

PROJECT: INSTITUTION

LECTURE: Guangzhou, 2026

SHEETS PAPER LAB PACK

Art as Simulation, Exhibition as Path

An explainer for the Guangzhou lecture on Institution's virtual museum, ALAN, Atrium, physics as medium, the physical building and world models. October 2026.

The sheets

32 sheets with speaker notes. Arrow keys or swipe.

The pavilion lab

Say how the room should hang. ALAN edits the rule sheet; the simulator checks it.

The pack

Paper and sheets as PDF, the lab to run offline, for review.

Abstract

An artwork placed in a pavilion of Institution's virtual museum lives inside a simulation: a Unity scene computes gravity, collision, light and sound for it, for the visitors around it and for any other body in the room. Today those rules are Earth's and nobody chose them. We treat the rules as part of the work. They are written on the pavilion's data sheet beside lighting and materials, in a schema of eighteen fields, and executed by an engine; the curator model ALAN edits them from a sentence and states in advance what it expects to happen, and the simulator scores that prediction. The second half of the claim is that an exhibition is a path. A visitor leaves a sequence of positions, dwell times and gaze targets; a robot leaves a trajectory; a world model produces a rollout. Under one event schema these are the same object, so a pavilion can test all three subjects under the same edited physics, and the physical building, which the concept by Clouds AO already treats as a mechanism, can enact the same rules with drones, a reconfigurable hall and a dark level. We describe the virtual museum, ALAN, Atrium, the live interpretation service and the physics lab, then the sensing

programme, the building, the commissions, the drone constellation and three experiments. Measured results include a benchmark of one robot task under eleven rule sheets, in which a forecast that assumes Earth is 6.36 m wrong in a room turned upside down, and the audit that found language-model fingerprints in our own fine-tuning data.

1. Introduction

The argument has two halves, art as simulation and exhibition as path.

1.1 Art as simulation

A work shown in Institution's virtual museum is placed in a pavilion, and a pavilion is a Unity scene. The scene computes, sixty or more times a second, where every body is, what it touches, how it is lit and what it sounds like. That computation is a simulation in the plain sense: a set of rules, applied to a set of bodies, advanced in time. The rules in use today are Earth's. Gravity points down at 9.81 m/s^2 , surfaces do not remember what pressed on them, and time runs at the same rate everywhere in the room. Nobody chose these values; they came with the engine.

The first half of the claim is that the rules are part of the work. Clouds AO's pavilion catalog already describes each pavilion type as a data sheet with fields for description, screen or art, users, control, lighting and materials [1]. We add a field for rules. It holds eighteen numbers: the magnitude and direction of gravity, tilt, a pulse, grip, bounce, viscosity, spin, a lift per theme, cohesion between works that share a meaning, touch and touch memory, softness, ripple, a time bubble and a time scale (section 7, appendix A). Because the sheet is executed, we call it a rule program. The curator model ALAN, which answers visitors from Institution's corpora (section 4), is also the editor of that sheet: a visitor or an artist speaks, ALAN returns a typed change set and a prediction, a validator clamps the change to the schema, and the simulator runs the same benchmark before and after to score the prediction (section 7.2). A rule sheet, a language editor and a verifier together are what the world-model literature of October 2026 calls an engine-plus-renderer system (section 9.1), with one difference: the world has art in it.

1.2 Exhibition as path

The second half of the claim concerns what an exhibition produces. A visitor to the virtual museum leaves a record of which pavilion they entered, which work they looked at, for how

long, and what they asked (section 5.2). The sensing programme adds gaze, pulse and heart rate for visitors who opt in, in the headset and later in the building, under one event schema (sections 5.3 and 5.4). That record is a path: a sequence of positions, dwells and gaze targets through a room. A robot moving through the same room leaves a trajectory, and a world model asked to predict the room produces a rollout. Under the same schema the three are one kind of object, and so the same measurement applies to all three: how far the path departs from what the rules imply, and how fast that error falls with repeated trials when the rules are changed (section 9.6). An exhibition is then a place where a human, a machine and a model can be compared on the same task with the same instrument, and a museum is a source of such paths that a laboratory cannot assemble: many visitors, repeated exposure, curated content, and consent. The same reading runs the other way (section 1.4): a visitor's heartbeat can set the pulse of the room's physics, so the affective record is not only read from the path but written into it.

1.3 Two spaces

Institution runs the virtual museum and is designing a physical one. The building, by Clouds AO, is a vertical atrium wrapped in a translucent screen, with galleries around the void and a gathering space at its foot; Institution's brief adds a hall of 1,000 to 3,000 m² with a stage that rises and lowers, retractable seating and an underground level for dark works (section 8.1) [2]. Every pavilion in the virtual museum has a physical counterpart in that plan. The building is therefore a second place where the same rule sheet runs, in three tiers: rules that matter can obey (a tilting platform, a turntable, a pressure floor), rules that machines enact for the bodies that are machines (drones in the void standing for the constellation of works, trolleys, winches), and rules that stay virtual for human bodies and are seen enacted rather than felt (section 8.2). Buildings that move already exist; ours would move by the exhibition's rule sheet.

1.4 The heartbeat

If the physics of a pavilion becomes a path, so does affective data. The rule sheet of section 7 has a pulse field: gravity oscillates by a fraction at a frequency. An artist can set that frequency, and so can the wearable that records a visitor's heart rate under the consent rules of section 5.5: a resting heart at 72 beats per minute is 1.2 Hz, and the whole hang breathes with the body standing in it. Affective data becomes physics, the physics moves the works, and the works' motion is a path that the same schema records. This gives a museum in which the semantic, the affective and the physical contextualise the exhibition together, in the headset and in the building, from one sheet.

2. Institution

Institution is a contemporary art institution founded by Fabio Cavallucci, Tom Cebic and Lucilla Grossi. It works in Italy and New York, and now in China. It programmes exhibitions, a course on the history of art taught from outside Europe, and an annual forum, and it runs five public software systems that it built itself. Everything described here is a way for a few people to run an institution that has a public voice, a virtual museum open to anyone with a browser, and an operations layer that remembers its decisions [7].

2.1 Programmes

The Decolonized History of Art is a course of 24 lessons given by 24 faculty, delivered in Italian, English and Chinese [7]. Its subject is the history of art told from outside the European canon, and its faculty come from outside that canon. Its transcripts are one of the four corpora from which ALAN answers (section 4.2), so the baseline ALAN reasons from is not the Western survey.

The Culture Commons Forum has run in two editions. The first edition was held at Santa Sofia, with roundtables moderated by invited designers and researchers [8]. The 2026 edition ran as a relay across eight hubs in eight time zones, from Tokyo and Shenzhen through Tanguiéta, Jakarta, Berlin and Guatemala City to Santa Sofia and New York, streamed on institution.tv with the original audio and three interpreted channels in English, Italian and Chinese [8]. The forum is where two of the systems in section 6 were used in public for the first time.

2.2 Geography

The institution's programmes and legal seats are in Italy, and its physical museum is briefed for New York [9, 2]. Work in China is anchored by Fabio Cavallucci's Changjiang Scholar appointment, whose first stated objective is to promote VR, AI and robotics in the arts and whose outcomes include a permanent innovation lab and immersive labs [9]. The lab of section 10 is the Guangzhou form of that work.

2.3 Five public systems

Five systems run in October 2026 [7, 10]:

- `vr.institution.art`, the virtual museum: a Unity WebGL world of pavilions reachable from a browser or a WebXR headset (section 3).

- curator.institution.art, ALAN: the AI curator that answers in public chat and inside the virtual museum (section 4).
- atrium.institution.art, Atrium: the operations layer on which the institution runs (section 6).
- translate.institution.it, the live interpretation service: a 13-language speech cascade with human interpreter consoles (section 6).
- institution.tv, the forum's livestream hub: two days, four audio channels, eight streams (section 6).

The sixth element is the physical museum by Clouds AO, Masayuki Sono and Ostap Rudakevych (section 8).

3. The virtual museum

The virtual museum is a world of pavilions placed in a constellation. A visitor enters through a browser or a headset, arrives in a hub, and moves between pavilions along curated paths. Each pavilion holds works bound to named slots in a reusable template scene, and everything that defines the world, from the position of a pavilion to the text of a work's description, lives in a database rather than in the application.

3.1 The world in October 2026

In October 2026 the world holds 23 pavilion rows, of which 16 are open to visitors. The open pavilions are built from about 20 reusable template scenes, with four further pavilions that exist only as compiled bundles: the hub, the Education Pavilion, a six-screen video pavilion and a brutalist sculpture pavilion [10]. Thirty-seven works sit in slots inside the pavilions: 32 videos, 4 articles and 1 sculpture. Every pavilion carries its description in English, Italian and Chinese, and a visitor whose browser is set to Chinese sees the Chinese text [12].

The topics shown on the deck are Institution Under Construction, with the subtopics Gaze and Memory and Body and Identity, and AI and Sustainability [7].

The client runs in WebGL 2 in a desktop browser and in WebXR in a headset browser; there is no native headset application [10]. The core download is about 119 MB per fresh visit, measured in 2026 [10]. Multi-user presence, spatial voice and avatars are in the code [13].

3.2 The Pavilion Catalog

The architecture of the virtual museum is by Clouds AO, the New York studio of Masayuki Sono and Ostap Rudakevych. Their Pavilion Catalog is a set of data sheets, one per pavilion type [1]. Each sheet carries the same fields: description, screen or art, users, control, lighting, materials and remarks, with a rendered interior, a plan and an exterior model. The catalog defines nine types: FLOCK, ARTICLE, CINEMA, ART, LECTURE, PERFORMANCE (in open, closed, chamber, sound, poetry and immersive variants), VIRTUAL REALITY, BLACK ROOM and EDUCATION [1, 14]. Every pavilion type is rendered as a white model, from the hub and the VR sphere to the performance and education shells [14].

