3.org/TR/2010/WD-emotionml-20100729/diffmarked.html.
21. Hobbs J.R., Gordon A.S. (2011) The Deep Lexical Semantics of Emotions. In: Ahmad K. (eds) Affective Computing and Sentiment Analysis. Text, Speech, and Language Technology, vol 45. Springer, Dordrecht, https://people.ict.usc.edu/~gordon/publications/EMOT08.PDF and https://www.researchgate.net/publication/227251103_The_Deep_Lexical_Semantics_of_Emotions.

6 AI Workflows

Technical Specification: Multimodal Conversation (MPAI-MMC) V2.2 assumes that implementations will be based on *Technical Specification: AI Framework (MPAI-AIF) V2.0*. specifying an AI Framework (AIF) where AI Workflows (AIW) composed of interconnected AI Modules (AIM) are executed.

Table 1 displays the full list of AIWs specified by MPAI-MMC V2.0. Click a listed AIW to access its dedicated page, which includes a its functions, reference model, I/O Data, Functions of AIMs, I/O Data of AIMs, and s table providing links to the AIW-related AIW, AIMs, and JSON metadata.

All previously specified MPAI-MMC AI-Workflows are superseded by those specified by V2.2.

Table 1 – AIWs of MPAI-MMC V2.2

Acronym	Title	JSON
MMC-AMQ	**Answer to Multimodal Question**	**X**
MMC-CAS	**Conversation About a Scene**	**X**
MMC-CPS	**Conversation with Personal Status**	**X**
MMC-CWE	**Conversation with Emotion**	**X**
MMC-HCI	**Human-CAV Interaction**	**X**
MMC-MQA	**Multimodal Question Answering**	**X**
MMC-TST	**Text and Speech Translation**	**X**
MMC-VMS	**Virtual Meeting Secretary**	**X**

NOTE: Conversation with Personal Status (MMC-CWP) is an enhanced or more complete version of Conversation with Emotion (MMC-CWE) specified in earlier version of the AIW. In MPAI's terminology, Personal Status includes not only Emotion (e.g., angry, sad, etc.) but Cognitive Status (surprised, confused, etc.) and Social Attitude (sarcastic, polite, etc.).

7 AI Modules

Table 1 provides the links to the specifications and the JSON syntax of all AIMs specified by ***Technical Specification: Multimodal Conversation (MPAI-MMC) V2.2***. All previously specified MPAI-MMC AI-Modules are superseded by those specified by V2.2.

Table 1 – Specifications and JSON syntax of AIMs used by MPAI-MMC V2.2

AIMs	Name	JSON
MMC-AQM	Answer to Question Module	X
MMC-ASR	Automatic Speech Recognition	X
MMC-AUS	Audio Segmentation	X
MMC-EDP	Entity Dialogue Processing	X
MMC-ESD	Entity Speech Description	X
MMC-ETD	Entity Text Description	X
MMC-MEF	Multimodal Emotion Fusion	X
MMC-NLU	Natural Language Understanding	X
MMC-PMX	Personal Status Multiplexing	X
MMC-PSE	Personal Status Extraction	X
MMC-PSI	PS-Speech Interpretation	X
MMC-PTI	PS-Text Interpretation	X
MMC-QAM	Question Analysis Module	X
MMC-SCM	Summary Creation Module	X
MMC-SIR	Speaker Identity Recognition	X
MMC-SPE	Speech Personal Status Extraction	X
MMC-STD	Speech Translation with Descriptors	X
MMC-TSD	Text-to-Speech with Descriptors	X
MMC-TST	Text and Speech Translation	X

MMC-TIQ	Text and Image Query	X
MMC-TTS	Text-To-Speech	X
MMC-TTT	Text-to-Text Translation	X
MMC-VLA	Video Lip Animation	X

8 Data Types

This page gives the links to the specification of Data Types of ***Technical Specification: Multimodal Conversation (MPAI-MMC) V2.2***. All previously specified MPAI-MMC Data Types are superseded by those specified by V2.2.

AMS-HCI Message	Cognitive State	Ego-Remote HCI Message	Emotion
Face Personal Status	Gesture Personal Status	Intention	Meaning
Personal Status	Social Attitude	Speech Basic Scene Descriptors	Speech Basic Scene Geometry
Speech Descriptors	Speech Object	Speech Overlap	Speech Personal Status
Speech Scene Descriptors	Speech Scene Geometry	Summary	Text Descriptors
Text Object	Text Segment	Text Word	

9 Datasets

9.1 Introduction

Testing the Conformance of MMC-CWE, MMC-MQA, and MMC-UST requires datasets to test Data, AIMs, and AIWs. The Data Formats belong to one of Text, Audio, Video, and JSON and should have the characteristics of Table 1:

Table 1 – Data Types for Conformance Testing of MMC-CWE, MMC-MQA, and MMC-UST

Data Type	Characteristics
Text	The texts files shall composed of **Unicode** characters.
Speech file	The speech files shall be conforming **.wav** files.
Video file	The video files shall be conforming **MP4** files.

