



N2647

2025/10/29

Source Requirements(PGP)

Title Use Cases and Functional Requirements: Pursuing Goals in metaverse (MPAI-PGM)
– Autonomous User Architecture (PGM-AUA) V1.0

Target MPAI-61

Abstract	1
1 Introduction	1
2 Definitions	2
3 Use Cases	3
3.1 Introduction	3
3.2 Dealing with scenes.....	3
3.2.1 Analyse context	3
3.3 A-User and H-User converse	4
3.3.1 A-User conveys message to H-User.....	4
3.3.2 Virtual Secretary.....	4
3.4 Marketing in the metaverse	5
3.4.1 Product promotion.....	5
3.4.2 Buying eyewear in the metaverse.....	6
3.5 Playing and dancing	6
3.5.1 Playing tennis in the metaverse.....	6
3.5.2 Dancing in the metaverse	7
3.6 Avatar dynamics.....	7
3.6.1 A-Avatar Walks from A to B	7
3.6.2 A-Avatar Follows a Walking Avatar.....	7
3.6.3 A- Avatars Walk Together	7
4 Functional Requirements.....	8
5 References	9
6 Annex 1 - The MPAI Metaverse Model.....	10

Abstract

This document, issued by Moving Picture, Audio, and Data Coding by Artificial Intelligence (MPAI), collects Use Cases and Functional Requirements relevant to the *planned* Technical Specification: Pursuing Goals in metaverse (MPAI-PGM) – Autonomous User Architecture (PGM-AUA) V1.0 that MPAI intends to develop using Responses to the companion [Call for Technologies: Pursuing Goals in metaverse \(MPAI-PGM\) – A-User Architecture \(PGM-AUA\) V1.0](#) [5].

1 Introduction

MPAI is an international, unaffiliated, and non-for-profit organisation established in September 2020 with the mission to develop standards for Artificial Intelligence (AI)-enabled data coding and technologies facilitating integration of data coding components into Information and

Communication Technology (ICT) systems [[The MPAI Statutes](#)]. So far, MPAI has developed 16 standards relevant to its mission covering various domains [[The MPAI Standards](#)] following a rigorous process [[The MPAI Patent Policy](#)]. Some of the domains covered are: execution environment of multi-component AI applications, portable avatar format, object and scene description, neural network watermarking, context-based audio enhancements, multimodal human-machine conversation and communication, company performance prediction, metaverse, and governance of the MPAI ecosystem. Eight Technical Specifications have been adopted by IEEE without modification and three more are in the pipeline. MPAI is executing an intense work plan [[The MPAI Workplan](#)] including several other standard projects – such as AI for Health and XR Venues. They are expected to become Technical Specifications in the next few months.

MPAI has initiated a new standard project called “Pursuing Goals in the metaverse (MPAI-PGM)” intended to enable Autonomous Users (A-User), i.e., Processes operating in a metaverse instance (M-Instance) that conforms with [Technical Specification: MPAI Metaverse Model \(MPAI-MMM\) – Technologies \(MMM-TEC\) V2.1](#) , to perform human-like activities in a portion of an M-Instance (see Annex 1 - for an introduction to MMM-TEC). The first MPAI-PGM standard project is Autonomous User Architecture (PGM-AUA).

This document contains collections of:

1. Use Cases describing realistic situations in which A-Users interact with another User that can be an A-User or an H-User, i.e., a User operating under direct human instructions.
2. Functional Requirements identifying some of the key functionalities extracted from the identified Use Cases that the A-User Architecture should support.

This document references four related documents:

1. [Call for Technologies: Pursuing Goals in metaverse \(MPAI-PGM\) – A-User Architecture \(PGM-AUA\) V1.0](#) [5] calls for technologies usable for Use Cases and supporting the Functional Requirements of this document.
2. [Framework Licence: Pursuing Goals in metaverse \(MPAI-PGM\) – Autonomous User Architecture \(PGM-AUA\) V1.0](#) [6] includes the IPR Guidelines that MPAI issues jointly to Functional Requirements.
3. [Template for Responses: Pursuing Goals in metaverse \(MPAI-PGM\) – A-User Architecture \(PGM-AUA\) V1.0](#) [7] designed to facilitate the drafting of a response.
4. [Tentative Technical Specification: Pursuing Goals in the metaverse \(MPAI-PGM\) – A-User Architecture \(PGM-AUA\) V1.0](#) [8]

2 Definitions

Terms	Definitions
A-Persona	The avatar of an A-User.
A-User	A User operating autonomously under the responsibility of a human.
Action	A standard activity performed by a Process.
AI Framework	(AIF) An environment enabling initialisation, dynamic configuration, and control of AI applications implemented as AI Workflows.
AI Modules	(AIM) A data processing element performing a Function by processing AIM-specific Inputs and producing AIM-specific Outputs.
AI Workflow	(AIW) A structured aggregation of AIMs performing a Function and receiving AIW-specific inputs and producing AIW-specific outputs.
Data Type	An instance of the Data Types defined by MPAI-AIF in 6.1.1.
Event	A logic composition of Process Action at a Time.