The first type, FLOCK, is the arrangement of all pavilions in three-dimensional space rather than a room, and its sheet describes a “transition from line to volumetric flock” [1]. The constellation is a formation rather than a map, and the idea has a history: Clouds AO’s concept designed the world as a luminous flock of contents over a horizon, with exhibition spaces forming “Interactive Clouds above the Rings” and gathering areas sunk into arenas beneath them [15]. The design brief that preceded it described a “flock-like disposition of artworks as a floating cloud/galaxy” [16]. In section 8 the void of the physical building is where the constellation is flown for real.

In Unity the catalog types became about twenty reusable scenes: Art types A to F, Lecture 1S and 5S, Sculpture, Performance types A to E, Video in single, three-screen and seven-screen variants, and Article, plus the hub and the Education Pavilion [10]. A new pavilion is usually one of these templates with new content, which is why the world grows without a new build.

3.3 The works and their pavilions

The sixteen open pavilions and their works, with the template each uses, are recorded in the render manifest of October 2026 [11]. The works, with their makers:

- Marina Abramović and Ulay, Light/Dark, video of the performance, in a single-video pavilion.
- Józef Robakowski, Attention: Light!, a film after a sketch Paul Sharits left in Poland in 1981, which turns a Chopin mazurka into colour changes, in a single-video pavilion.
- Karolina Wojtas, steals from net, restaged photographs found online, in a performance pavilion of type E.
- Artiom Constantinov, Pain, Suffering, audio-reactive video, in a performance pavilion of type D.
- New-Territories (François Roche) with Emanuele Coccia, Impermanences Tractatus, a sculpture made from invasive algae for the Venice Architecture Biennale and designed to vanish, in the sculpture pavilion.

- Suzanne Lacy, *By Your Own Hand*, video of the Quito bullring in which two hundred men read letters written by women who suffered violence, in a seven-screen video pavilion and in a six-screen pavilion built for the work.
- Katarzyna Kozyra, *The Rite of Spring*, seven video screens, in a seven-screen video pavilion.
- Mattia Cupelli, *SYNC/IDENTITY*, performance, in a performance pavilion of type D.
- Analivia Cordeiro, *BodyWays*, a movement work, in a performance pavilion of type D.

The list spans film, performance, photograph, sound, sculpture and text, and each form needs a different pavilion type, which the catalog gives it [7].

The museum also holds lectures and texts. Thomas Seeley's lectures on how a honeybee swarm chooses a new home sit in a five-screen lecture pavilion, a tiered amphitheatre of about 37 by 38 metres [10]. Jeff Hawkins' articles on cortical columns, a sociocracy article on collective governance, and Fabio Cavallucci's essay "Institution: A Copernican Revolution for the Arts" sit in article pavilions. Seeley and Hawkins are in the museum because they are in ALAN: the institution's curatorial method takes decision-making in swarms and prediction in the cortex as working models, and the museum shows its sources [7].

The Education Pavilion is the newest. Clouds AO designed it as a ring of workshop rooms around two tiered halls: a large lecture hall under one surround screen, for about 900 people at 90 tables of ten, a small hall for about 90, and workshop rooms for 20 to 60 people whose enclosure transparency can be changed according to the activity [14]. Its compiled bundle is the largest in the world, 40 MB, with a main shell of about 105 by 139 by 27 metres [10]. The course programme for it is the next content job.

3.4 How a visitor moves

A visitor signs in with a Google account and arrives in the Hub Pavilion. From there four things are possible: fly out to the overview, where topics and subtopics float as labels in space and can be zoomed in and out; read the pavilion's text; press NEXT to follow the path a curator set through the portals; or ask ALAN [10, 7]. There is no menu of exhibitions. Movement between pavilions is a scripted flight of about three seconds along cubic Bézier segments through waypoints, with the character controller disabled during flight [10]. Inside a pavilion the visitor walks, looks, reads and asks. Fifteen directed portal links join the pavilions into a graph, and the sequence a curator sets through that graph is the argument of the exhibition [17]. The flights follow from the space being a constellation rather than a corridor.

Other visitors appear as avatars with voice in proximity, and there is text chat, comments and likes. That part is still being tuned for larger numbers [13, 18].

3.5 Platform architecture

The platform is a Unity WebGL client over a backend-as-a-service, a dedicated multiplayer server and the separate ALAN backend.

- Client: Unity 2022.3 LTS with the Universal Render Pipeline, compiled to WebGL 2 for the browser and reaching headsets through WebXR [10]. The project compiles only for the WebGL target.
- Data: Supabase, which holds a Postgres database (topics, subtopics, pavilions, portals, content, profiles, likes, playlists), a storage bucket with 35 pavilion bundles, realtime channels for chat and notifications, and authentication through Google sign-in [19, 10].
- Media: video is streamed from content delivery networks onto in-world screens whose size is set from the video's aspect ratio; images and articles load by URL [19, 10].
- Multiplayer: Unity Netcode for GameObjects over WebSockets with TLS to a dedicated server build of the same project, running in AWS London at 30 ticks per second; movement is client-authoritative [10].
- Voice: Photon Voice, spatial, with MFCC-based lip-sync driving visitor avatars [19, 10].
- Hosting: the build on S3 behind CloudFront, last deployed October 2026; a staging deployment on Amplify [10].
- ALAN: HTTPS calls to curator.institution.art, token-gated. The client sends identity, location and the visitor's words only [20, 21].

Retrieval, memory and the voice rules live on the server, where they can be governed and changed without rebuilding the world.

Two architectural facts matter later. Every open pavilion is downloaded and instantiated at boot, which is simple and does not scale past a few dozen pavilions [10]. The dedicated server already loads the pavilion scenes with their colliders, but there is no server-side physics stepping of world objects and no networked rigid bodies [10]. The server is the natural host for the server-authoritative physics of section 7, which would be an extension of it.

3.6 A pavilion is data

A pavilion in this system is three things: a template scene, one database row, and a list of works bound to named slots in that scene [12, 10]. The row holds the pavilion's name, its template, its coordinates in the constellation as a semicolon-separated triple, its subtopic, its visibility and scheduled opening, its ambient sound, and its description in three languages. The content rows hold the works: title, body text, type (video, article, image, 360-degree

video or sculpture), media URL, the slot the work goes into (Video;1, Article;2 and so on) and the creator.

At boot the client fetches every table, filters the visible pavilions, downloads each distinct template bundle from storage, loads it additively, binds the content rows to the tagged slots, places a copy of each pavilion's exterior at its coordinates in the dark star field, and lands the visitor in the hub [10]. Switching pavilion moves the full scene to the exterior's position and activates it.

Curators work in a web CMS with five tabs: Pavilions, Topics, Positions, Users and Bundle [12]. Positions is the walk graph, with "comes from" and "leads to" relations and branching. Topics and subtopics carry map coordinates that place them on the overview. The Bundle tab accepts custom asset bundles for pavilions that depart from the templates.

The design rule from the data architecture was that content changes without an application update and that the database stays in plain text so that an AI can read it [19]. ALAN reads it, and the same property allows a pavilion's rules, in section 7, to be written as fields on a sheet rather than as code.

4. ALAN as a system of layers

ALAN stands for Associative Learning Affective Art Network. In production it is six layers: a model, a corpus, retrieval, a voice and disclosure policy, memory, and output. The same six layers serve four surfaces: the public chat at curator.institution.art, the companion inside the virtual museum, the staff assistant on email and Slack, and the rooms of Atrium [7]. ALAN runs on Amazon Bedrock, and answers from Institution's own corpora, built from worldwide curatorial sources rather than the Western survey; it answers from documents and says when it has none [7].

4.1 Layer 1: the model

The model is a stock Claude model called through Amazon Bedrock's Converse API [22, 20]. A fallback chain degrades a provider error to the previous Sonnet and then to Haiku rather than to silence [22]. Temperature is 0.35. Embeddings for the vector index are OpenAI's text-embedding-3 family [22]. Speech is produced in the browser by speech synthesis in the virtual museum, and by server-side synthesis and transcription in the public chat [10, 22]. The backend is one FastAPI service on a single EC2 instance in us-east-1, with Postgres, S3 and Secrets Manager behind it [22, 23].

The model controls fluency and cost. What ALAN knows is in the corpus and how it speaks is in the policy, so a change of model changes neither. The whole backend runs on one small server, and the cost is in the model calls rather than in the machine [7].

4.2 Layer 2: the corpus

The corpus makes ALAN the institution's. It is four corpora: authored notes written by the institution's curators; a curator reference corpus of curatorial sources; transcripts; and answers the cohort has reviewed [23]. At the 2026 review it held 1,519 documents in about 21,000 chunks, organised along 45 named research paths [24].

Its sources are the institution's exhibitions, publications and recorded encounters, the course, and curatorial sources assembled from outside the Western canon [7]. The course is taught by 24 faculty from outside the European canon, so the baseline ALAN reasons from is not the Western survey. The reference corpus is assembled on the same principle. The project's own terms for this are non-Western and under-covered voices, cohort thinking, and conceptual mapping over movement labels [23].

The corpus is maintained by nightly loops. One re-reads a rotating 10 percent slice of the corpus; one re-reads a 1 percent slice of cohort answers; research loops run on artists where the corpus is thin [23]. In public chat, if the corpus has nothing, ALAN may search the web, and that answer is labelled as outside the corpus, quarantined and cached, and never promoted into the main corpus [23]. In the virtual museum there is no web search at all [20].

4.3 Layer 3: retrieval

Retrieval decides what the model is allowed to see for a given question. The index is a Pinecone vector store with a lexical fallback in Postgres [20, 23]. In the virtual museum, retrieval runs three passes, scoped from the inside out: first the documents of the artwork the visitor is standing at, then the documents of the pavilion, then the public reference corpus, which comprises the reference source types such as encyclopaedia entries, art databases and journals [20]. Each pass returns up to eight chunks. Only the public tier is searched in the museum, enforced by namespace and by a disclosure filter applied before any prompt is built, with a post-condition that drops anything that slipped through [20]. If the three passes return nothing, ALAN says it has nothing.