Image File	The Image file shall be conforming
Emotion	Emotion files shall be JSON files conforming with the **Emotion** JSON Schema.
Intention	Emotion files shall be JSON files conforming with the **Intention** JSON Schema.
Meaning	Emotion files shall be JSON files conforming with the **Meaning** JSON Schema.

Humans shall carry out Conformance Testing by visual and auditory inspection. Appropriate software may replace humans as Conformance Testers.

Conformance Testing Datasets are publicly available upon registration.

9.2 Text with Emotion

9.2.1 Coherent scenarios

Happy	1. Today was a wonderful day. I spent quality time with my parents, and the restaurant was excellent as well. I look forward to seeing them again!
	2. I'm so excited about Christmas. This year, my girlfriend and I are going to celebrate the holiday together. We'll decorate our room, and it'll be so much fun.
	3. Today I watched a movie called 'The Pianist.' Not only was it touching, but also very absorbing. Now I feel very happy thanks to the memorable experience.
	4. The weather is awesome these days. It is not too cold, not too hot, and the sun shines beautifully. I look forward to the picnic that is scheduled this weekend.
	5. Nowadays my business is running very smoothly. There are no unexpected issues arising, and my employees are working very diligently. I am very relieved.
Angry	1. Today my coworkers treated me really badly. They blamed me for the things that were neither my responsibility nor the result of my actions. This is so unfair.
	2. I am angry with my sister. She not only does not finish her chores, but forces me to do the chores for her. This is not a new occasion, but this time I can't,
	stand it.
	3. Yesterday I had an argument with a friend of mine. He always wants me to listen to him very carefully and provide advice, but when I'm in need of the help of the same sort, he doesn't fulfill his duty at all. I'm furious about this.
	4. These days consumer price is skyrocketing. However, the government and political parties are busy blaming the external variables, not trying hard to solve.
	the problem that ordinary citizens are facing. Why is there no one trying to be responsible?
	5. Because of my superior in my workplace, I am doing monotonous tasks all day long these days. I have to look at thousands of boring images and classify them each day, which drives me crazy. I cannot but blame my superior.
Neutral	1. Seoul is the capital city of the Republic of Korea. It is a city of almost ten million residents. According to "The Global Livability Index" Seoul is ranked the fourth most livable city in Asia as of 2023.
	2. There is a famous proverb, "Honesty is the best policy." In essence, it suggests that honesty is the most effective and beneficial approach in various aspects of life.
	3. There is a famous saying, "Don't judge a book by its cover." This advises people not to form an opinion or make assumptions about someone or something based solely on its outward appearance.
	4. Global warming refers to the long-term increase in Earth's average surface temperature due to human activities, primarily the emission of greenhouse gases. Greenhouse gases trap heat in the Earth's atmosphere, leading to the warming effect.
	5. Inflation is a general increase of the prices of goods and services in an economy. This is usually measured using the consumer price index (CPI).

9.2.2 Incoherent scenarios

Text	Meaning	Speech	Face	Sentences
Happy	Happy	Angry	Angry	I'm headed to a yoga class now, and then I have a cozy evening planned with a good book. Life is good, for sure.
Happy	Happy	Neutral	Neutral	With a big scoop of ice cream in hand, I laughed and played in the park, feeling super happy as the sun shone brightly overhead.
Angry	Angry	Happy	Happy	Witnessing my neighbor being rude and disrespectful to an old stranger asking for directions, I couldn't be sane, because that old man was my father.
Neutral	Neutral	Happy	Happy	A political party is an organization that coordinates candidates to compete in a particular country's elections. It is common for the members of a party to hold similar ideas about politics.
Neutral	Neutral	Angry	Angry	According to Max Weber, a state is a compulsory political organization with a centralized government that maintains a monopoly of the legitimate use of force within a certain territory.

9.3 Audio and Video with Emotion

9.3.1 Neutral

MPAI emotions neutral 1 audio.240309.1041.wav
MPAI emotions neutral 1 video.240309.1041.mp4
MPAI emotions neutral 1.240309.1041.mp4
MPAI emotions neutral 2 audio.240309.1041.wav
MPAI emotions neutral 2 video.240309.1041.mp4
MPAI emotions neutral 2.240309.1041.mp4
MPAI emotions neutral 3 audio.240309.1041.wav
MPAI emotions neutral 3 video.240309.1041.mp4
MPAI emotions neutral 3.240309.1041.mp4
MPAI emotions neutral 4 audio.240309.1041.wav
MPAI emotions neutral 4 video.240309.1041.mp4
MPAI emotions neutral 4.240309.1041.mp4
MPAI emotions neutral 5 audio.240309.1041.wav
MPAI emotions neutral 5 video.240309.1041.mp4
MPAI emotions neutral 5.240309.1041.mp4

