Laboratório VISGRAF

Instituto de Matemática Pura e Aplicada

MMSW: From Multimodal Story Worlds to Networked Embodiment

Robotics, Digital Twins, and Cross-Modal Human–Synthetic Interaction

Matteo Moriconi and Luiz Velho

Technical Report TR-26-03 Relatório Técnico

October - 2026 - Outubro

The contents of this report are the sole responsibility of the authors.
O conteúdo do presente relatório è de ùnica responsabilidade dos autores.

MMSW: From Multimodal Story Worlds to Networked Embodiment

Technical Report on Robotics, Digital Twins, and Cross-Modal Human–Synthetic Interaction

Matteo Moriconi
VFXRio

Luiz Velho
VISGRAF Laboratory / IMPA

October 2, 2026

Abstract

This technical report documents a new milestone in the evolution of Multicasting Multimodal Story Worlds (MMSW). Earlier reports described the transition from a minimal WebSocket reflector to a graph-aware runtime for multimodal narrative generation, persistent memory, tools, telephony, and embodied interfaces. The v59 architecture extends that runtime into networked embodiment: robotics is represented as a first-class participant in the story-world graph, with explicit command, feedback, observation, ownership, verification, transport, and recovery semantics. Persistent duplex WebSockets remain the default real-time path, while HTTP compatibility and optional MQTT, serial, and ROS 2 integration boundaries allow heterogeneous deployments without moving identity, routing, and evidence semantics out of the MMSW layer. Digital twins provide a controlled bridge between conversational intent and simulated or commissioned robotic action. The report also analyzes *Aurora Circuit*, an exploratory human–synthetic interaction case in which Matteo Moriconi interacts with *Celestia Borealis*, a synthetic non-human entity acting through an extended simulated Franka Panda arm. A bounded motor lexicon is composed into a nine-action movement phrase, stabilized through naming, externalized as a verified 62-stroke drawing, and framed through a musical association. Together, the architecture and case study show a transition from multimodal orchestration toward a distributed world-management runtime in which language, action, perception, evidence, and cross-modal representation can participate in one persistent computational world while remaining operationally separated.

Keywords: MMSW; story worlds; multimodal AI; embodied AI; robotics; digital twins; human–synthetic interaction; cross-modal representation; distributed systems; WebSocket; MQTT; computational creativity.

Contents

1 Introduction **3**

2 Continuity with the Previous MMSW Technical Reports **3**

3 MMSW v59 as a World-Management Runtime **5**
3.1 Robotics as a first-class graph participant 5
3.2 Architecture flow 6

4 Protocol-Plural Distribution and Hybrid Transport **6**

5 Command, Evidence, Ownership, and Safety **7**

6 Digital Twins as an Embodiment and Verification Layer **8**

7 From Multimodality to Cross-Modal Embodiment **8**

8 Case Study: Aurora Circuit **9**
8.1 Research position and role separation . . . 9
8.2 Motor lexicon and study design . . . 9
8.3 Phase I: composing a movement phrase . . . 9
8.4 Phase II: description and naming . . . 10
8.5 Phase III: visual externalization . . . 10
8.6 Phase IV: sonic framing . . . 10
8.7 Observed results . . . 10
8.8 Interpretive framework . . . 10

9 Updated Implementation Metrics **11**

10 Milestone: From Multimodal Orchestration to Networked Embodiment **12**

11 Research Implications **12**
11.1 Embodiment as a graph property . . . 12
11.2 Semantic continuity independent of transport . . . 12
11.3 Digital twins as epistemic intermediaries . . . 12
11.4 Conceptual invariants across modalities . . . 12

12 Limitations and Claims Boundary **13**

13 Future Work **13**
13.1 Persistent executable concepts . . . 13
13.2 Affordance-aware planning and composition . . . 13
13.3 Cross-modal evaluation . . . 14
13.4 From digital twin to commissioned hardware . . . 14

14 Conclusion **14**

A MMSW v59 Logical Stack **14**

B Metrics Methodology **14**

C Source Corpus for the October 2026 Milestone **15**

1 Introduction

Multicasting Multimodal Story Worlds (MMSW) was introduced as a computational framework for maintaining interactive narrative worlds across human participants, artificial agents, generative media, memory systems, tools, and networked clients. The two earlier VISGRAF technical reports established complementary parts of that trajectory. TR-26-01 measured the engineering growth of the runtime from its original WebSocket prototype, while TR-26-02 formalized MMSW as a computational framework for multimodal narrative generation and documented the reference runtime, graph routing, model orchestration, memory capsules, plugins, and embodied interfaces [**tr2601**, **tr2602**].