H-Persona	The avatar of an H-User.
H-User	A User a User operating under direct human instructions.
Process Action	The request to another Process to perform an Action on a Process or Items.
Spatial Attitude	The Position and Orientation of an object and their velocities and accelerations

3 Use Cases

3.1 Introduction

M-PAI-PGM Use Cases describe examples of realistic situations in which an Autonomous User (A-Users) - with a high degree of autonomy - interact with another User that can be an A-User or a Human User, a User directly driven by a human (H-User) in human-like activities in the M-Instance, such as moving around or conversing.

Use Cases are described with the following conventions:

- Humans have regular human names, e.g., Bob.
- Their Users have a human name possibly by A (Autonomous) or H (Human), e.g., A-Bob.
- Users' Personae (rendering of Users as avatars) have the human's name preceded by P, e.g., P-Bob.

3.2 Dealing with scenes

3.2.1 Analyse context

A-User has MM-Added its A-Persona at an M-Location containing a Scene populated by:

- Context
1. Audio Objects, inanimate Visual Objects (not resulting from the rendering of an A- or H-User), and AV Objects.
 2. A- and H-Personae.

To Interpret the Scene for its own use by MM-Capturing, Describing and Interpreting its:

- Goal
1. Audio Objects in terms of:
 - Coordinates of Source.
 - Music, sirens, noises, etc.
 - Identifier of the Audio Object ID (e.g., “All You Need is Love”)
 2. Speech Objects in terms of:
 - Coordinates of Source.
 - Speech attributes (e.g., language).
 - Text in Speech.
 - Identities of Personae uttering Speech.
 3. Visual Objects in terms of:
 - Inanimate objects identified by:
 1. Identifiers of the visual object classes (“car”)
 2. Identifiers of specific instances of visual objects in the class (“my car”).
 - Animate objects identified by:
 - Identifiers (“my friend John”)
 - Personal Statuses.

Constraints Scene analysis may be driven by specific Sub-Goals (e.g., “find my friend John”).

3.3 A-User and H-User converse

This section includes Use Cases where A-Users and H-Users converse.

3.3.1 A-User conveys message to H-User

Entities	<ul style="list-style-type: none"> - Charlie is a human seeking to sign a contract with Alice, in need to send a personal message to Alice but currently busy. - A-Bob is a Process acting as Charlie’s assistant rendered as P-Bob. - Alice is a human currently in the Metaverse Grand Square (MGS) represented by P-Alice and rendered as P-Alice.
Goal statement	<ul style="list-style-type: none"> - Bob should MM-Move B-Persona from the current place to MGS. - Should take a reasonable amount of time along an undefined Trajectory. - When at MGS, A-Bob should: <ul style="list-style-type: none"> - Look for P-Alice. - Approach P-Alice politely. - Convey the message “Charlie is expecting a response to his contract proposal”. - Close conversation politely.

3.3.2 Virtual Secretary

MPAI has already specified a Virtual Meeting Secretary as an AI Workflow (MMC-VMS) that combines AI Modules [6] having the ability to understand the conversation at a virtual meeting and draft a Summary and implemented. Unlike the currently specified Virtual Secretary who does not have the capability to generate meaningful messages, the Virtual Secretary of this Use Case should have this capability.

Entities	<ul style="list-style-type: none"> - David is a human renting virtual meeting rooms. - A-Edward is David’s Virtual Secretary (A-User) tasked to draft a report of a meeting at MLoc3. - Meeting participants are a mix of H-Users and A-Users all rendered as Personae.
Goal statement	<ul style="list-style-type: none"> - Should produce the meeting Summary. - Should behave in a human fashion but without hiding its artificiality. - Should treat H-Users and A-Users with the same attitude. - When drafting a report item, should confirm comprehension with as little disruption as possible. - Should behave appropriately for the context, i.e., it should display a Personal Status and perform Actions that are appropriate to the context. - Should comply with Users’ requests, i.e., should perform the requested Actions. - May ask clarification questions (by MM-Adding Utterances) without disrupting the discussion: - Preferable to Use Portable Avatars to communicate to an A-User (not disruptive). - Not desirable to use Portable Avatars to communicate to an H-User (disruptive). - A request for clarification by the Virtual Secretary may or may not be made public.