4.4 Layer 4: voice and disclosure

Every document in the corpus carries a disclosure tier and every retrieved chunk is assigned an evidence role, and the policy that governs both is a written document rather than a prompt

[25]. The six tiers are: public direct, public paraphrase, public link only, public silent tissue, operator only, and private archive. The filter on tiers runs at retrieval, so private material cannot reach a public answer by accident. The six roles are: primary source, supporting evidence, bridge, voice calibration, prosodic template, and counterposition. A primary source can be quoted. A calibration text teaches tone and is never quoted. A counterposition that contradicts the primary source has to be acknowledged [25].

After generation a pre-publish check runs. It rejects a set of forbidden verbs (explores, interrogates, grapples with, evokes, and the rest of that register), checks that every quoted span has a source, and matches five-word sequences against the calibration texts so that a tone model is not reproduced as an answer [25, 23]. Answer length is budgeted: 400 characters for a first answer, 900 on a follow-up, 1,500 on a deeper one [22, 20]. The critical posture of the voice was shaped by the published writing of the critic Alessandra Mammi, with her written permission [20, 7].

Grounding is mechanical: the tier filter and the pre-publish check decide what an answer can contain, and when ALAN has no documents it says so, which tells the visitor where each answer came from.

4.5 Layer 5: memory

Memory is small. Each visitor has a short summary, rewritten every six turns from the last eight, that persists across sessions, so the next visit resumes where the last one stopped [22, 20]. For every turn the system infers a register, among factual, operational, scholarly, contemplative and lyric, and uses it to set the pacing and length of the answer; the register is not stored across sessions [20]. Raw conversation beyond the summary is not kept. No visitor data is used to train a model, and each institution that runs ALAN has its own deployment with no corpus shared across them [23].

4.6 Layer 6: output

In the public chat the output is text with citations. In the virtual museum it is speech, synthesised in the browser, with the sources carried separately in the response payload rather than read aloud [10, 20]. The staff assistant answers on email and Slack after verifying the sender's identity, and the Atrium rooms use the same voice policy on every staff draft (section 6).

4.7 Fine-tuning work: research models

Alongside the production system the institution fine-tunes models. All of these are research lines. ALAN in production answers through Claude with retrieval.

Voice adapters. The base is DeepSeek-R1-Distill-Qwen-32B, an open 32-billion-parameter distillation of the R1 reasoning model into a Qwen architecture [26]. The method is supervised fine-tuning with QLoRA [27]: the base is quantised to 4-bit NormalFloat, and low-rank adapters of rank 64 in the first version and 128 thereafter are trained on top, with ChatML formatting and 4,096-token sequences [28]. The data is Institution's own writing. Seven adapter versions were trained up to 2026. Hardware moved from an H100 for the first version, to L40S instances on AWS for versions three to six, to A100-class machines for version seven [28]. One direct preference optimisation run, with beta 0.1 for one epoch at a learning rate of $5e-6$, made the voice worse and was reverted [28].

A corpus audit in 2026 found that the training completions used for the first three versions had been largely drafted by a language model, so the adapter had learned a model's punctuation as the institution's voice [28]. Version seven was retrained on 489 checked essay rows, with continued pre-training followed by supervised fine-tuning, and the fingerprint went away: the em-dash rate fell from 14.0 to 2.1 per 1,000 words [28]. A version-six loader bug, recorded in the configuration, meant that model had trained on the 156 baseline pairs only [28].

Artwork model. In 2026 the research team fine-tuned VideoLLaMA-3 [29], a video-language model, on five representative artworks with labels written by curators for style, colour and emotion [30]. The sample set included Kozyra's *The Rite of Spring*, Ortega's *Controller of the Universe*, Lacy's *De tu puño y letra* and Robakowski's *Attention: Light!* [31]. The training was sharded across several GPUs, and the model was deployed to Hugging Face with a basic front end [30]. It is a prototype that reads style, colour and emotion in a work; it is not wired to ALAN.

Both study what a voice is and what a model can read in a work; neither is in the production answer path.

4.8 ALAN inside a pavilion

Inside the virtual museum, the client holds no intelligence and the companion knows where the visitor stands. The Unity client calls four routes on the curator backend, status, session, location and chat, with a token in the header [10, 21]. When the visitor arrives at or leaves a pavilion, the client tells the server; when a work comes into focus, the client sends which one; when the visitor speaks, the words go up [10]. The server returns a spoken answer of

400, 900 or 1,500 characters depending on the depth of the exchange, the sources separately, and the updated memory summary [20, 21].

ALAN appears as a chrome “Ask ALAN” sphere that follows the visitor at 3.7 metres in VR, with collision avoidance against the pavilion geometry, and as a chat panel with a spatial keyboard in headset mode [10]. It has no humanoid avatar and no lip-sync of its own. It answers in English, Italian and Chinese, the three languages of the course. It remembers the visitor between visits as the short summary of section 4.5, not as a transcript.

Everything ALAN knows about where the visitor is comes from events the world already produces, which makes the exhibit a sensor (section 5).

5. Semantic and affective data

A work has two layers that a visitor reads at once: what it represents and what it does. The first is semantic: material, date, place, figures, composition, provenance. The second is affective: what the colour, the composition and the symbols do to the person standing in front of it. Current systems describe the first layer well and the second badly, and when they try the second they project one culture’s responses onto everyone. The ALAN paper [32] separates the two layers in the architecture and recombines them afterwards. In production, the semantic layer is the corpus described in section 4. The affective layer is a data programme. It needs the exhibit itself to become the instrument.

5.1 Two pathways, integrated after separate processing

The paper, “ALAN: A Demonstration of an Emotionally-Aware AI Companion with Human Expert Guidance”, was written by Xia Wang, Zeyang Zhang, Fabio Cavallucci, Tom Cebic and Jonathan Sprinkle and submitted to the ECAI demonstration track [32]. Its starting observation is that object recognition on a Basquiat returns “person, crown” and misses the anger and the historical weight. A model trained on one body of data projects one culture’s reading onto every work, and that reading is usually the Western one.

The design answers with a disentangled affective-semantic engine. Two attention pathways process the same input independently: one captures the objective and factual elements of the work, the other extracts emotional cues from colour, composition and contextual symbolism. The two are integrated only after separate processing, so the model can distinguish the literal content of a work from the experience it is built to produce [32, section 2]. The inspiration the paper names is the voting mechanism of cortical columns [33]. The

encoders are a Vision Transformer for images, CLAP for audio and SPECTER for text; the paper's core model is a fine-tuned LLaMA-3-70B with LoRA adapters, which is the research configuration, not the production one (section 4 gives the production stack).

The cultural argument is the paper's central rule. A bird is a symbol of mourning in one culture and of fortune in another. The affective reading is cultural, so the training data must be, and the ground truth must come from curators in the regions the works and the visitors come from. The paper states this as "human-aided fine-tuning": partner curators supply structured, expert-annotated data for the interpretations an AI cannot learn on its own [32, section 3]. A sociocratic veto lets a curator flag a problematic narrative, after which the interface shows a revised one co-authored with experts. The paper is a demonstration plan with no quantitative experiments; its contribution is the architecture and the governance. The data programme below is how the affective side gets its training signal.

5.2 What the exhibit already records

Before any sensor is attached to anybody, the virtual museum records five signals as side effects of events the world produces for other reasons [34, 35].

Position. The client tells the ALAN server when a visitor arrives at or leaves a pavilion and which work is in focus. The calls are `POST /vr/alan/location` on arrival and departure and `SetArtworkContext` when a work is selected; each is timestamped [35]. **Dwell.** Time between arrival and departure per pavilion, and between focus and the next focus per work, falls out of the position events by subtraction. Nothing extra is recorded to obtain it.

Questions. Every question a visitor asks ALAN is stored per session, with its language, in the `VRConversationTurn` table [34]. **Register.** For every turn the server infers a register (factual, operational, scholarly, contemplative, lyric) and uses it to set pacing, that is, how fast and how long ALAN answers. The register is not stored across sessions; it is discarded at the end of the session [34, `VRConversationTurn.inferred_register`]. **Reactions.** Likes, saving to a playlist and comments exist as Supabase tables (`likes`, `playlists`) and are coming to the client [36].

5.3 Sensing: in the headset, on the body, in the building

The sensing programme is the AI Guide plan, a document that has kept the same phase structure through its versions.

WHERE	SIGNAL
In the headset	head pose
In the headset	eye tracking and the gaze target on a work, on headsets that provide it
On the body	pulse and heart rate from a wearable the visitor chooses to wear
In the building	image recognition so the guide follows the visitor from work to work; NFC or QR tags at the work
In the building	sensors for emotional impact in the room, with artists making installations that use them

Table 5.1. The sensing programme.

The phases in the plan run: Phase 1, an MVP guide; Phase 2, an enhanced guide with image recognition and speech; Phase 3, a mobile guide and preparation of the physical institution, where the guide follows the visitor through the building by recognising the work in front of them; Phase 3.B, emotional-impact integration, where sensors in the physical space detect emotional impact and artists build installations that use them; Phase 4, continuous refinement with artist-specific models; and Phase 5, which the plan calls “Artistic Intelligence”, a new corpus of affective training data with classification categories beyond standard sentiment analysis [37, phase table]. Phases 1 and 2 run in VR: the guide answers from the corpus, hears speech and speaks, and the client knows which work is in focus. Phase 3 and everything after it come next. No biometric data is collected in either space [38, finding 9].

Eye tracking is the next signal because the gaze target on a work, how long it is held and in what order, is the closest thing to an affective reading a headset can give without a wearable, and current headsets provide it. The body comes after, through a device the visitor chooses. The building comes last, because a camera is a different instrument from a headset, and the plan starts with tags and devices the visitor carries (5.5).