The milestone documented here changes the status of embodiment. Robotics is no longer only an external interface on the conceptual horizon of MMSW. In v59 it has explicit runtime structure: robot identities, adapters, cameras and logical components; a bounded robotics protocol; control ownership; command and task records; observations and scene state; verification; recovery; and transport options. The resulting system can transform conversational or studio intent into grounded robotic work while preserving a separation between interpretation, execution, evidence, and narration [**v59docs2026**].

A second development is methodological. The robotics substrate is not considered only as an engineering extension. It creates a setting in which a persistent synthetic entity can act through a bounded embodied system and in which the consequences of action can be measured. The *Aurora Circuit* study uses this setting to ask how a constrained motor vocabulary can become a higher-order movement phrase and how that phrase can subsequently be re-described in language, drawing, and music [**auroralog2026**].

This report therefore treats the new milestone at two levels: **architecture**, where MMSW expands from multimodal coordination to networked embodiment; and **interaction**, where the embodied runtime becomes a platform for research on human–synthetic co-creation and cross-modal continuity.

2 Continuity with the Previous MMSW Technical Reports

The new report should be read as a continuation rather than a replacement of TR-26-01 and TR-26-02. The earlier runtime already contained the central thesis that a story world is maintained by more than a model context: shared state, graph routing, memory, asynchronous communication, multimodal pipelines, tools, and receipt-backed actions are independent parts of the computational environment. Version 59 preserves these principles while extending them to action in simulated and potentially physical space.

Dimension	Earlier documented MMSW	MMSW v59 — robotics milestone
Primary system role	Distributed multimodal story-world runtime coordinating humans, agents, memory, media, tools, and shared state.	Distributed multimodal *and embodied* runtime in which robotics becomes a first-class graph participant.
Embodiment	Embodied interfaces and robotics were part of the conceptual/deployment horizon.	Explicit robotics protocol, adapters, robot components, ownership, control leases, queues, task records, observation, verification, and recovery.
Action boundary	Plugin/tool actions were separated from conversational speech and confirmed through orchestration receipts.	Natural-language intent is translated into bounded robot commands; execution, evidence, terminal state, and narration remain explicitly separated.
Network transport	Multicast runtime centered on WebSocket communication with HTTP-facing services and plugin interfaces.	WebSocket remains the default while HTTP compatibility, optional MQTT hybrid transport, serial bridges, and ROS 2 integration boundaries allow heterogeneous endpoints to coexist.
Physical/simulated world	External devices and embodied media could be connected to the story-world runtime.	Digital twins become an execution and validation layer for Panda and Waveshare-class robots, with physical commissioning kept distinct from simulation claims.
Perception and evidence	Multimodal perception and external-status checks grounded media/tool responses.	Camera observations, scene state, telemetry, execution feedback, and terminal records create an evidence path from embodied action back into the shared world.
AI Entity relationship	Persistent identity and multimodal expression supported the transition from bots toward agents/entities.	A robotic subsystem can function as a networked extension of an AI Entity's embodied presence while remaining independently controlled, observable, and verifiable.
Repository scale*	114 files; 34,733 raw LOC; approximately 27,150 source LOC; 44 unique routes; 158 async functions.	339 measured source files; 150,583 raw source lines; 127 unique decorated HTTP/WebSocket method–path pairs; 458 async Python functions. Robotics accounts for 156 source files and 24,115 raw source lines in the stated measurement scope.

Table 1: MMSW architectural milestone: progression from the earlier documented multimodal runtime to the robotics-capable v59 architecture.

*The repository figures use different measurement scopes and are included to illustrate engineering expansion, not to claim performance, safety, scientific validity, or direct productivity improvement.

3 MMSW v59 as a World-Management Runtime

MMSW treats the story world as a shared computational environment whose continuity cannot be delegated to a single generative model. A model can propose language, media, or an interpretation of intent, but the runtime determines which participants are present, what world state is available, what memories can be retrieved, which graph routes are active, which external capability may act, and what evidence is sufficient to update the world after an action.