3.4 Marketing in the metaverse

This section introduces Use Cases related to promotion and sale of products.

3.4.1 Product promotion

Entities	<ul style="list-style-type: none"> - Fiona is a human walking around the M-Instance, possibly meeting friends. - A-Gary is a company’s User with the Goal of promoting a product. - Henry is a human walking around the M-Instance, possibly meeting friends. - A-Ian is a security company’s User having the Goal of ensuring that H-Users are not disturbed by impertinent A-Users or even H-Users. - John is a human from the security company taking over Ian’s role in difficult situations.
Goal statement	<p>A-Gary has the following Goals:</p> <ul style="list-style-type: none"> - Attract H-Users, e.g., by: - Stopping passers-by. - Attracting attention. - Intruding into conversations. - Acknowledge its artificiality. - Determine a suitable conversation opener by detecting the Personal Status and availability of passers-by. - Determine how to try again after a first unsuccessful approach. - Determine when to give up to avoid security action. - Keep an eye on the surroundings to check for approaching security A-Users (mall cops). - Apply its persuasion training to sustain conversation and convince the target H-User by dynamically changing words, intonation, face, and gestures (Personal

	Status). - Call in a company H-User or A-User for coordinated actions to re-enforce or assist. Ian has the following Goals: - Detect gatherings of people that look “suspicious” (e.g., A-Users bothering H-Users). - Apply convincing arguments to desist. - Refer to higher-level security if convincing fails.
--	--

3.4.2 Buying eyewear in the metaverse

Entities	<ul style="list-style-type: none"> - Kate is a human wishing to buy eyewear. - A-Liam is a User acting as a sales agent in an optical store. - Mary is a human taking over the contact with the customer when requested by Liam.
Goal statement	<ul style="list-style-type: none"> - Kate: to buy eyewear that suits her face in outdoor conditions. - A-Liam: to sell high-priced eyewear rather than discounted eyewear. - Mary: to take over and conclude the sale if Liam fails at the highest price.
Workflow	<ul style="list-style-type: none"> - Kate explains her needs - A-Liam places samples on the counter - Kate tries on various samples in different lights and observes the effect. - A-Liam assesses and comments on Kate’s appearance. - Kate negotiates the cost of selected eyewear. - A-Liam applies negotiation training but fails. - A-Liam calls Mary when the negotiation fails. - Mary closes the sale with Kate.

3.5 Playing and dancing

3.5.1 Playing tennis in the metaverse

Context	A virtual tennis court populated by: <ul style="list-style-type: none"> - An H-User, a Process representing a human located on a footboard and animating a humanoid simulating bones and muscles animated by sensors (MoCap?). - An A-User, an autonomous process animating the muscles and skeleton of a humanoid correctly representing a human body as muscles and a skeleton.
Goal	Both Users have the Goal of winning games and sets per lawn tennis rules.
Constraints	The movements of the human could be captured by: <ul style="list-style-type: none"> - A footboard that senses the human’s movements - generates reactions simulating those of a real surface (wood, grass etc.). - A special racket The movements of the autonomous avatar are computed by the autonomous process and used to animate the avatar. The legs of the autonomous avatar cannot exceed the assigned velocity and acceleration maxima. Each avatar holds a virtual racket that simulates the behaviour of a real racket. Similarly, the arms of the autonomous avatar cannot exceed the assigned velocity and acceleration maxima.

3.5.2 Dancing in the metaverse

Entities	<ul style="list-style-type: none"> - Angela is a dance student wearing a haptic dress. - Franz is a human running a dancing school in an M-Instance. - A-Kurt is one of Franz's A-Users with haptic capabilities and dancing know how. - Franz delegates A-Kurt to dance with Angela.
Goal statement	<p>A-Kurt should</p> <ul style="list-style-type: none"> - Dance Samba with Angela. - Lead the dance. - Notify Angela of any imperfection. - Angela should follow Kurt when dancing.