5.4 One work, two sensors: a single event schema

The reason to build the virtual museum first is that the same work can be sensed the same way in both spaces. We define one event:

visitor · work · time · position · gaze target · dwell · question · affect
(if opted in)

In VR the engine gives the position, the headset gives the gaze target, the arrival and focus events give the dwell, the chat gives the question, and a wearable, if the visitor turned it on, gives the affect signal. In the building, image recognition or a tag at the work gives the position, eye-tracking glasses or a headset give the gaze target, dwell comes from the same arrival and departure events, and the wearable is the same device. The fields do not change; only the sensor behind each field does.

The pairing rests on a rule carried over from the VR design: every pavilion has a physical counterpart, and the platform always knows where the body of the work is. The project documents call it phygital integration, “every node has its digital twin” [39]. A work therefore has one identity in both spaces, and a visitor who stood in front of it in the headset and later in the gallery produces two events with the same `work` value.

What the event feeds, in order of how soon each use is possible:

1. ALAN’s pacing and suggestions, which already use position and register today.
2. The curators’ reading of a show: how a room was actually walked, which works held attention, where people left.
3. Artists’ installations that respond to the room, which is Phase 3.B of the plan and the subject of section 7.
4. The affective corpus with cultural labels: events from visitors of different cultures, labelled by curators from those cultures, which is the dataset the ALAN paper describes.

5.5 Consent and data rules

Seven rules bind the programme in 5.3 and 5.4.

1. **Opt-in.** Any body signal is collected only when the visitor turns it on. The default is off.
2. **Ownership.** Affective data belongs to the visitor. Sharing it with an institution or with other visitors is a visible choice the visitor makes [37].
3. **Not stored.** The per-turn register is discarded at the end of the session. ALAN’s memory of a visitor is a short summary rewritten every six turns, not a transcript; raw conversation beyond that summary is not kept across sessions [34].
4. **Not used.** Visitor data is never used to train a model. Each institution that runs ALAN has its own deployment, and there is no corpus shared across deployments.

5. **Separation.** The public index and the operator index are separate systems; a public question cannot reach operator-only material, and the private archive is outside every index (section 4.3).
6. **Processing.** The VR platform has a data-processing agreement and a GDPR Article 28(3) record of processing with its platform developer, naming Supabase, AWS, Photon and Unity as sub-processors [40].
7. **Building.** Cameras raise questions a headset does not. The physical programme begins with tags at the work and devices the visitor carries, and adds cameras only after a separate review.

5.6 Exhibition as path

The event in 5.4 is one row and a visit is many rows: a visitor arrives, looks, stays, asks, moves on, in a sequence the exhibition shaped but did not dictate. That sequence is the visitor's path through the show, and it is the most complete record an exhibition can produce of what it did to a person. Art as simulation is the pavilion as a world whose rules are authored (section 7); exhibition as path is the record a body leaves in that world.

Two consequences follow. A path is the natural unit for comparing how a show works in the headset and in the building, because the schema is the same in both; twin fidelity between the two spaces is measured on paths (section 10.4). A path is also what a robot leaves in a room (a trajectory) and what a world model produces when it predicts a room (a rollout). Section 9 puts a visitor, a robot policy and a world model in the same pavilion under the same rules and fits the same curve to each of their paths. The affective programme and the robotics programme share a data structure, and the structure closes on itself: the heart rate the wearable records is a field on the rule sheet (section 7.1), so the body's own rhythm can move the works it stands among.

6. Atrium

Atrium is the operations layer: the system Institution runs on. Everything in sections 3 to 5 is public-facing; Atrium is where the staff plan, draft, check and record the institution's work, and where the agents that help them are governed. It was rewritten in 2026 from a single model loop per turn into a typed multi-agent engine, and it has been in daily use since [23, 41].

6.1 The engine

Atrium is a typed engine on LangGraph. A request becomes a Goal. The Goal becomes a Plan, the Plan becomes Steps, every Step passes a set of Checkpoints, and the whole run is recorded as a Case that later Goals on the same subject can read [23]. The agents inside a run are named by function: a planner, a critic, a researcher, a drafter, a verifier and an aggregator. One composer turn is one Goal, which is one task on the dashboard with its plan, steps and checkpoints visible, rather than a streaming chat [41].

Role rooms. There are eight rooms, one for each kind of work an institution does: curatorial, grants, development, education, press, support, artist relations and archive, plus a daily briefing. Each room is a voice profile, a tool bundle and a set of Goal-category preferences, rather than a persona overlaid on one orchestrator [23].

Checkpoints. Governance lives in the checkpoints: budget, plan coherence, drift from the Goal, value, voice, faithfulness, citation integrity and human approval. The voice checkpoint applies ALAN's voice-and-disclosure rules (section 4.4) to every staff draft. The approval checkpoint pauses any step whose action class is outward or irreversible, an external email for instance, until a person approves it [41].

Autonomy. There are three levels. L0 drafts only. L1 allows side effects when the engine's confidence clears a gate, with audit. L2 runs a background loop under daily caps. New projects default to L2 for project agents and employees run at L1 with audit through the gateway (6.2) [23].

Costs. A ledger records tokens, tool calls and wall-clock for every run, per model tier (Haiku, Sonnet, Opus 4.7). The default cap is

25perinstitutionperday. L1budgetswereraisedto60toolcalls,

1.50 and 900 seconds per Goal during the first walkthrough after a budget checkpoint aborted a run at 114% [41].

Isolation and knowledge. One Postgres database, with row-level security per institution, so the same engine serves several institutions without their data meeting. Knowledge comes from a Bedrock knowledge base built from eight of Institution's shared drives: 286 of 316 documents indexed as of 2026 [23]. Every researcher's `search_institution_kb` call grounds against that corpus.

Tools. Google Workspace: Drive, Docs, Sheets, Slides and Calendar through an OAuth connector with document, spreadsheet, drive-file and calendar scopes; Gmail send through a delegated service account with an audit log and a revert path (Drive trash, restore from

snapshot, a retraction follow-up for sent mail) [23, 42]. Each project gets its own Drive folder under `/Atrium Projects/{title}/`.

Studies. A Study is a persistent research object on one subject, an artist for instance, that holds pinned facts, threads and sources across months and can be promoted to a Project with one click. Creating a Study on an artist fires four parallel Goals (identity card, recent works, constellation, opportunities) and pins each result as a dated, signed institutional position [23, 41].

6.2 The staff gateway

Anyone in the institution writes to ALAN by email at `alan@institution.art` or in Slack. Identity is verified before anything else happens, because a spoofed sender would otherwise have a tool-using agent at its disposal. For email, the message must carry Google's DKIM and DMARC pass aligned to the From domain; a spoofed message is rejected and forwarded once to the operator. For Slack, the request must carry a valid signing-secret HMAC and the sender must be on the workspace e-mail allowlist [42]. Verified employees run at L1 with audit; the three admins run at L2.

Most requests are answered in seconds by a fast loop with real tools: Gmail, Calendar and Drive reads, the knowledge base, web search and fetch, and writes that create native Docs, Sheets, Slides and Calendar events. Anything that takes a `DEEP:` prefix goes to the full engine. An email is never sent directly: it is staged, and goes out only after the requester replies with the word "send". Every reply copies the operator. Two bugs found on the first day in production are recorded with the code: replies were cut at 240 characters because the gateway read a summary field meant for dashboard cards, and overlapping poll runs sent duplicate replies until a single-flight lock and per-message persistence were added [42].

6.3 Live interpretation and the forum hub

Two Atrium services ran in public at the Culture Commons Forum in 2026 [43, 44].

translate.institution.it is a self-hosted live interpretation service built in 2026. A host captures the program audio; the pipeline transcribes it, translates it per target language and synthesises a voice, and listeners pick one of thirteen languages (English, Italian, Spanish, French, German, Portuguese, Dutch, Polish, Swedish, Japanese, Korean, Chinese, Arabic) in a browser and hear audio with captions. Partial transcripts arrive as instant captions; finalised segments arrive as caption plus audio, 2 to 5 seconds behind the speaker. The cost is a few dollars per language-hour, against

120 per hour for commercial live interpretation [43].

institution.tv is the forum's public hub: a static site with two days, four audio channels (original, English, Italian, Chinese) and eight YouTube streams, with a 19-row schedule shown in the forum's own time zone and the viewer's. For a venue in Shenzhen where YouTube is not reachable and the public stream runs 20 to 40 seconds late, the service provided a direct room feed at `/room/{en,it,zh}` on the translate box, so a physical room could hear the interpretation without the public platform [44]. The forum ran on the human interpreters; the AI cascade remains available for events where a human is not booked.

7. Physics as medium

A pavilion's physics is authored like its lighting. The lab that implements it runs in a browser (`demo/pavilion_physics_lab.html`), and the standalone copy in `demo/standalone/`.

7.1 The rule sheet

The Clouds AO catalog describes a pavilion type in six fields: description, screen or art, users, control, lighting, materials [1]. We add a seventh, rules, and give it a schema of eighteen numeric fields with ranges and units (appendix A): gravity magnitude, upside down, tilt and tilt direction, pulse and pulse frequency, grip, bounce, viscosity, spin, a lift per theme, cohesion, touch and touch memory, softness, ripple, time bubble and time scale. Every field has a range the validator enforces, so a sheet is always executable. Earth is the sheet with gravity at 9.81 m/s^2 and every other field at its neutral value.

The bodies in the room are works, a visitor, a ball and a wheeled robot. Each work carries a six-dimensional theme vector over water, stone, light, body, machine and memory, taken from its labels. The floor and the ceiling are 72×72 grids that record pressure and ripple, so that a dent forms where load accumulates and persists for as long as touch memory says.