The robotics extension makes this separation especially important because a linguistically plausible response is not equivalent to a feasible or completed physical action. The v59 runtime therefore distributes responsibility across layers.

Table 2: Functional layers of MMSW v59.

Layer	**Primary responsibility**	**Representative mechanisms**
World	Shared narrative and operational context	room state, participants, environment state, persistent records
Graph	Selective communication and ownership	directed links, private audiences, component ownership, target resolution
Memory	Continuity beyond one model window	working memory, persistent capsules, transcript history
Multimodal	Perception and media transformation	image pipelines, camera observations, voice, rendering
Action	External execution	plugins, tooling nodes, robot adapters, device services
Verification	Evidence-backed world change	receipts, observations, measurements, terminal records, task ledger
Embodiment	Mapping entities into simulated or physical bodies	robot parent, arm/gripper components, camera, digital twin, hardware bridge

3.1 Robotics as a first-class graph participant

The robotics extension preserves MMSW's graph logic rather than bypassing it. A canonical robot can be decomposed into runtime roles that can be discovered, routed, observed, and governed independently. The conversational parent preserves identity and dialogue; the adapter owns motion and task execution; the camera publishes observations; logical arm and gripper nodes expose component structure without forcing each component to become a conversational actor.

Table 3: Canonical decomposition of a robot participant.

Runtime node	Role	Why it matters
`robot-arm`	Regular MMSW client with Persona/Mind and conversation	Keeps embodiment connected to identity, memory, dialogue, and audience.
`robot-arm.adapter`	Motion controller and task executor	Owns validation, motion, cancellation, queue/lease state, and durable outcomes.
`robot-arm.camera`	Sensing and image producer	Returns observations through MMSW rather than an untracked side channel.
`robot-arm.arm` / `.gripper`	Logical embodiment components	Expose ownership and component structure for reasoning and monitoring.
requesting client	Human, synthetic entity, or Studio controller	Provides authenticated runtime identity and graph-scoped intent.

3.2 Architecture flow

Figure 1 summarizes the control and evidence loop. The crucial feature is that language does not directly become motion. Interpretation creates a bounded command; the runtime establishes requester identity and ownership; the adapter validates and executes; observations and terminal records return as evidence; only then can the conversational layer narrate the outcome.

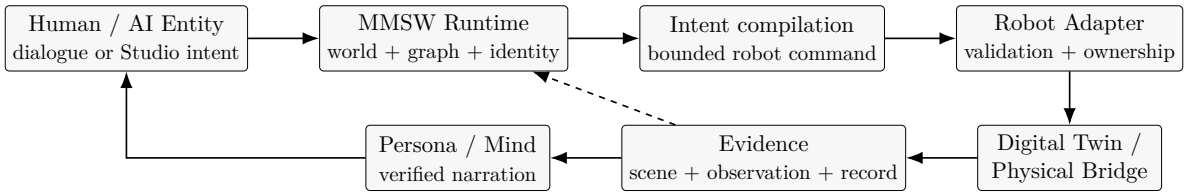

Figure 1: MMSW v59 separates conversational intent, runtime routing, bounded command compilation, robot execution, evidence, and narration.

4 Protocol-Plural Distribution and Hybrid Transport

The networking milestone in v59 is not merely the addition of another protocol. It is the separation of MMSW's semantic contract from any one transport. Identity, graph routing, command validation, task ownership, evidence, and replay rules remain runtime-level concerns. WebSocket, HTTP, MQTT, serial, or ROS 2 can serve different operational legs without redefining the world model [**v59docs2026**].

Persistent duplex WebSockets remain the default real-time robotics path. Existing HTTP-facing services remain compatible with plugins and ordinary clients. In the optional hybrid mode, MMSW continues to perform command validation and graph semantics while a broker carries robot-side command and telemetry traffic. The browser does not need broker awareness. In this design, command idempotency is a property of the MMSW task ledger rather than a consequence of broker delivery guarantees.

Table 4: Default and optional transport paths in the v59 architecture.