9.3.2 Angry

MPAI emotions angry 5.240309.1041.mp4
MPAI emotions angry 5 video.240309.1041.mp4
MPAI emotions angry 5 audio.240309.1041.wav
MPAI emotions angry 4.240309.1041.mp4
MPAI emotions angry 4 video.240309.1041.mp4
MPAI emotions angry 4 audio.240309.1041.wav
MPAI emotions angry 3.240309.1041.mp4
MPAI emotions angry 3 video.240309.1041.mp4
MPAI emotions angry 3 audio.240309.1041.wav
MPAI emotions angry 2.240309.1041.mp4

MPAI emotions angry 2 video.240309.1041.mp4
MPAI emotions angry 2 audio.240309.1041.wav
MPAI emotions angry 1.240309.1041.mp4
MPAI emotions angry 1 video.240309.1041.mp4
MPAI emotions angry 1 audio.240309.1041.wav

9.3.3 Happy

MPAI emotions happy 1 audio.240309.1041.wav
MPAI emotions happy 1 video.240309.1041.mp4
MPAI emotions happy 1.240309.1041.mp4
MPAI emotions happy 2 audio.240309.1041.wav
MPAI emotions happy 2 video.240309.1041.mp4
MPAI emotions happy 2.240309.1041.mp4
MPAI emotions happy 3 audio.240309.1041.wav
MPAI emotions happy 3 video.240309.1041.mp4
MPAI emotions happy 3.240309.1041.mp4
MPAI emotions happy 4 audio.240309.1041.wav
MPAI emotions happy 4 video.240309.1041.mp4
MPAI emotions happy 4.240309.1041.mp4
MPAI emotions happy 5 audio.240309.1041.wav
MPAI emotions happy 5 video.240309.1041.mp4
MPAI emotions happy 5.240309.1041.mp4

9.3.4 Incoherent

MPAI emotions angry text happy voice.240309.1041.mp4
MPAI emotions angry text happy voice audio.240309.1041.wav
MPAI emotions angry text happy voice video.240309.1041.mp4
MPAI emotions happy text angry voice.240309.1041.mp4
MPAI emotions happy text angry voice audio.240309.1041.wav
MPAI emotions happy text angry voice video.240309.1041.mp4
MPAI emotions happy text neutral voice.240309.1041.mp4
MPAI emotions happy text neutral voice audio.240309.1041.wav
MPAI emotions happy text neutral voice video.240309.1041.mp4
MPAI emotions neutral text angry voice.240311.0915.mp4
MPAI emotions neutral text angry voice audio.240311.0915.wav
MPAI emotions neutral text angry voice video.240311.0915.mp4
MPAI emotions neutral text happy voice audio.240309.1041.wav
MPAI emotions neutral text happy voice video.240309.1041.mp4

9.4 Emotion JSON Files

The JSON files below represent Happy, Angry, and Neutral Emotions.

```
{
“EmotionType”:{
“emotionDegree”:”high”,
“emotionName”:”happy”,
“emotionSetName”:”MPAI Basic Emotion Set”
}
}
{
“EmotionType”:{
```

```
“emotionDegree”:”high”,
“emotionName”:”happy”,
“emotionSetName”:”MPAI Basic Emotion Set”
}
}
{
“EmotionType”:{
        “emotionDegree”:”high”,
“emotionName”:”happy”,
“emotionSetName”:”MPAI Basic Emotion Set”
}
}
```