One equation moves every body:

$$\mathbf{a}_i = f_i, \mathbf{g}(t) - \nu, \mathbf{v}_i - 2\boldsymbol{\Omega} \times \mathbf{v}_i -$$

$$\begin{aligned} & \|\boldsymbol{\Omega}\| \times (\|\boldsymbol{\Omega}\| \times \|\mathbf{x}_i\| + \\ & \|\mathbf{a}_{\text{slope}}\| + \|\mathbf{a}_{\text{cohesion}}\| + \|\mathbf{a}_{\text{query}}\| + \\ & \|\mathbf{a}_{\text{policy}}\| \end{aligned}$$

Here $g(t)$ is gravity with its tilt and pulse, f_i is the lift factor from the work’s themes (works about light can weigh less than works about stone), v is viscosity, Ω is the room’s spin with its Coriolis and centrifugal terms, the slope term is the gradient of the dent field under the body, cohesion pulls works with similar theme vectors together, the query term pulls works toward a visitor whose request names their themes, and the policy term is the robot’s own control. The pulse term is where the affective layer enters the physics: pulse and pulse frequency can be set from a visitor’s heart rate, 72 beats per minute being 1.2 Hz, so the hang breathes with the person in the room. Integration is symplectic Euler at 120 Hz in the live view and 60 Hz in the benchmark. The engine is hand-written and runs in the browser; the same rules can be ported to a Unity pavilion, which is one of the capabilities section 10 asks for [3].

The sheet is a program rather than a setting because it is executed, it has a schema with types and ranges, and a language model edits it under validation. Those three properties make language-authored physics usable, and they are why the world-model literature treats engine-held rules as the robust half of the field (section 9.1).

7.2 ALAN as editor

The editor is ALAN, the same model that answers visitors from Institution’s corpora (section 4). Given a sentence such as “turn the room upside down and make the floor remember where people stood”, ALAN returns three things: a change set over the eighteen fields, a one-line summary a visitor can understand, and one to three predictions, each naming a benchmark metric, a direction, an expected value and a reason. The prompt carries the schema with ranges and descriptions, the current sheet, the current benchmark figures, and the instruction to change only what the request implies. The reply is JSON rather than code; ALAN never executes anything.

A validator clamps every changed value to its range and reports any field outside the schema, which is then ignored. The benchmark runs before and after the change, and each prediction is scored on direction (did the metric move the way ALAN said) and on absolute error against the measured value. ALAN’s running score for the session is shown beside its answers. When

no model is reachable, an offline parser handles the example requests with the same schema, and the source of every change is labelled.

Language models revert to Earth values when asked about worlds that are not Earth (section 9.2): a frontier model scored 0 of 15 in a world where quantities decay, and slid back to standard ratios [4]. In the pavilion that tendency becomes a number on the screen, per request. The artist gets a medium with a feedback loop: say it, see it, and see whether the system understood its own change.

7.3 The benchmark

The task is fixed: a wheeled robot pushes a ball to a goal ring while a visitor walks the room, and the robot must keep clear of the visitor. Only the rule sheet changes. Each rule sheet is run on the same twelve seeds for 25 s at 60 Hz, and four figures are recorded: deliveries out of twelve, median time to delivery, close calls in seconds per trial (time spent inside the visitor’s exclusion radius), and the Earth-forecast error, which is how far a predictor that assumes Earth physics is wrong about the ball’s position 2 s ahead. The figures below were measured in October 2026 and exported from the lab (`demo/benchmark_presets_2026-10-26.jsonl`) [5].

RULE SHEET	DELIVERED, OF 12	MEDIAN TIME, S	CLOSE CALLS, S PER TRIAL	EARTH-FORECAST ERROR, M	WHAT HAPPENED
Earth	11	6.6	0.16	0.00	the baseline
Moon	7	13.3	0.17	0.97	every push floats; slow
Underwater	2	21.4	0.32	1.12	drag stops the ball short
Ice floor	3	15.8	0.39	1.19	no grip; the robot slides
Zero gravity	0	—	0.00	0.23	wheels have nothing to push against
Upside down	10	8.5	0.16	6.36	same task on the ceiling; the Earth

RULE SHEET	DELIVERED, OF 12	MEDIAN TIME, S	CLOSE CALLS, S PER TRIAL	EARTH-FORECAST ERROR, M	WHAT HAPPENED
					forecast is 6 m wrong
Tilted floor	6	8.1	0.30	0.38	the ball rolls downhill, away from the goal
Spinning room	3	8.7	0.49	0.36	Coriolis bends every push; more close calls
Memory-foam floor	12	6.4	0.10	0.04	dents hold the ball on course
Light rises, stone sinks	11	6.6	0.16	0.00	works lift; the ball does not
Time bubble	7	6.6	0.24	0.11	slow near the visitor; close calls rise

Two rows matter most. In zero gravity the robot never delivers, for a physical reason: its wheels have nothing to push against. Upside down it delivers ten of twelve, because the task is unchanged for a robot that is now on the ceiling, but the Earth forecast is 6.36 m wrong. A model that assumes Earth is confidently wrong about the room it is in, and the pavilion is a place where that error can be measured on any predictor, including the ones that drive robots. The lab also runs dose-response sweeps, in which one field (gravity from $-2\times$ to $+4\times$ Earth, tilt, grip, viscosity, spin, softness, the time bubble) is stepped through nine values and the four figures are plotted against it, so that the sensitivity of the task to each rule is a curve rather than a row.

7.4 Commissions as sheets

What artists make with this is a rule. The commissions are written in the catalog's own format, for artists to answer, with the same five lines each: what the visitor does, what the work does, how it runs in VR, how it runs in the building, and what kind of artist it needs.

COMMISSION	THE VISITOR	THE WORK	IN VR	IN THE BUILDING	RULE	ARTIST
Meaning Pulls	says “water and memory”, anywhere in the show	the works that carry that meaning come to the visitor: the constellation re-forms around them, light and sound follow, the rest of the show goes quiet	the pavilions drift together; the visitor is the centre of a new constellation	nothing heavy moves: the drones of the void gather overhead, each work’s light and sound rise or fall by the same potential, and the floor lights the path to the nearest one	cohesion and query	writer, with a light and sound artist
Floor Memory	walks the show, for weeks	the floor records pressure; the dent field shows underfoot and in VR; at the end it is milled in marble	the floor yields and remembers	a pressure floor, projection, then stone	touch memory, softness	sculptor
Turning Room	throws and catches in a sunken arena	the room rotates; throws curve; the Earth forecast is projected and is	the room spins	a turntable at 2 to 3 rpm; the forecast on the screen	spin	sound artist

COMMISSION	THE VISITOR	THE WORK	IN VR	IN THE BUILDING	RULE	ARTIST
		wrong in public				
Slow Near Me	approaches a dancer or a machine	everything near the visitor slows by one function; machines and dancers obey it together	exact	machines exact; dancers by ear	time bubble	choreographer

Each has a lineage: drone flocks run as public works, live simulations with visitors as boundary conditions, rooms that react by a law, language driving stage lights [6]. The new thing is the medium: a rule on a sheet, edited in language, verified by a simulator, and enacted in two spaces by the same potential.

8. The physical building

The physical museum is designed by Clouds AO. The building can enact the rule sheet of section 7 with a hall that reconfigures, a flock of drones that holds the constellation, and a dark level for the rules a body cannot feel.

8.1 The design and the brief

Clouds AO's Physical Space Interiors organises the building around a vertical atrium: a void wrapped in a translucent screen, with galleries arranged around it on each floor and a gathering space with benches at its foot [2]. The screen shows content inward to the galleries and outward to the street, so the building's programme is visible from outside. The renders in the deck (sheets 28 and 29) are from that document.

Institution's brief, written for a New York site, asks for a large hall of 1,000 to 3,000 m² with a ceiling of 6 to 10 m, a stage that rises and lowers in the middle, a retractable seating system,

sound isolation, mezzanines, and an underground level for works that need darkness: cinema, concerts, immersive content [2].

One rule carries over from the virtual museum: every pavilion has a physical counterpart, and the platform always knows where the body of a work is (section 5.4). The building is the second space of the same exhibition, and the void is what makes the rest of this section possible.

8.2 The building as mechanism

Buildings that move already exist. The Shed in New York has an outer shell on rail wheels that telescopes out over its plaza in minutes to double the enclosed space [69]. The Perelman Performing Arts Center in New York has three theatres whose walls, floors and seating reset into dozens of configurations behind a translucent marble skin [70]. Ours would move by the exhibition's rule sheet. Each element of the brief maps onto one class of rule.

The hall. The rising stage is the tilt and pulse platform of the rule sheet. Retractable seating turns a lecture into an arena into an open floor, which is the sequence the Education Pavilion already runs in the virtual museum between its large hall, its booths and its piazza (section 3.3). A surround screen shows the state of the virtual room at the scale of that hall, about 900 people, so that the rules the floor cannot enact are seen enacted. Mechanics: a 6° tilt gives a lateral acceleration of $g \sin 6^\circ = 1.03 \text{ m/s}^2$; an 8 m platform tilted 6° rises 0.84 m from edge to edge; a rotation of 0.3 rad/s, which is 2.9 rpm, gives a Coriolis acceleration of $2\Omega v = 0.6 \text{ m/s}^2$ on a body moving at 1 m/s [3]. A rotating or tilting platform with people on it needs a safety review before any of these figures becomes a design.

The flock as constellation. The FLOCK sheet that opens the catalog (section 3.2) is the arrangement of every work in space: tens of drones stand for thousands of points in the virtual museum. The points are quantised to drones by k-means in position, each drone standing for a cluster of works with its brightness set by the cluster's mass; drones are assigned to clusters by the Hungarian method so that the formation re-forms with minimum travel; and each drone flies a minimum-snap trajectory under separation and person-exclusion constraints [3]. When a visitor speaks, the constellation re-forms: the cohesion and query terms of the equation of motion move the cluster centres, and the drones follow. This is how Meaning Pulls (section 7.4) runs in the building without moving anything heavy: the drones gather overhead, each work's light and sound rise or fall by the same potential, and the floor lights the path to the nearest one. The flock needs a netted flight box, which section 10 lists. The flock is also a path. Drones above the galleries can light the route of one visitor's exhibition: the works that answer what that visitor asked, in the order the curator set, drawn

overhead as a corridor of light that moves ahead of them and dissolves behind. Two visitors in the same hall see two corridors. The path that the sensing layer records (section 5.4) is the path the flock draws, so the exhibition each person takes is visible to everyone in the room as a line of light above it.