3.6 Avatar dynamics

3.6.1 A-Avatar Walks from A to B

Context	<p>An M-Instance locally populated by randomly moving</p> <ul style="list-style-type: none"> - A-Personae and H-Personae - Static and dynamic objects.
Goal	A-User should move its A-Persona "on foot" from A to B.
Constraints	<p>The A-User should not drive its Persona in ways that make it:</p> <ul style="list-style-type: none"> - Leave a specified M-Location of the M-Instance. - Collide with any Persona or Visual Object in the M-Instance. - Take not more than a time T to reach the destination M-Location.

3.6.2 A-Avatar Follows a Walking Avatar

Context	<p>An A-User is rendered</p> <ul style="list-style-type: none"> - As Persona₁ MM-Added at an M-Location with a Spatial Attitude. - In a portion of the M-Instance randomly occupied by A-Personae randomly walking in the M-Instance as in A-Avatar Walks from A to B. - Following an H-User walking in the M-Location
Goal	Of A-User: drive A-Persona to follow H-Persona.
Constraints	<p>Persona₂ shall:</p> <ul style="list-style-type: none"> - Stay at distance d from Persona₁ with a tolerance of $\pm\Delta d$. <p>Persona₂ shall not:-</p> <ul style="list-style-type: none"> - Intrude upon the personal space D of any other Persona. - Collide with any moving or static Object. - Exceed assigned maximum velocity and acceleration.

3.6.3 A- Avatars Walk Together

Context	<p>Two A-Users are rendered:</p> <ul style="list-style-type: none"> - As two Personae MM-Added at nearby M-Locations with respective Spatial Attitudes. - In a portion of the M-Instance where A-Personae and H-Personae randomly move in the M-Instance at random M-Locations.
Goal	Two A-Users intend to go to the same M-Location "walking together", i.e., by accompanying each other to reach a target M-Location.

Constraints	<p>Each A-Persona moves like any other Persona in A-Avatar Walks from A to B with the Constraints of:</p> <ul style="list-style-type: none"> - Staying at distance d from the other Persona. - With a tolerance of $\pm\Delta d$, where Δd is smaller than d, e.g., $d=1\text{m}$ and $\Delta d=50\text{ cm}$.
-------------	--

4 Functional Requirements

Note: A-User refers to an implementation of PGM-AUA.

Table 1 – List of Functional Requirements

1.	The A-User shall be implementable with limited computing resources
2.	The A-User shall be able to
1.	Understand its surroundings including
1.	Audio features
2.	Visual Features
3.	Speech features
2.	Converse with one User (either A-User and H-User) in contexts relevant to an MMM-TEC M-Instance.
3.	Move in space in a human fashion understanding its constraints.
4.	Assume (as mandated or autonomously decided).
1.	Personality
2.	Ethics
3.	Communication style
4.	Appearance
5.	Display a Personal Status coherent with its conversation partner.
3.	The A-User shall be implementable
1.	For operation in an MMM-TEC M-Instance.
2.	With relatively small Language Models.
3.	Assuming access the features available in MPAI standards.
4.	Assuming access to specialised knowledge.
4.	The A-User should
1.	Rely on existing MPAI standards or extensions thereof.
2.	Also be able to deal with more than one User at a time in contexts likely to be relevant to an MMM-TEC M-Instance.
5.	Optional functionalities
1.	The A-User could interact with more than one User at the same time.
2.	The A-Use has haptic capabilities

The main functionalities that PGM-AUA should provide are listed above in point 1 to 4. Point 5 gives examples of functionalities that are not part of the call but could be accepted for consideration if the response proposing it has sufficient maturity for the planned A-User Architecture standard.

5 References

1. MP AI; The MP AI Statutes; <https://mpai.community/about/statutes/>
2. MP AI; The MP AI Standards; <https://mpai.community/standards/>
3. MP AI; MP AI Patent Policy; <https://mpai.community/about/the-mpai-patent-policy/>
4. MP AI; The MP AI Workplan; <https://mpai.community/standards/work-plan/>
5. MP AI; [Call for Technologies : Pursuing Goals in metaverse \(MP AI-PGM\) – Autonomous User Architecture \(PGM-AUA\) V1.0.](#)
6. MP AI; [Framework Licence; Pursuing Goals in metaverse \(MP AI-PGM\) – Autonomous User Architecture \(PGM-AUA\) V1.0.](#)
7. MP AI; [Template for Respondents: Pursuing Goals in metaverse \(MP AI-PGM\) – Autonomous User Architecture \(PGM-AUA\) V1.0.](#)
8. MP AI; [Tentative Technical Specification: Pursuing Goals in metaverse \(MP AI-PGM\) – A-User Architecture \(PGM-AUA\) V1.0.](#)
9. MP AI; [Technical Specification: AI Framework \(MP AI-AIF\) V2.2.](#)