9.5 Meaning JSON Files

Sentence 1: Today was a wonderful day! I spent quality time with my parents, and the McDonald restaurant was excellent, too. I’m looking forward to seeing them again!

```
{
    “meaning”: {
        “POS_tagging”: {
            “POS_tagging_set”: “CST’s Part-Of-Speech tagger (Brill, with adaptations)”,
            “POS_tagging_result”: “Today/RB was/VBD a/DT wonderful/JJ day/NN !/. I/PRP
spent/VBD quality/NN time/NN with/IN my/PRP$ parents/NNS ,/, and/CC the/DT
McDonald/NNP     restaurant/NN was/VBD excellent/JJ ,/, too/RB ./. I’m/NNP looking/VBG
forward/RB to/TO seeing/VBG them/PRP again/RB !/.”
        },
        “NE_tagging”: {
            “NE_tagging_set”: “CST’s named entity recogniser”,
            “NE_tagging_result”: ” [Today,misc,uncertain] was a wonderful day ! I spent
quality time with my parents, and the [McDonald,person,likely] restaurant was excellent , too .
I’m looking forward to seeing them again!”
        },
        “dependency_tagging”: {
            “dependency_tagging_set”: “CG-dependency,
https://edu.visl.dk/visl/en/parsing/automatic/dependency.php “,
            “dependency_tagging_result”: “<ß>nToday [today] <*> <atemp> ADV @ADVL>
#1->2nwas [be] <mv> V IMPF 1/3S @FS-STA #2->0na [a] <indef> ART S @>N #3-
>5nwonderful [wonderful] ADJ POS @>N #4->5nday [day] <dur> <per> <idf> <nhead> N S
NOM @<SUBJ #5->2n! [!] PU @PU #6->0n</s>n<ß>nI [I] <*> PERS 1S NOM @SUBJ> #1-
>2nspent [spend] <cjt-head> <mv> V IMPF @FS-STA #2->0nquality [quality] <f-q> <f-phys>
<comp1> <first> <idf> <comp1> <ncomp> N S NOM @>N #3->4ntime [time] <ac-cat>
<temp> <per> <num+> <second> <comp2> <idf> <nhead> N S NOM @<ACC #4->2nwith
[with] PRP @<ADVL #5->2nmy [I] <poss> <refl> <det> PERS 1S GEN @>N #6->7nparents
[parent] <Hfam> <def> <nhead> N P NOM @P< #7->5n, [,] PU @PU #8->0nand [and] <clb?>
<co-fin> KC @CO #9->2nthe [the] <def> ART S/P @>N #10->12nMcDonald [McDonald] <*>
<Proper> <first> <ncomp> N S NOM @>N #11->12nrestaurant [restaurant] <inst> <second>
<def> <nhead> N S NOM @SUBJ> #12->13nwas [be] <cjt> <mv> V IMPF 1/3S @FS-STA
#13->2nexcellent [excellent] <Q:good> ADJ POS @<SC #14->13n, [,] PU @PU #15->0ntoo
[too] ADV @<ADVL #16->13n. [.] PU @PU #17->0n</s>n<ß>nI-m [I-m] <*> <unit> <ac-
sign> <heur> <idf> <nhead> N S NOM @SUBJ> #1->2nlooking [look] <mv> V PCP1 @ICL-
ADVL #2->0nforward [forward] <adir> <advl-close> ADV @<ADVL #3->2nto [to] <advl-
```

close> PRP @<ADVL #4->2nseeing [see] <vq> <v.contact> <vtk+ADJ> <mv> V PCP1 @ICL-P< #5->4nthem [they] PERS 3P ACC @<ACC #6->5nagain [again] <atemp> ADV @<ADVL #7->5n! [!] PU @PU #8->0n</ß>”
            },
            “SRL_tagging”: {
                  “SRL_tagging_set”: “HanLP,   https://hanlp.hankcs.com/en/demos/srl.html”,
                  “SRL_tagging_result”: “Today/ARG1 was/PRED (a wonderful day)/ARG2 ! I/ARG0 spent/PRED (quality time)/ARG1 (with my parents)/ARG2, and (the McDonald restaurant)/ARG1 was/PRED excellent/ARG2, too/ARGM-ADV. I/ARG0’m looking/PRED forward/ARGM-DIR (to seeing them again)/ARG1!”
            }
      }
}

Sentence 2: I’m really excited about Christmas! This year, my girlfriend and I are gonna celebrate the holiday together. We’re gonna decorate our room, and it’ll be so much fun!