The dark level. The underground level takes the black-room works and the rules a body cannot feel: inverted gravity, per-theme lift, time dilation. They are seen enacted, by projection and by the machines that obey them, rather than felt.

The galleries. A pressure-sensing floor section gives the touch and touch-memory rules a physical form, and trolleys move the works that machines may move. Every work has its twin in the virtual museum, so a visitor’s path through the gallery is recorded with the same schema as a path through the headset.

8.3 Three tiers

Every rule on the sheet falls into one of three tiers in the building [3].

TIER	WHO DOES IT	RULES	WHERE
Mechanism	matter	tilt (platform), rotation (turntable), friction and bounce (materials), softness (foam), touch and memory (pressure floor), a small gravity pulse (motion platform)	the hall, the galleries
Enactment	a machine	any rule, for bodies that are drones, trolleys, winches or arms	the void, the galleries
Sim-only	nobody, for human bodies	inverted gravity, per-theme lift, time dilation	the dark level, the screen

In the enactment tier the actuator tracks the trajectory the simulated body would have under the rule sheet, projected onto the set of accelerations that are safe for the machine and for the people near it:

$$\mathbf{a}_{\text{phys}} = \Pi_{\text{safe}}(\mathbf{a}_R)$$

The difference between what the machine did and what the simulation asked is logged, and that log is one of the twin-fidelity measures of section 10.4.

9. World models and robots

The evidence is the frontier survey compiled for this lecture, with every arXiv record checked through the arXiv API [45].

9.1 The field in October 2026

World models come in two architectures. Pixel-native models take a prompt, a text or an image, and generate a navigable video of a world. Genie 3 and its product form Project Genie, Odyssey-3, Runway’s GWM Worlds 2 and Tencent’s Hy-Nomas are of this kind [45]. They are interactive, they look alive, and their physics is learned from video, which means it is implicit and cannot be set exactly. The second architecture keeps the state and the rules in an engine and uses a neural model only for appearance. Matrix-Game 4.0, published in October 2026, is the clearest current example: a game engine holds the state and the rules, a coding agent turns the player’s language into “new rules registered into the engine”, and a causal video model renders the result at 480p and 60 frames per second with 112 ms of latency on one consumer GPU [46]. The survey summarises the distinction as “interactive is not physically controllable”: every pixel-native model above accepts text, but the only systems whose physics can be set exactly are engine-backed or are simulators [45].

A pavilion is a world model of the second kind. Its state and rules live in an engine (the browser rigid-body engine of section 7 today, a Unity scene in the live world, a standard robotics engine after the port of section 10). The rules are written on the pavilion’s sheet as eighteen typed fields. ALAN edits the sheet from language, as Matrix-Game’s coding agent does for games, and a validator and a benchmark check the edit and ALAN’s own prediction of its effect (section 7.2). The pavilion differs from a game in its content rather than its architecture. A pavilion holds works, and each work carries a theme vector that the equation

of motion treats as a force, so meaning is part of the dynamics. Visitors are boundary conditions: the engine treats a visitor as a body with an exclusion radius, and the time bubble and the safety terms are functions of where the visitor is. The room has a curated path, the sequence of works a curator intends and the sequence a visitor actually takes, which the sensing layer of section 5 records as one event schema. The public comes back, which makes the room a source of repeated, consented human data rather than a one-off study.

9.2 Measured failures: Earth priors

The documented failure of learned world models is that they carry Earth's physics into worlds that are not Earth. The survey collects the measurements [45].

Video models fall under Earth-like gravity whatever they are asked. A study of video generators under reduced gravity found an effective g of 1.81 m/s^2 where the prompt specified sub-Earth values [45]. PhysEditWorld, published in 2026, built a dataset in Unreal Engine 5 by replaying the same action traces at gravities from 0.05 to 20 times Earth, about 60 million frames, with gravity passed to the video model as a text token. Zero-shot, the Wan2.2 model ordered gravities correctly 33% of the time, which is chance; after low-rank fine-tuning the ordering reached 100% and the coefficient of determination rose from 0.066 to 0.57 [47]. The range covers no negative and no tilted gravity. PhysWeep, published in 2026, audits the slope from the requested physical parameter to the realised one and finds it indistinguishable from zero for open image-to-video generators, two of three of which “converge to one of a small number of fixed, wrong values selected by the sampling seed” [48].

Language models show the same reversion. Parallel Physical Worlds, 2026, places frontier language models in three counterfactual worlds, one where force is proportional to velocity, one Aristotelian, one with decay, and scores 6 of 15, 6 of 15 and 0 of 15 respectively; the models slide back to standard-physics ratios [4]. QuantiPhy reports that a vision-language model's score on quantitative physical reasoning falls from 56.1 to 15.4 when the prior it is told to use is counterfactual [49]. On the measurement side, World Models' Last Exam in Physics, October 2026, poses 40 measured tasks across mechanics, optics, fluids, thermal, electromagnetic and surface-tension physics; the best of eight models scores 57.76 of 100 [50]. CausalCascade, October 2026, tests nine image-to-video models on 2,832 counterfactual variants and reports “weak selective responsiveness” [51].

Two further results apply. GAUGE, 2026, measures three robotics engines and six video models against real measurements and finds “no uniformly faithful physics engine”, so an engine is a specification rather than ground truth, and every result must log the engine, its

version and its solver [52]. Counterfactually robust physical world models, 2026, show that imposing structure, an energy and a dissipation port, lets a model answer changed-law queries correctly where unstructured models fail on held-out parameters [53]. The pavilion’s rule sheet is such a structure, written explicitly rather than learned.

ALAN is a language model, so its predictions about an edited room will tend to regress toward Earth. Section 7.2 makes this a measured property: every edit carries a stated prediction and the simulator scores it. The Earth-reversion index of section 9.6 is the same measurement applied to any predictor.

9.3 The robotics standard: a world model as policy evaluator

Robotics has settled on a standard for whether a world model is a good simulator. Roll a set of robot policies out inside the world model, predict each policy’s success, run the same policies in reality or in a trusted simulator, and report the correlation between predicted and real success: Pearson r , Spearman ρ , and the rank-violation count MMRV [45]. The 2026 numbers are consistent. WEAVER, 2026, a multi-view latent world model, reports $\rho = 0.870$ against real success and a 38% improvement when used on top of a $\pi 0.5$ policy [54]. SC3-Eval, 2026, evaluates robot foundation models through self-consistent video generation and reports $r = 0.929$ over seven policies [55]. RoboWorld, 2026, a neural simulator, reports $r = 0.989$ and $\rho = 0.970$ against RoboArena [56]. Physical Canvas, October 2026, adds touch to vision and raises the correlation from 0.602 to 0.885 [57]. DreamDojo (2026, $r = 0.995$) and Ctrl-World (ICLR 2026) belong to the same line, as does the evaluation of robot policies inside a Veo world simulator over more than 1,600 real trials [45].

The policy sets are small, seven or eight policies, and the numbers are self-reported [45]. None of the evaluations tests under shifted physics: the world model is judged on Earth, with Earth-trained policies, in Earth-like scenes. The pavilion can add the same statistic under a changed rule sheet.

9.4 The gap: no benchmark perturbs the physics

Robustness benchmarks for robot policies have multiplied in 2026. LIBERO-Plus (CVPR 2026) reports ten models dropping “from over 95% to below 30%” under seven perturbation types; LIBERO-PRO and LIBERO-VPro (2026), RoboArena, RoboTwin 2.0, RoboDojo (2026) and RoboWits (2026, 208 mutated tasks on which vision-language-action models collapse) extend the pattern, and an inductive-generalisation protocol published in 2026 gives an accepted out-of-distribution ladder [45]. What they perturb is visual appearance, layout, language and objects. None perturbs the physics itself: gravity direction, viscosity, a rotating

frame, a time-dependent law [45]. Simulators can do it, and inverted gravity in half of 4,096 parallel worlds is a tutorial-level feature of MuJoCo Warp [45], but no benchmark reports policies or world models on that axis.

9.5 Adapting to unseen physics at test time

The most active robotics thread of autumn 2026 is adapting a frozen model to physics it was not trained on. T2-z, October 2026, keeps a world model frozen and optimises an environment latent at test time to match a few interactions, adapting to unseen friction, mass and what its authors call “human-defined environmental rules”; prediction error falls from 15.20 to 7.00 and task success rises from 33.7% to 73.0% [58]. SCOUT, 2026, adapts through dynamics-prediction error in a shared belief latent and transfers from simulation to reality [59]. PhysiCalWorld, October 2026, derives a “physics descriptor token” from one exploration clip and generalises beyond the training ranges of mass, damping and friction [60]. FADA, 2026, aligns humanoid dynamics from about two minutes of target data [61].

The phrase “human-defined environmental rules” in T2-z joins this literature to the pavilion. A rule sheet is a human-defined environmental rule, and so is a meaning-dependent force. Works that move toward a visitor who names their theme are a legible class of unobservable dynamics that only a semantic agent can announce, and the kind of shift these methods are built to absorb. The pavilion is a generator of such shifts, continuous in severity and arbitrary in form.

Humans carry an Earth prior as well. Interception studies in virtual reality show that people catch balls under Earth gravity better than under zero gravity and assume acceleration even for constant-velocity targets; internal models of gravity shape time-to-contact judgements; and reviews describe gravity as a prior so strong that full adaptation to Earth-discrepant visual gravity is “nearly impossible” in VR on Earth [45, 62]. No published work measures humans and world models adapting to the same counterfactual physics side by side [45]; a multiplayer pavilion can.