Annex A The MPAI Metaverse Model

Technical Specification: MPAI Metaverse Model (MPAI-MMM) – Technologies (MMM-TEC) V2.1 defines a metaverse instance (M-Instance) as an ICT system populated by Processes performing actions on things or other Processes and Items representing a variety of things. An important type of Process is a User representing a human either directly (called an H-User) or indirectly (called an A-User). A human subscribed to an M-Instance can deploy Users in that M-Instance that may or may not also be rendered as Personae, possibly human-like avatars. A Process can influence a part of an M-Instance called (M-Environment) by performing a [Process Action](#), defined as

- One of the 29 *Actions* specified by MMM-TEC V2.1
- Followed by a set of constructs each composed of a sequence of
 - Complement: Nil (before a direct object), At, From, To, and With.
 - Process or Item
- Followed by *IfEvent*, a logic combination of Process Actions followed by *At*

Examples of Action performed by a User are:

- MM-Add an Item, i.e., place it somewhere in an M-Instance.
- MM-Move an Item, i.e., displace it.
- MM-Capture data from the real world.
- MM-Actuate Item to the real world
- Identify data, convert data into an Identified Item.

A Process performs a Process Action if it:

- *Can perform it*, i.e., it has the necessary Capabilities, and
- *May perform it*, i.e., it has the necessary Rights, where Rights are expressed by a Deontic Expression (May, May Not, Must) followed by a set of Process Actions, each preceded by *At* Time and *IfEvent*, and by one of the adjectives Internal, Acquired, or Granted. For example, a User may speak in a specific place, and
- *Is allowed to perform it* by the Rules in force in the M-Environment, expressed by a Deontic Expression followed by a set of Process Actions, each preceded by *At* Time and *If* For example, in a private space a User is not allowed to perform any Process Action without being Granted appropriate Rights.

Internal Rights are obtained at registration time (e.g., visitor Rights), Acquired Rights are typically obtained through Transactions (e.g., by buying a ticket to a concert), and Granted Rights are given by another Process (e.g., my friend’s User gives my User Rights to enter his space).

A Process unable to perform a Process Action may request another Process to perform it on its behalf. Table 2 provides links to the specification of the general Process Action, i.e., when the Action results from the request to another Process.

Table 2 - MMM-TEC V2.1 Process Actions

Authenticate	Author	Convert	Discover	Execute	Hide
Identify	Inform	Interpret	MM-Add	MM-Animate	MM-Move
MM-Send	Modify	MU-Actuate	MU-Add	MU-Animate	MU-Move
MU-Send	Post	Property Change	Register	Resolve	Rights Change
Transact	UM-Capture	UM-Send	Validate		

A-Users can be endowed with some or all of the following Functions:

Table 3 - The Functions that an A-User can perform

Function	Description	Acronym
<u>Communication</u>	Exchanges Data with other PAAIs or humans.	Cm
<u>Goal setting</u>	Receives instructions or establishes to reach a goal.	Gs
<u>Planning</u>	Subdivides the activities to reach the goal in structured plans.	Pl
<u>Conclusion</u>	Draws a Conclusion.	Cc
<u>Decision</u>	Decides to implement the Conclusion.	Dc
<u>Action</u>	Implements the Decision, i.e, performs a Process Action.	Ac
<u>Representation</u>	Digitally represent captured information.	Rp
<u>Description</u>	Represents captured Data as Descriptors, e.g., AV Scene Descriptors.	Ds
<u>Interpretation</u>	Analyses Descriptors to obtain Interpretations, e.g., recognises speech.	In
<u>Explanation</u>	Describes the path that led to the performance of a Function.	Ex
<u>Storage/Retrieval</u>	Stores or retrieves Experiences.	Sr
<u>Learning</u>	Improves the performance of one or more Functions by the experience.	Le

Note that:

1. The Functions need not be performed in a sequential order. Communication may happen before almost all Functions; Conclusion may be reached after Planning or after Interpretation, etc.
2. Some Functions act on and produce Data of a specified format, but the way a Data instance is produced depends on the implementation of the Function.
3. Some Functions, e.g., Storage and retrieval, and Learning, are internal to the A-User and may or may not have a standard format.
4. An active A-User continuously “senses” the environment and produces Representations, but not necessarily high-order Functions, such as Descriptors and Interpretations.