{
      “meaning”: {
            “POS_tagging”: {
                  “POS_tagging_set”: “CST’s Part-Of-Speech tagger (Brill, with adaptations)”,
                  “POS_tagging_result”: ” I’m/NNP really/RB excited/VBD about/IN Christmas/NNP !/.nThis/DT year/NN ,/, my/PRP$ girlfriend/NN and/CC I/PRP are/VBP gon/VBG na/TO celebrate/VB the/DT holiday/NN together/RB ./. We’re/NNP gon/VBG na/TO decorate/VB our/PRP$ room/NN ,/, and/CC it’ll/NN be/VB so/RB much/JJ fun/NN !/.”
            },
            “NE_tagging”: {
                  “NE_tagging_set”: “HanLP,   https://hanlp.hankcs.com/en/demos/srl.html”,
                  “NE_tagging_result”: ” I’m really excited about Christmas/DATE ! This year, my girlfriend and I are gonna celebrate the holiday together. We’re gonna decorate our room, and it’ll be so much fun! “
            },
            “dependency_tagging”: {
                  “dependency_tagging_set”: “CG-dependency, https://edu.visl.dk/visl/en/parsing/automatic/dependency.php “,
                  ” dependency_tagging_result”: “n<ß>nI-m [I-m] <*> <unit> <ac-sign> <heur> <idf> <nhead> N S NOM @NPHR #1->0nreally [really] <ly> <ameta> <ADJ:real+ly> ADV @>A #2->3nexcited [excited] <np-close> ADJ POS @N< #3->1nabout [about] <pp-temp> PRP @A< #4->3nChristmas [Christmas] <*> <temp> <per> <nhead> N S NOM @P< #5->4n! [!] PU @PU #6->0n</s>n<ß>nThis [this] <*> <dem> DET S @>N #1->2nyear [year] <per> <dur> <def> <nhead> N S NOM @ADVL> #2->10n, [,] PU @PU #3->0nmy [I] <poss> <det> PERS 1S GEN @>N #4->5ngirlfriend [girlfriend] <cjt-head> <Hfam> <def> <nhead> N S NOM @SUBJ> #5->8nand [and] <co-subj> KC @CO #6->5nI [I] <cjt> <*> PERS 1S NOM @SUBJ> #7->5nare [be] <vch> <aux> V PR -1/3S @FS-STA #8->0ngonna [going=to] <complex> <aux> V PCP1 @ICL-AUX< #9->8ncelebrate [celebrate] <mv> V INF @ICL-AUX< #10->9nthe [the] <def> ART S/P @>N #11->12nholiday [holiday] <temp> <per> <def> <nhead> N S NOM @<ACC #12->10ntogether [together] ADV @<ADVL #13->10n. [.] PU @PU #14->0n</s>n<ß>nWe-re [We-re] <*> <Hmyth> <rem> <heur> <idf> <nhead> N S NOM @SUBJ> #1->3ngonna [going=to] <cjt-head> <complex> <aux> V PCP1 @FS-STA #2->0ndecorate [decorate] <v.contact> <mv> V INF @ICL-AUX< #3->2nour [we] <poss> <det> PERS GEN 1P @>N #4->5nroom [room] <Lh> <am> <def> <nhead> N S NOM @<ACC #5->3n, [,] PU @PU #6->0nand [and] <clb?> KC @CO #7->5nit-ll [it-ll] <heur> <idf> <nhead> N S NOM @SUBJ>

```
#8->9nbe [be] <cjt> <mv> V SUBJ @FS-STA #9->2nso [so] <aquant> ADV @>A #10-
>11nmuch [much] <quant> DET ABS S @>N #11->12nfun [fun] <sem-c> <percep-f> <idf>
<nhead> N S NOM @<SC #12->9n! [!] PU @PU #13->0n</ß>”
            },
            “SRL_tagging”: {
                  “SRL_tagging_set”: “HanLP,    https://hanlp.hankcs.com/en/demos/srl.html”,
                  “SRL_tagging_result”: ” I/ARG1 ’m/PRED (really excited about
Christmas)/ARG2! This year, (my girlfriend and I) /ARG0 are gonna celebrate/PRED (the
holiday)/ARG1 together/ARGM-MNR. We/ARG0 ’re gonna decorate/PRED (our room)/ARG1,
and it/ARG1 ’ll/ARGM-MOD be/PRED (so much fun)/ARG2 !”
            }
      }
}
```

Sentence 3: Today I watched a movie called ‘The Pianist.’ It was not only touching, but really absorbing, too. Now I’m feeling really happy, thanks to this memorable experience.

```
{
      “meaning”: {
            “POS_tagging”: {
                  “POS_tagging_set”: “CST’s Part-Of-Speech tagger (Brill, with adaptations)”,
                  “POS_tagging_result”: ” Today/RB I/PRP watched/VBD a/DT movie/NN
called/VBN ‘/” The/DT Pianist/NNP ./. ‘/POS It/PRP was/VBD not/RB only/RB touching/VBG
,/, but/CC really/RB absorbing/VBG ,/, too/RB ./.nNow/RB I’m/NNP feeling/NN really/RB
happy/JJ ,/, thanks/NNS to/TO this/DT memorable/JJ experience/NN ./.”
            },
            “NE_tagging”: {
                  “NE_tagging_set”: “HanLP,    https://hanlp.hankcs.com/en/demos/srl.html”,
                  “NE_tagging_result”: ” Today I watched a movie called ‘The
Pianist.’/WORK_OF_ART It was not only touching, but really absorbing, too. Now I’m feeling
really happy, thanks to this memorable experience.”
            },
            “dependency_tagging”: {
                  “dependency_tagging_set”: “CG-dependency,
https://edu.visl.dk/visl/en/parsing/automatic/dependency.php “,
                  ” dependency_tagging_result”: “n<ß>nToday [today] <*> <atemp> ADV
@ADVL> #1->3nI [I] <*> PERS 1S NOM @SUBJ> #2->3nwatched [watch] <DL:bio> <mv>
V IMPF @FS-STA #3->0na [a] <indef> ART S @>N #4->5nmovie [movie] <sem-w>
<DL:bio> <idf> <nhead> N S NOM @<ACC #5->3ncalled [call] <vtk+N> <vtk+ADJ>
<vtk+N> <vtk+PROP> <vq> <v.contact> <DL:bio> <mv> <np-close> V PCP2 PAS @ICL-N<
#6->5n-The [-The] <heur> <DL:bio> <idf> <nhead> N S NOM @<SC #7->6nPianist [Pianist]
<*> <Proper> <DL:bio> <nhead> N S NOM @<OC #8->6n. [.] PU @PU #9->0n<ß>n- [-] PU
@PU #1->0n</ß>n</s>n<ß>nIt [it] <*> PERS NEU 3S NOM @SUBJ> #1->2nwas [be]
<DL:bio> <mv> V IMPF 1/3S @FS-STA #2->0nnot [not] ADV @>A #3->4nonly [only] <ly>
<ADJ:on+ly> <advl-close> ADV @<ADVL #4->2ntouching [touching] <DL:bio> ADJ POS
@<SC #5->2n, [,] PU @PU #6->0nbut [but] KC @CO #7->5nreally [really] <ly> <ameta>
<ADJ:real+ly> ADV @ADVL> #8->9nabsorbing [absorb] <v.contact> <DL:bio> <mv> V
PCP1 @ICL-N<PRED #9->1n, [,] PU @PU #10->0ntoo [too] <advl-close> ADV @<ADVL
#11->9n. [.] PU @PU #12->0n</s>n<ß>nNow [now] <*> <atemp> ADV @ADVL #1->0nI-m
[I-m] <*> <unit> <ac-sign> <DL:bio> <heur> <nhead> N S NOM @NPHR #2->1nfeeling [feel]
<v.contact> <v-cog> <DL:bio> <mv> <np-close> V PCP1 @ICL-N<PRED #3->2nreally
[really] <ly> <ameta> <ADJ:real+ly> ADV @>A #4->5nhappy [happy] <jpsych> <DL:bio>
```