9.6 The pavilion as test-bed: exhibition as path

Put a human visitor, a robot policy and a world model in the same pavilion, change the rules the same way for all three, and measure the same thing.

The path. For a visitor, the path is the sequence of positions, dwells and gaze targets that the event schema of section 5.4 records: visitor, work, time, position, gaze target, dwell, question, and an affect signal where the visitor has opted in. For a robot, the path is its trajectory. For a world model, the path is its rollout. One schema holds all three, and the

exhibition is the set of paths a room produces. The data the museum collects for curators is the data the experiment needs.

Adaptation curve. For a subject S (human, world model or policy) and an edit x , the error on trial k is fitted as

$$e_k = e_{\infty} + (e_0 - e_{\infty})e^{-k/\kappa_S}$$

and the time constant κ_S and the floor e_{∞} are compared across subjects on the same edits [3]. Section 9.5 names the methods to compare on the model side; the human side has no published counterpart.

Earth-reversion index. From a predictor’s rollout under the true gravity g_{true} , fit the effective gravity $\hat{g}_{\text{eff}}$ by least squares on free-flight segments, $y(t) = y_0 + v_0 t + \frac{1}{2} \hat{g} t^2$, and define

$$\mathrm{ERI} = \mathrm{clip}\left(\frac{|\hat{g}_{\text{eff}} - g_{\text{true}}|}{|g_{\text{Earth}} - g_{\text{true}}|}, 0, 1\right)$$

so that 0 means the predictor follows the edit and 1 means it stays on Earth [3]. The same construction applies to friction from deceleration fits, viscosity from velocity decay and rotation from the curvature of rolling paths. The requested-versus-realised slope and R^2 of PhysWeep [48] and the ordering accuracy of PhysEditWorld [47] are reported alongside for parameter-conditioned models.

Policy correlation under shifted physics. Over at least eight policies, report Pearson r , Spearman ρ and MMRV between world-model-predicted and simulator-measured success, as WEAVER and RoboWorld do [54, 56], and report it per rule sheet. No published evaluation reports it under shifted physics [45].

Safety. Because people share the room, the safety numbers sit in the same table: minimum separation $d_{\min} = \min_t (\|x_R - x_V\| - r_R - r_V)$; the time-to-collision distribution when closing, $TTC = -d/\dot{d}$ for $\dot{d} < 0$, and the fraction below one second; contact count, relative speed at contact and transferred energy $\frac{1}{2} m_{\text{eff}} \|v_{\text{rel}}\|^2$; and the safety-filter intervention rate and false-intervention rate [3]. These are a proxy inspired by ISO/TS 15066’s power-and-force limiting and speed-and-separation monitoring, rather than a compliance measure [63].

Protocol. Deterministic replay with a fixed time step and seeded random numbers; pre-registered splits of interpolation (within the training range of Earth plus randomisation), extrapolation (outside it) and new law form (a velocity-, position- or meaning-dependent law); at least three seeds; confidence intervals on every reported number; the engine, version and solver logged with every result [52, 3]. For the human part, an ethics protocol and the consent rules of section 5.5.

Data source. A museum has many visitors, repeated exposure, multiplayer sessions and curated content. For robotics this is unusual: human data at a scale laboratories struggle to recruit, collected with consent in a public venue.

9.7 Three experiments

E1. Specification match and the counterfactual-physics ladder. Hypothesis: structured or calibrated predictors degrade gracefully on shifts of magnitude but fail on new law forms, and pixel-native world models revert to Earth. Method: the same scenes and the same logged action traces replayed under graded edits, following the PhysEditWorld protocol [47], along the axes the rule sheet provides: gravity magnitude from 0.25 to 4 times Earth, gravity direction from 0 to 180 degrees, friction, restitution, viscosity, rotation rate, a local lift field, a time-dilation bubble, a meaning-dependent force. Splits as in the protocol above. Metrics: error against severity, the Earth-reversion index, adaptation curves. Baselines: an analytic Earth predictor, a system-identification predictor that fits g , μ and v from the first N steps, a learned dynamics model, video world models zero-shot and fine-tuned on a few edited rollouts, and the test-time adaptation methods of section 9.5 [45]. The browser lab runs the sweeps for gravity, tilt, friction, viscosity, rotation, softness and the time bubble with twelve seeds and reports the Earth-model error (section 7.3); the comparison predictors and the port to a standard engine come next.

E2. Language-specified physics edits, verified. Hypothesis: a typed specification with a verification loop gives better calibration than free-form code generation and than no verification, and makes the Earth reversion of section 9.2 a measured, correctable property

of ALAN. Method: a set of 300 to 500 visitor-style requests (“make it feel like the room is underwater”, “let ideas about care rise”) with gold typed specifications from two or three annotators and reported inter-annotator agreement. ALAN outputs a specification and a numerical prediction with a range; the simulator scores both. Metrics: specification exact match, execution success, prediction error and calibration (Brier score, expected calibration error), correct clarification rate on ambiguous requests, correct refusal of edits that break the safety envelope [45]. Comparison: Code2Worlds-2 [64] and Text2Sim [65] style code generation, a raw language model writing engine markup, a human expert. The lab scores single predictions today; the request set, annotators and comparison systems come next.

E3. Human-robot adaptation and safety under altered physics. Hypothesis: humans, policies and world models adapt to the same edit with measurably different time constants, and a latent safety filter reduces close calls under edited physics at a measurable cost in task time. Method: a within-subjects study with about 24 to 30 participants under an ethics protocol; VR visitors share a pavilion with a simulated robot performing a pushing or handover task; conditions are Earth against edited physics, no filter against a latent safety filter, and ALAN announcing the change against silence. Measures: the safety set above, the human adaptation curve, task load, perceived safety and trust, and optional gaze and affect signals from the sensing programme. A physical follow-up uses a tabletop arm on a tilting or rotating platform for the edits that exist physically [45].

9.8 What has not been published

Each leg of the test-bed exists separately. Interactive worlds steered with text exist. Engine-held rules written by a language agent and rendered neurally exist, in Matrix-Game 4.0 [46]. Text-to-simulation by language agents exists: Text2Sim [65], GS-Agent [66], Code2Worlds-2 [64]. Gravity as a controllable variable for world models exists, in PhysEditWorld [47]. Requested-versus-realised audits exist, in PhysWeep [48]. Altered-law reasoning tests for language models exist: DiscoverPhysics (2026, 22 N-body worlds with screened and fractional-power gravity, where the best agents solve about half and low trajectory error does not imply a correct explanation) [67], and Parallel Physical Worlds [4]. World models as policy evaluators exist, and so does test-time adaptation to unseen physics [54, 56, 58, 59, 60].

What has not been published, as far as three independent searches found by October 2026 [45], is the combination: a language agent that edits physical laws rather than rescaling a constant, with inversion, local per-concept lift, viscosity fields, rotating frames, time dilation near a person and meaning-dependent forces; each edit compiled to a typed specification and checked in simulation, with the agent’s own quantitative prediction scored against the outcome; Earth-trained predictors and robot policies scored on adapting to those edits; and

real humans sharing the room, measured on the same curve. The closest comparisons are PhysEditWorld, DiscoverPhysics, PhysWeep, T2-z and Matrix-Game 4.0. Section 10 adopts what the survey asks for: a deterministic physics backend in a standard engine rather than hand-written browser code, continuous graded parameters rather than anecdotes, an open environment and logged data, standard baselines, a physical twin for the subset of edits that exist physically, pre-registered splits, multiple seeds, confidence intervals and an ethics protocol for the human part [45].

For an audience in Guangzhou the practical toolset is largely local and open. Tencent publishes Hy-Nomas, an open interactive video world model that runs on one consumer GPU, with HY-World 2.0 and WorldPlay; Skywork publishes Matrix-Game; Alibaba publishes RynnWorld-4D and Qwen-RobotWorld; Unitree publishes the UnifoLM-WLA world-language-action model; AgiBot publishes Genie Envisioner 2.0 with a built-in reward model. All of these publish weights or demos on domains reachable from mainland China [45, 68]. The pavilion’s comparison predictors and policies can be drawn from them.

10. The Guangzhou lab

We want to build the lab here, with you. The work of sections 7 to 9 needs the following.

10.1 Capabilities

The test-bed needs five capabilities.

Server-authoritative physics in the virtual museum. The live world runs physics on each client; the rule sheet of section 7 runs in a browser lab. The pavilions need one authoritative simulation on the server so that every visitor, every robot policy and every world model sees the same room under the same edit, with deterministic replay and the same seeds. This is a port of the rule program to a standard engine with per-world gravity and the rotating-frame forces added as per-step terms, as section 9.8 requires.

Eye tracking in headsets. Head pose is recorded; gaze target, dwell and order of attention are the next headset signal (section 5.3) and the path data of section 9.6 depends on them.

One event schema across the virtual museum and the building. The schema is defined (section 5.4); the building side, position from a tag or image recognition and gaze from eye-tracking glasses, needs its instrumentation.

A Chinese-language corpus and voice for ALAN. ALAN answers in Chinese today from a corpus assembled and from curatorial sources outside the Western canon (section 4.2). A corpus that is Chinese in its sources, not only in its output language, is a curatorial job as much as a technical one, and it is the part of the lab that depends most on who is in the room.

Fine-tuning and world-model evaluation on local GPUs. The fine-tuning work of section 4.7 ran on rented accelerators abroad. The comparison predictors of E1 and the test-time adaptation methods of section 9.5 are open and run locally; the evaluation needs a cluster where the data stays.

10.2 Equipment

The list follows the three tiers of section 8.2: what matter does, what machines enact, what stays virtual. Quantities depend on the space.