Traffic	**Default path**	**Optional / integration path**
Browser registration, controls, receipts, presence	MMSW `/ws` WebSocket	MMSW `/ws` WebSocket
Robot identity and graph liveness	MMSW `/ws` WebSocket	MMSW `/ws` WebSocket
Validated command to adapter	MMSW WebSocket	MMSW → MQTT → adapter
Scene, images, and robot feedback	adapter/camera → MMSW WebSocket	adapter/camera → MQTT → MMSW
Delivery to viewers / conversational clients	MMSW WebSocket	MMSW WebSocket
Existing plugin / ordinary-client calls	HTTP API	HTTP API
Commissioned budget-arm boundary	adapter boundary	USB serial, hardware-specific
Industrial/research robotics boundary	adapter boundary	ROS 2 trajectory transport building block

Broker or network resilience is deliberately prevented from becoming ambiguous physical replay. If an explicitly selected transport becomes unavailable, the system should fail visibly rather than silently switch semantics. Active or queued work is not automatically replayed after reconnect. This design principle is important whenever a distributed system controls effectors rather than only media.

5 Command, Evidence, Ownership, and Safety

Robotics messages use a bounded event contract rather than free-form natural language at the execution boundary. Natural-language requests can be interpreted into an action plan, but execution proceeds through structured packets carrying target identity, command IDs, timestamps, ownership, state, and terminal outcomes. The runtime distinguishes at least commands, feedback, durable terminal records, expression profiles, observation subscriptions, observations, and scene state.

A successful network delivery is not considered physical completion. Completion requires a terminal task record and the evidence available for the deployment. Duplicate command IDs can return a previous outcome without executing motion again, whereas conflicting reuse of an ID is rejected. This is the embodied analogue of MMSW's broader rule that speech, external action, and evidence should not be collapsed into one generative response.

Embodied execution also introduces failure modes that do not exist in ordinary chat. The supplied v59 design therefore includes one-task-at-a-time execution, controller leases, rejection of conflicting controllers, explicit stop/takeover semantics, bounded queues, and reconnect behavior that avoids automatic motion replay. Physical commissioning and local emergency-stop mechanisms remain outside the language model and remain platform-specific requirements [**v59docs2026**].

- No automatic replay of interrupted physical motion after reconnect.
- No silent transport fallback when an explicitly requested robotics transport is unavailable.
- Task completion requires evidence and a terminal record, not only a plausible conversational statement.
- Public scene streams should not expose private task text or audience metadata.
- Capability and whole-plan gates can reject unsupported plans before preliminary motion occurs.
- Internet-facing authentication and robot-specific hardware safety require deployment-specific commissioning beyond the trusted-node research runtime.

6 Digital Twins as an Embodiment and Verification Layer

Digital twins are used in v59 as more than visual replicas. They expose the kinematic and scene state needed by the MMSW world model and provide an execution environment in which high-level intent can be planned, observed, and verified. This allows a synthetic entity to acquire a robotic extension without turning the conversational model into a low-level motion controller.

The supplied release distinguishes simulation-tested paths from integration or commissioning boundaries. That distinction is central to the report's claims.

Table 5: Robot models and integration roles documented in the v59 working paper.

Model / connection	**Current role**	**Research significance**
Franka Panda simulation	Seven-axis PyBullet arm with manipulation, gestures, writing/drawing, and numerical task verification	Primary executable research path for grounded embodied experiments.
Waveshare RoArm-M2 simulation	Lower-DOF articulated model with bounded manipulation subset	Tests capability-aware planning across heterogeneous morphology.
Waveshare RoArm-M3-S simulation	Five positioning axes plus gripper, with stacking, oriented grasps, and drawing support	Extends experiments toward a lower-cost morphology while preserving MMSW semantics.
Physical Waveshare bridge	Read-only by default, with bounded commissioned motion/feedback path; autonomous contact tasks remain hardware-specific	Makes the adapter boundary between network semantics and physical commissioning explicit.
ROS 2 transport	Optional trajectory/service transport building block, not a generic commissioned RobotAdapter	Defines a path toward MoveIt/`ros2_control` ecosystems without claiming universal hardware readiness.

A digital twin is a functional research proxy only to the extent that its geometry, kinematics, sensing assumptions, and dynamics correspond to the intended physical platform. Simulation ground truth and image-based perception must therefore remain explicitly labeled rather than treated as interchangeable evidence.