ADJ POS @<SC #5->3n, [,] PU @PU #6->0nthanks to [thanks=to] <insertion> <complex> PRP @<ADVL #7->3nthis [this] <dem> DET S @>N #8->10nmemorable [memorable] <DL:bio> ADJ POS @>N #9->10nexperience [experience] <f-psych> <percep-f> <DL:bio> <def> <nhead> N S NOM @P< #10->7n. [.] PU @PU #11->0n</ß>"
},
"SRL_tagging": {
"SRL_tagging_set": "HanLP, https://hanlp.hankcs.com/en/demos/srl.html",
"SRL_tagging_result": "Today/ARG-TMP I/ARG0 watched/PRED a movie/ARG1 called/PRED 'The Pianist.' It/ARG1 was/PRED (not only touching, but really absorbing, too)/ARG2. Now/ARG-TMP I/ARG0 'm feeling/PRED (really happy)/ARG1, thanks to this memorable experience."
}
}
}

9.6 Question Text Files

Q1: What is the tool in the picture?
Q2: What is the nickname of the person in the picture?
Q3: What is the job of the person on the left hand-side in the picture
Q4: What is the family name of the person in the centre of the picture?
Q5: What is the name of the square in the picture?

9.7 Question Speech Files

Q1.wav
Q2.wav
Q3.wav
Q4.wav
Q5.wav

9.8 Images for Question

Images for Q1	**Q1-1.jpg** **Q1-2.jpg** **Q1-3.jpg**
Image for Q2	**Q2-Joseph Gordon Levitt.jpg**
Image for Q3	**Q3-1.jpg** **Q3-2.jpg**
images for Q4	**Q4-1.jpg** **Q4-2.jpg** **Q4-3.jpg**
1 image for Q5	**Q5-1.jpg**

9.9 Meaning JSON Files

Sentence 1: What is the tool in the picture?
{
"meaning": {
"POS_tagging": {
"POS_tagging_set": "CST's Part-Of-Speech tagger (Brill, with adaptations), https://cst.dk/online/pos_tagger/uk/",
"POS_tagging_result": "What/WP is/VBZ the/DT tool/NN in/IN the/DT picture/NN ?/."

```
},
“NE_tagging”: {
“NE_tagging_set”: “CST’s named entity recogniser,
https://cst.dk/online/navnegenkenderCSTNER/uk/”,
“NE_tagging_result”: ” [What,misc,uncertain] is the tool in the picture ?”
},
“dependency_tagging”: {
“dependency_tagging_set”: “CG-dependency,
https://edu.visl.dk/visl/en/parsing/automatic/dependency.php “,
” dependency_tagging_result”: “<ß>nWhat [what] <clb> <*> <interr> INDP S/P @SC> #1-
>2nis [be] <mv> V PR 3S @FS-QUE #2->0nthe [the] <def> ART S/P @>N #3->4ntool [tool]
<tool> <def> <nhead> N S NOM @<SUBJ #4->2nin [in] <advl-fs> PRP @<ADVL #5->2nthe
[the] <def> ART S/P @>N #6->7npicture [picture] <pict> <repr> <def> <nhead> N S NOM
@P< #7->5n? [?] PU @PU #8->0n</ß>”
},
“SRL_tagging”: {
“SRL_tagging_set”: “HanLP,    https://hanlp.hankcs.com/en/demos/srl.html”,
“SRL_tagging_result”: ” What/ARG2 is/PRED (the tool in the picture)/ARG1 ?”
}
}
}
```

Sentence 2: What is the nickname of the person in the picture?