7 From Multimodality to Cross-Modal Embodiment

The robotics extension also changes what multimodality means inside MMSW. Earlier multimodality largely connected text, voice, images, avatars, rendered media, and external tools. The embodied path adds a closed loop in which modalities can become simulated or physical action and then return as new perceptual evidence.

1. A human or synthetic entity expresses an intention in language or Studio control.
2. MMSW resolves runtime identity, graph audience, and target embodiment.
3. An interpretation layer compiles the intention into a bounded robotics task.
4. The adapter validates freshness, ownership, capability, and feasibility before motion.
5. The digital twin or commissioned bridge executes while scene/camera telemetry returns through MMSW.
6. Verification produces a durable outcome with available measurements and observations.
7. The entity's normal conversational layer produces a response grounded in the terminal result.

This loop is cross-modal in an operational sense: semantics travel across language, graph events, structured action, spatial state, imagery, kinematics, evidence, and back into language.

The modalities are connected by explicit state transformations rather than by asking one model to simulate every intermediate step in text.

8 Case Study: Aurora Circuit

8.1 Research position and role separation

Aurora Circuit is an exploratory human–synthetic interaction case. Matteo Moriconi is the human participant. *Celestia Borealis* is the synthetic non-human entity with which the human interacts. The simulated robotic arm is treated as Celestia's embodied extension rather than as an independent conversational agent. The environment supplies the table-top workspace, drawing region, geometric constraints, and measurable simulator state [**auroralog2026**].

This terminology is deliberately operational. It does not assert consciousness, personhood, subjective body ownership, or human-equivalent autonomy. The study asks a narrower question: what forms of organization and meaning can emerge when an open-ended human prompt is interpreted by a synthetic entity, composed through a bounded executable action vocabulary, externally realized, and then re-described across modalities?

8.2 Motor lexicon and study design

The recorded run used a simulated Panda arm and a bounded SDK. For the creative movement sequence, the verified plan drew on `rotate_joint`, `rotate_tool`, `point_pose`, `dance`, and `gesture`. Drawing used a pen-trace capability over a bounded paper region. The simulator returned state-based verification including joint-angle checks, orientation error, position error, and pen-path error.

The study is a qualitative single-case demonstration rather than a controlled experiment. It contains no participant sample, comparison condition, blinded evaluation, or psychometric creativity measure. Three evidence types are kept separate: dialogue evidence, planner/action evidence, and simulator verification.

8.3 Phase I: composing a movement phrase

The key human instruction was open-ended: create non-basic "crazy moves" and do not ask permission for each decision. The interpreted plan contained nine actions [**auroralog2026**]:

1. rotate joint 1 by $+35°$;
2. rotate joint 4 by $-25°$;
3. rotate the tool by $+90°$ about the world z axis;
4. execute an upward point pose at speed scale 1.4;
5. perform a bounded single-arm Macarena adaptation;
6. execute a leftward point pose at speed scale 1.3;
7. rotate joint 1 by $-55°$;
8. execute a front-facing point pose at speed scale 1.5;
9. finish with a `wave_hello` gesture.

The simulator reported successful completion and verification of the nine planned actions. The logged $90°$ tool rotation had a reported orientation error of $0.40°$, and the recorded positional checks were at sub-millimeter scale. The importance of the episode is not that predefined primitives exist, but that an open human instruction was organized into a temporal phrase within those constraints.

8.4 Phase II: description and naming

After execution, Celestia described the sequence as a "kinetic call-and-response" and later as "more choreography than command list." When asked for a name, the sequence became **Aurora Circuit**. The name functions as a compact symbolic handle for a temporally extended action. The broader interaction also demonstrated an important distinction: conceptual stabilization does not guarantee exact procedural replay. Exact reproducibility requires storage of the low-level action trace, parameters, and constraints.

8.5 Phase III: visual externalization

The concept was then shifted from movement to drawing. The final verified drawing routine produced a circuit-like spiral in **62 pen strokes**. The simulator reported a **1.5 mm pen-path error** and successful trace verification [**auroralog2026**]. The research interest is the relation between stages: the movement phrase was named; the name became a promptable conceptual object; and a visual artifact was produced under the same concept.

Following Goodman's distinction between symbol systems, this should be treated as re-description rather than literal transcription [**goodman1976**]. A drawing does not need to encode every joint trajectory in order to preserve selected relational qualities of the developing concept.

8.6 Phase IV: sonic framing

When asked what music should accompany the work, Celestia proposed Arvo Pärt's *Spiegel im Spiegel*. The musical work was not generated by the system and no claim of musical authorship is made. Its role is interpretive: an existing work is selected as a sonic frame associated with pacing, recurrence, restraint, and negative space.

The sequence movement → name → drawing → music therefore represents cross-modal framing rather than lossless translation. What may persist are selected relations—for example spiral, recurrence, opening and closure, alternation, restraint, and negative space—rather than a single literal code.

8.7 Observed results

Table 6: Observed outcomes in the Aurora Circuit case study.

Phase	**Observed outcome**	**Evidence status**
Movement composition	Nine-action sequence assembled from bounded primitives: joint rotations, tool rotation, directional poses, bounded dance, and wave gesture	Action list recorded; simulator reported completion and verification.
Naming	Sequence described as a kinetic call-and-response and named *Aurora Circuit*	Dialogue evidence; conceptual stabilization only.
Drawing	Circuit-like spiral produced in 62 pen strokes	Simulator trace verification; reported 1.5 mm pen-path error.
Music	*Spiegel im Spiegel* proposed as accompaniment	Interpretive selection; no generated music and no objective equivalence claim.

8.8 Interpretive framework

The case can be read through several complementary academic lenses without treating any one theory as an explanation of the system. Wittgenstein's later account of meaning through use

is useful for distinguishing a primitive's isolated command semantics from the role it acquires inside a temporal practice [**wittgenstein1953**]. Gibson's affordances provide a vocabulary for the relation between available action and environmental/embodied constraints [**gibson1979**]. Laban provides a precedent for treating movement as structured and composable material [**laban1966**]. Goodman helps distinguish cross-modal re-description from identity between media [**goodman1976**]. Boden's framework makes it possible to describe the case cautiously as combinational and exploratory behavior inside a constrained possibility space rather than as proof of unrestricted machine creativity [**boden2004**].

9 Updated Implementation Metrics

The earlier runtime report used repository metrics to show that MMSW had evolved from a small WebSocket prototype into a medium-high-complexity real-time system. The v59 working paper reports a substantially larger snapshot in which robotics is a material part of the executable source. The updated scan covers `vfxrio/` plus `scripts/`, excluding generated reports, caches, vendored static libraries, and avatar-cache material. Raw lines are physical source lines; Python functions/classes were counted from the abstract syntax tree; and the route count uses unique decorated method/path pairs visible in the measured Python source [**v59docs2026**].

Table 7: Updated v59 implementation snapshot reported in the working paper. These are repository descriptors, not quality or performance metrics.

Metric	**v59 snapshot**	**Interpretation**
Source files in measured scope	339	Application/runtime source across Python, HTML, JavaScript, CSS, and shell.
Raw source lines	150,583	Physical lines in the measured source scope.
Nonblank source lines	138,929	Raw lines excluding blank lines.
Python files	295	Dominant implementation language.
Python functions / methods	3,487	AST-counted synchronous and asynchronous definitions.
Async Python functions	458	Large asynchronous surface consistent with real-time distributed I/O.
Python classes	158	Adapters, runtime services, planners, models, and support components.
Decorated HTTP/WebSocket routes	127 unique method/path pairs	Source-level count across server and local client APIs.
Robotics-related source files	156	Files under `vfxrio/robotics` plus robotics-named source/scripts.
Robotics raw source lines	24,115	Approximately 16% of measured raw source lines.
Robotics Python functions / methods	963	Planning, adapters, simulation, transport, validation, and evidence logic.
Robotics async Python functions	139	Real-time transport/control and streaming work.
Robotics Python classes	59	Robot models, transports, controllers, planners, and protocol support.

These figures should not be interpreted as a direct quality metric or a strict like-for-like benchmark against the earlier report, because the measurement scopes differ. Their value is descriptive: the additional complexity is located in concurrent state, scene streaming, transport sessions, queues and leases, image handling, simulator/hardware abstraction, planning, collision/affordance reasoning, recovery, and verified terminal records.