What/WP is/VBZ the/DT nickname/NN of/IN the/DT person/NN in/IN the/DT picture/NN ?/.

```
{
“meaning”: {
“POS_tagging”: {
“POS_tagging_set”: “CST’s Part-Of-Speech tagger (Brill, with adaptations)”,
“POS_tagging_result”: ” What/WP is/VBZ the/DT nickname/NN of/IN the/DT person/NN in/IN
the/DT picture/NN ?/.”
},
“NE_tagging”: {
“NE_tagging_set”: “HanLP,     https://hanlp.hankcs.com/en/demos/ner.html “,
“NE_tagging_result”: “”
},
“dependency_tagging”: {
“dependency_tagging_set”: “CG-dependency,
https://edu.visl.dk/visl/en/parsing/automatic/dependency.php “,
” dependency_tagging_result”: ” <ß>nWhat [what] <clb> <*> <interr> INDP S/P @SC> #1-
>2nis [be] <mv> V PR 3S @FS-QUE #2->0nthe [the] <def> ART S/P @>N #3->4nnickname
[nickname] <ac-cat> <def> <nhead> N S NOM @<SUBJ #4->2nof [of] <np-close> PRP @N<
#5->4nthe [the] <def> ART S/P @>N #6->7nperson [person] <H> <def> <nhead> N S NOM
@P< #7->5nin [in] <advl-fs> PRP @<ADVL #8->2nthe [the] <def> ART S/P @>N #9-
>10npicture [picture] <pict> <repr> <def> <nhead> N S NOM @P< #10->8n? [?] PU @PU
#11->0n</ß>”
},
“SRL_tagging”: {
“SRL_tagging_set”: “HanLP,    https://hanlp.hankcs.com/en/demos/srl.html”,
“SRL_tagging_result”: ” What/ARG2 is/PRED (the nickname of the person in the
picture)/ARG1 ?”
}
}
}
```

Sentence 3: What is the job of the person on the left hand-side in the picture?

```
{
“meaning”: {
“POS_tagging”: {
“POS_tagging_set”: “CST’s Part-Of-Speech tagger (Brill, with adaptations)”,
“POS_tagging_result”: ” What/WP is/VBZ the/DT job/NN of/IN the/DT person/NN on/IN the/DT left/VBN hand-side/JJ in/IN the/DT picture/NN ?/.”
},
“NE_tagging”: {
“NE_tagging_set”: ” https://cst.dk/online/navnegenkenderCSTNER/uk/”,
“NE_tagging_result”: ” [What,misc,uncertain] is the job of the person on the left hand-side in the picture ?”
},
“dependency_tagging”: {
“dependency_tagging_set”: “CG-dependency, https://edu.visl.dk/visl/en/parsing/automatic/dependency.php “,
” dependency_tagging_result”: ” <ß>nWhat [what] <clb> <*> <interr> INDP S/P @SC> #1->2nis [be] <mv> V PR 3S @FS-QUE #2->0nthe [the] <def> ART S/P @>N #3->4njob [job] <pos-soc> <sem-c> <def> <nhead> N S NOM @<SUBJ #4->2nof [of] <np-close> PRP @N< #5->4nthe [the] <def> ART S/P @>N #6->7nperson [person] <H> <def> <nhead> N S NOM @P< #7->5non [on] <advl-fs> PRP @<ADVL #8->2nthe [the] <def> ART S/P @>N #9->11nleft [left] ADJ POS @>N #10->11nhand-side [hand-side] <Lsurf> <HH> <geom> <heur> <def> <nhead> N S NOM @P< [hand-side] <heur> <def> N S NOM @P< #11->8nin [in] <advl-fs> PRP @<ADVL #12->2nthe [the] <def> ART S/P @>N #13->14npicture [picture] <pict> <repr> <def> <nhead> N S NOM @P< #14->12n? [?] PU @PU #15->0n</ß>”
},
“SRL_tagging”: {
“SRL_tagging_set”: “HanLP,    https://hanlp.hankcs.com/en/demos/srl.html”,
“SRL_tagging_result”: ” What/ARG2 is/PRED (the job of the person on the left hand-side in the picture)/ARG1 ?”
}
}
}
```

Sentence 4: What is the family name of the person in the centre of the picture?

```
{
“meaning”: {
“POS_tagging”: {
“POS_tagging_set”: “CST’s Part-Of-Speech tagger (Brill, with adaptations)”,
“POS_tagging_result”: ” What/WP is/VBZ the/DT family/NN name/NN of/IN the/DT person/NN in/IN the/DT centre/NN of/IN the/DT picture/NN ?/.”
},
“NE_tagging”: {
“NE_tagging_set”: ” https://cst.dk/online/navnegenkenderCSTNER/uk/”,
“NE_tagging_result”: ” [What,misc,uncertain] is the family name of the person in the centre of the picture ?”
},
“dependency_tagging”: {
```