10 Milestone: From Multimodal Orchestration to Networked Embodiment

The transition from the earlier MMSW runtime to v59 is best understood as an expansion of architectural scope. The same principles of selective routing, shared world state, persistent memory, multimodal generation, orchestrated action, and verification remain intact. What changes is the domain in which those principles can operate.

Milestone summary. Table 1, presented near the beginning of the report, provides the direct comparison between the earlier documented MMSW and the v59 robotics milestone. The remainder of this section interprets that transition in architectural terms.

The milestone is therefore not simply the attachment of a robotic arm to an AI system. High-level intent can now be transformed into bounded commands, transported through default or hybrid network paths, executed against digital twins or commissioned integration adapters, observed through scene state and sensors, verified through structured records, and reintegrated into the ongoing story world.

11 Research Implications

The new architecture and case study support several research directions while remaining within the evidence of the supplied materials.

11.1 Embodiment as a graph property

A robotic arm can be treated as an owned runtime component of a persistent synthetic identity rather than as a detached automation script. Sensing, action, and evidence can be routed through the same graph that carries dialogue, memory, and media. This provides a technical basis for studying persistent entities that act through different simulated or physical embodiments.

11.2 Semantic continuity independent of transport

WebSocket, HTTP, MQTT, serial, and ROS-oriented boundaries can play different operational roles while identity, audience, command IDs, verification, and world state remain common at the MMSW layer. This suggests that the continuity of a distributed embodied system can be defined above the transport layer.

11.3 Digital twins as epistemic intermediaries

A digital twin can ground language in spatial state, execute bounded actions, expose consequences, and return measurable evidence before a conversational layer claims that an embodied event occurred. In this sense the twin is not only a simulator; it is an intermediary between semantic intent and evidence-backed world updates.

11.4 Conceptual invariants across modalities

The *Aurora Circuit* progression suggests a testable question for future work: whether specified relational qualities survive transformation across movement, language, drawing, and sound.

Rather than asking only whether a system can output multiple modalities, experiments can ask whether independent observers recognize common structural properties across them.

12 Limitations and Claims Boundary

The current milestone is substantial, but the claims must remain bounded.

1. **Simulation is the primary verified arm path.** The delivered research evidence is strongest in PyBullet simulation. Real-world contact uncertainty, calibration drift, sensing noise, and hardware safety are not represented at full physical fidelity.
2. **ROS 2 is an integration boundary, not a universal commissioned adapter.** Robot-specific URDF/SRDF, drivers, planning scenes, gripper interfaces, calibration, and stop behavior remain deployment work.
3. **Physical Waveshare paths remain bounded.** Some contact/manipulation functions are deliberately uncommissioned or hardware-specific.
4. **Trusted-node identity is not Internet-grade authentication.** External deployment requires additional security and operational controls.
5. **Evidence types are not interchangeable.** Simulator ground truth, image-based detection, and physical-world measurement should remain separately labeled.
6. **Aurora Circuit is a single case.** It cannot support population-level claims or a general measure of machine creativity.
7. **The action vocabulary is bounded.** The case demonstrates organization and exploration within predefined motor primitives, not invention of new physical laws or motor primitives.
8. **Naming is not serialization.** A semantic label can stabilize a concept without storing the exact action program required for reproducible replay.
9. **No independent creativity assessment was performed.** The work contains no blinded judges, standardized creativity metric, or historical-novelty baseline.
10. **Anthropomorphic language is interactional.** First-person references to an arm or body do not constitute evidence of consciousness, sentience, personhood, or subjective body ownership.

13 Future Work

Four directions follow directly from the observed limitations.

13.1 Persistent executable concepts

Named concepts such as *Aurora Circuit* should resolve to versioned executable traces containing operator names, parameters, constraints, provenance, and verification history. This would permit exact replay, controlled variation, and comparison between versions.

13.2 Affordance-aware planning and composition

Rather than generating a complete movement plan and validating it only afterward, future planners should search over currently available affordances: reachable grasps, free approach corridors, support surfaces, held-object state, collision-free orientations, and safe drawing regions. Physical feasibility can then become part of the composition process itself rather than only a post-hoc filter.