“dependency_tagging_set”: “CG-dependency, https://edu.visl.dk/visl/en/parsing/automatic/dependency.php “,
” dependency_tagging_result”: ” <ß>nWhat [what] <clb> <*> <interr> INDP S/P @SC> #1->2nis [be] <mv> V PR 3S @FS-QUE #2->0nthe [the] <def> ART S/P @>N #3->5nfamily [family] <HH> <comp1> <comp1> <ncomp> N S NOM @>N #4->5nname [name] <ac-cat> <comp2> <def> <nhead> N S NOM @<SUBJ #5->2nof [of] <np-close> PRP @N< #6->5nthe [the] <def> ART S/P @>N #7->8nperson [person] <H> <def> <nhead> N S NOM @P< #8->6nin [in] <advl-fs> PRP @<ADVL #9->2nthe [the] <def> ART S/P @>N #10->11ncentre [centre] <Labs> <inst> <def> <nhead> N S NOM @P< #11->9nof [of] <np-close> PRP @N< #12->11nthe [the] <def> ART S/P @>N #13->14npicture [picture] <pict> <repr> <def> <nhead> N S NOM @P< #14->12n? [?] PU @PU #15->0n</ß>”
},
“SRL_tagging”: {
“SRL_tagging_set”: “HanLP, https://hanlp.hankcs.com/en/demos/srl.html”,
“SRL_tagging_result”: ” What/ARG2 is/PRED (the family name of the person in the centre of the picture)/ARG1 ?”
}
}
}

Sentence 5: What is the name of the square in the picture?

{
“meaning”: {
“POS_tagging”: {
“POS_tagging_set”: “CST’s Part-Of-Speech tagger (Brill, with adaptations)”,
“POS_tagging_result”: ” What/WP is/VBZ the/DT name/NN of/IN the/DT square/NN of/IN the/DT picture/NN ?/.”
},
“NE_tagging”: {
“NE_tagging_set”: ” https://cst.dk/online/navnegenkenderCSTNER/uk/”,
“NE_tagging_result”: “[What,misc,uncertain] is the name of the square of the picture ?”
},
“dependency_tagging”: {
“dependency_tagging_set”: “CG-dependency, https://edu.visl.dk/visl/en/parsing/automatic/dependency.php “,
” dependency_tagging_result”: ” n<ß>nWhat [what] <clb> <*> <interr> INDP S/P @SC> #1->2nis [be] <mv> V PR 3S @FS-QUE #2->0nthe [the] <def> ART S/P @>N #3->4nname [name] <ac-cat> <sem-c> <def> <nhead> N S NOM @<SUBJ #4->2nof [of] <np-close> PRP @N< #5->4nthe [the] <def> ART S/P @>N #6->7nsquare [square] <Lh> <geom> <def> <nhead> N S NOM @P< #7->5nof [in] <np-close> PRP @N< #8->7nthe [the] <def> ART S/P @>N #9->10npicture [picture] <pict> <repr> <def> <nhead> N S NOM @P< #10->8n? [?] PU @PU #11->0n</ß>”
},
“SRL_tagging”: {
“SRL_tagging_set”: “HanLP, https://hanlp.hankcs.com/en/demos/srl.html”,
“SRL_tagging_result”: ” What/ARG2 is/PRED (the name of the square of the picture)/ARG1 ?”
}
}
}

9.10 Intention JSON Files

Q1: What is the tool in the picture?

```
{
“Intention”:{
“qtopic”: “tool”,
“qfocus”:”What”,
“qLAT”:”tool”,
“qSAT”:”ETC”,
qdomain”:”everyday life”
}
}
```

Q2: What is the nickname of the person in the picture?

```
{
“Intention”: {
                    “qtopic”: “person”,
“qfocus”:”What”,
“qLAT”:”nickname”,
“qSAT”:”PS_NAME”,
“qdomain”:”famous people”
}
}
```

Q3: What is the job of the person on the left hand-side in the picture

```
{
“Intention”:{
“qtopic”:”person”,
“qfocus”:”What”,
“qLAT”:”job”,
“qSAT”:”CV_OCCUPATION”,
“qdomain”:”famous people”
}
}
```

Q4: What is the family name of the person in the centre of the picture?

```
{
“Intention”:{
“qtopic”:”person”,
“qfocus”:”What”,
“qLAT”:”family name”,
“qSAT”:”PS_NAME”,
“qdomain”:”famous people”
}
}
```

Q5: What is the name of the square in the picture?

```
{
“Intention”:{
“qtopic”:”square”,
“qfocus”:”What”,
“qLAT”:”square”,
“qSAT”:”LC_TOUR”,
“qdomain”:”traveling”
}
}
```