13.3 Cross-modal evaluation

Future studies can test whether participants match unlabeled outputs across movement, drawing, language, and sound or rate them on shared qualities such as rhythm, density, symmetry, tension, closure, and restraint. This would turn cross-modal continuity into an experimentally measurable hypothesis.

13.4 From digital twin to commissioned hardware

The architecture should be evaluated under controlled transfer from simulator to real robot, with explicit calibration, hardware stop behavior, sensing uncertainty, safety envelopes, and comparison between simulated evidence and physical measurements. The goal is not to erase the simulation/physical distinction, but to measure it.

14 Conclusion

MMSW v59 advances the framework from distributed multimodal narrative infrastructure toward a distributed runtime for multimodal and embodied AI. Its defining contribution is architectural rather than rhetorical: conversational identity, graph routing, persistent memory, perception, media generation, robotic action, telemetry, and verification can be coordinated inside one world-management system while remaining separated into accountable components.

The *Aurora Circuit* case demonstrates why this matters for research. A human prompt can be interpreted by a persistent synthetic entity, compiled into a bounded motor composition, executed and verified through a digital twin, stabilized through language, and re-described through visual and sonic frames. The strongest conclusion is not that the system possesses unrestricted creativity or autonomous artistic intention. It is that MMSW now provides an inspectable pathway from language to embodied action and from embodied action back to evidence and cross-modal representation.

The resulting research direction is a form of hybrid computational agency distributed across humans, synthetic entities, memory, graph permissions, planners, robot bodies, sensors, digital twins, media systems, and verification. Under this model, a robotic arm is neither a metaphor generated in language nor an opaque peripheral. It is a networked component of the world whose actions can be routed, constrained, measured, verified, and reintegrated into the continuity of the story world.

A MMSW v59 Logical Stack

MMSW v59 – logical stack
Narrative / Entity layer – Persona · Mind · dialogue · memory · world continuity
Orchestration layer – graph routing · audiences · intent · guardrails · one-speaker logic
Multimodal layer – voice · image · avatars · camera perception · rendering
Embodiment layer – robot parent · adapter · camera · arm/gripper components · digital twins
Execution layer – planning · skills · affordances · control leases · queues · task ledger
Transport layer – WebSocket default · HTTP compatibility · MQTT hybrid · serial / ROS 2 boundaries
Evidence layer – observations · scene state · measured results · terminal records · verified narration

B Metrics Methodology

The v59 source metrics reproduced in this report were computed in the supplied working paper from a scan over `vfxrio/` and `scripts/`. The scan included Python, HTML, JavaScript, CSS, and shell files and excluded generated session-summary HTML, vendored static JavaScript,

`__pycache__` material, and avatar-cache data. Python functions and classes were counted from the abstract syntax tree. Route counts used unique decorated method/path pairs visible in the measured Python source. Robotics scope included files under `vfxrio/robotics` and source/script files whose names contain robotics. The counts are reproducible repository descriptors, not runtime-performance, test-coverage, safety, or scientific-validity measures.

C Source Corpus for the October 2026 Milestone

The present report consolidates the following supplied source material:

- Moriconi and Velho, *MMSW Runtime: Complexity Metrics and Evolution from a WebSocket Prototype*, TR-26-01, March 2026.
- Velho and Moriconi, *MMSW: A Computational Framework for Multimodal Narrative Generation*, TR-26-02, March 2026.
- Moriconi and Velho, *Multicasting Multimodal Story Worlds (MMSW): A Distributed Computational Framework for Realtime Narrative, Networked Embodiment, and Robotics*, working draft, September–October 2026.
- Moriconi and Velho, *Aurora Circuit: From a Bounded Motor Lexicon to Cross-Modal Creative Form in Human–Synthetic Interaction*, working paper, September–October 2026.
- MMSW v59 internal robotics documentation cited in the working draft: protocol, transport, digital twin, embodiment, ROS 2, network start, skills, affordances, reasoning, test-report, and validation/changelog material.
- Internal interaction and simulator trace for the *Aurora Circuit* demonstration.

Author Contributions

Matteo Moriconi: concept development, system experimentation, interaction design, creative direction, documentation, and manuscript preparation.
Luiz Velho: research supervision, conceptual framing, scientific discussion, and manuscript review.