- Headsets with eye tracking, for the visitor path and for E3.
- Optical motion capture, for the physical twin: the position of every body in the room, logged against the simulation.
- A pressure-sensing floor section, for the touch and touch-memory rules and the Floor Memory commission.
- A netted flight box with micro-drones, for the constellation of section 8.2: tens of drones standing for thousands of points in the virtual museum.
- Omni-wheel trolleys, for works that machines move under the enactment tier.
- A small motion platform, for tilt and for the small gravity pulse: the mechanism tier.
- Projection, so that the full state of the virtual museum is visible in the room, as the translucent screen of the building design would show it.
- Opt-in wearables for pulse and heart rate, under the consent rules of section 5.5.
- A GPU cluster for fine-tuning, world-model evaluation and the server-side simulation.

10.3 Roles

- An engine and Unity engineer, for the server-authoritative port and the live world.
- A robotics and controls engineer, for the drones, trolleys and platform, the safety filter and the physical twin.
- A machine-learning engineer, for fine-tuning, the comparison predictors and the adaptation methods.
- An affective-computing researcher, for the affective layer of section 5 and the human measures of E3.
- Curator-researchers for non-Western and Chinese-language corpora.

- Artists in residence, who answer the commission sheets of section 7.4 and write new ones.
- A producer.
- A data-protection lead from the first day, because the sensing programme and the human experiments depend on consent being designed in from the start.

10.4 Research programme

The programme is the three experiments of section 9.7, E1 specification match and the counterfactual ladder, E2 verified language edits, E3 adaptation and safety with people in the room, and three further studies that the building side needs. Twin fidelity: the error between the simulated room and the motion-captured physical room, per tier, for the edits that exist physically. Legibility: whether visitors read a rule program as physics rather than as video, measured as motor adaptation, and whether they identify theme clusters in a moving constellation better than chance. The affective corpus: the dataset of section 5, responses to works labelled by curators from the cultures the visitors come from, which the ALAN paper describes and which is still to be made.

10.5 What comes next

AREA	NEXT
The virtual museum	multiplayer and voice at scale; ALAN on the two-dimensional map; the Education Pavilion programme
Sensing	eye tracking; the wearables and the heartbeat pulse; the building side of the event schema
The building	the physical museum with Clouds AO
Art and research	the four commissions; the flock flown for real; the test-bed experiments E1 to E3

11. Limitations, risks and open questions

11.1 ALAN as editor

ALAN's predictions are scored, and the score is low. In a test run of the standalone demo in October 2026, through Amazon Bedrock on Claude Sonnet 5, ALAN's predictions were right in direction on 12 of 19 counts across the seven example requests: right on every count for the touch-sensitive surface and for the return to Earth, right on two of three for the upside-down room, and wrong on all three for the memory-foam floor and for the time bubble, where it expected effects the benchmark did not show. One request in seven returned malformed JSON and fell back to the offline parser, as designed. There is no request set, no annotators and no calibration measure yet; that is experiment E2. The Earth-reversion tendency of section 9.2 applies to ALAN as to any language model, and the upside-down room is the case where it shows: ALAN expected a delivery rate of 3 of 12 where the room gave 10, because the task is unchanged for a robot on the ceiling.

11.2 The affective layer

The affective layer is a design and a data programme. No biometric data is collected in either space, no eye tracking is deployed, and the cultural-label corpus is still to be made. The ALAN paper has no quantitative experiments. The consent rules of section 5.5 are written and bind the programme.

11.3 Fine-tuning

The voice adapters were trained in a research line outside the production system. Their main result is the audit that found language-model fingerprints in the training data, which is a result about data provenance rather than about voice. The VideoLLaMA-3 model was trained on five works and is a prototype.

11.4 Novelty

Section 9.8 rests on three searches through the arXiv API in October 2026; a fourth search may find a prior combination. Each leg of the test-bed exists separately, and what is new is the combination, the content and the data source.

References

Internal documents are cited by name and date and by their path in the lecture's research folder (`GUANGZHOU_VR_AI_LECTURE_2026/research/`) or in the repository; Drive identifiers

are given where the document lives on the institution's shared drives. Published work is cited by title, authors as recorded in the survey, and arXiv identifier or venue.

1. Clouds AO (Masayuki Sono, Ostap Rudakevych), Pavilion Catalog, issue Final; Full Catalog Matrix revised in the Education Pavilion issue, 2026. Page renders: `research/style/catalog/` and `research/style/edu_workshop/`.
2. Clouds AO, Physical Space Interiors, concept design; with Institution's briefs "Building deck, New York" and "notes for the physical space in NYC"; summarised in `research/drive/CLOUDS_AO_DESIGN.md` §2.
3. Rooms as rule programs: art, physics simulation, coupled spaces, drones and robots. Internal concept and mathematics document, October 2026. `research/CONCEPT_ART_PHYSICS_MATH.md`.
4. Dong Zhang, Parallel Physical Worlds. arXiv 2607.00276.
5. Pavilion Physics Lab, benchmark export of October 2026, `demo/benchmark_presets_2026-10-26.json`; lab source `demo/pavilion_physics_lab.html` and the standalone build `demo/standalone/index.html`.
6. Physical and virtual art after drone flocks and reactive rooms: a lineage dossier compiled for this lecture, October 2026.
7. Tom Cebic, "Art as Simulation, Exhibition as Path", lecture deck, October 2026, sheets 1 to 32 and their speaker notes (`deck/src/project/slides/`).
8. Institution, Culture Commons Forum Full Report (52 pp., Drive `1ptltNHUXca7WAIxNYl3Weu4kv89DBady`) and the Culture Commons Forum 2026 programme and hub list, `institution_tv/site/config/streams.json` and `content.json`, 2026.
9. Institution, China Update, 2026 (Changjiang Scholar project approval, objectives and outcomes). Board pack `05_Legal_and_Corporate/Institution_China_Update_2026-02.txt`.
10. VR world research notes, `research/local/vr_world_notes.md`, compiled October 2026 from the Unity source (`Assets/Scripts/Alan/AlanClient.cs`, `AlanCompanionController.cs`, `SceneLoading/ScenePreloader.cs`, `PlayerPavilionMover.cs`, `MapFlyer.cs`, `NetCode/`), the audit documents of 2026, and live HTTP checks of October 2026.
11. Institution VR world, public database snapshot of October 2026 (`institution-vr-clean/OfflineSnapshot/pavilions.json`, `content.json`, `pavilion_authors.json`) and the render manifest `research/vr_renders/manifest.md`, October 2026.
12. Nils Gallist, Institution CMS Description. Drive `1wITH38bEyZqRoOpjTQhG1uuGq3EUOEJ8`; export `research/drive/raw/nils_institution_CMS_description_2025-03-29.txt`.

13. Lucilla Grossi, Institution Platform Development Schedule, October 2026. Drive 1iABxfREoseKoSmoWrstLJou7X5TaeP3GQ1itQzbjT3U.
14. Clouds AO, Education Pavilion, issue Additional, 2026, including the Full Catalog Matrix and the Education Pavilion program diagram. File: Institution-Education-Workshop Pavilions_260427.pdf; page renders in research/style/edu_workshop/.
15. Clouds AO, “Institution / Museum of Tomorrow, Progress, Concept Design”, reproduced as slides 22 to 30 of the Institution Virtual Space Deck. Drive 1vcWPa0f78oQbAtbuKmVKzzmor-NJpHOL0zhDx5n8RHE; summarised in research/drive/CLOUDS_AO_DESIGN.md §1.
16. Lucilla Grossi, Digital Space Roadmap. Drive 1mzLY_NL-rJI9dlvX65jx6NUvcndxO3hX.
17. Hub inventory of the virtual museum, 2026: the full topic, pavilion, portal and content graph with coordinates (stars-project/docs/unity/HUB_INVENTORY_2026-09-19.md and hub_positions_2026-09-19.json).
18. Institution, “notes on VR”, QA notes. Drive 1RTyKWLhHbMDt7LuzgQi8wo0PVYVFincR.
19. Nils Gallist (Extended Reality), Data Architecture Analysis: Institution Lab APS Virtual Environment, revision 226. Drive 1_QQItaJuhAF06KyhO9pS8NQLVXVjt1Z00Ptxhh5valA; export research/drive/raw/nils_data_architecture_analysis.txt.
20. Curator backend, VR companion service, ai_assistant/app/services/vr_alan.py (three-pass retrieval, public-tier scope, Bedrock generation, pre-publish check, memory compaction), read October 2026.
21. Institution, VR ALAN Unity Integration contract, ai_assistant/docs/vr-alan-unity-integration.md (605 lines).
22. Curator backend configuration, ai_assistant/app/core/config.py (model identifiers and fallbacks, temperature, embedding model, answer budgets, memory settings), read October 2026.
23. Institution, Curator Assistant project context, AI_Stuff/CLAUDE.md, last verified 2026, with the Atrium and gateway sections dated 2026.
24. Curator corpus review, 2026: document and chunk counts and research paths, as recorded in research/local/LOCAL_DOSSIER.md §1.2 and the platform cards of the October 2026 pack.
25. Institution, Curator Voice and Disclosure Policy, ai_assistant/app/core/canonical/curator_voice_and_disclosure_policy.md, and its implementation ai_assistant/app/core/voice_policy.py.
26. DeepSeek-AI, “DeepSeek-R1: Incentivizing Reasoning Capability in LLMs via Reinforcement Learning”, arXiv 2501.12948 (the R1-Distill-Qwen-32B checkpoint).

27. Tim Dettmers, Artidoro Pagnoni, Ari Holtzman, Luke Zettlemoyer, “QLoRA: Efficient Finetuning of Quantized LLMs”, arXiv 2305.14314.
28. Voice adapter research line: MODEL_CARD.md, engine/config.py (adapter versions v1 to v6, DPO configuration, hardware), V7_PREREGISTRATION.md (2026) and VOICE_CORPUS_AUDIT.md (2026), in the stage3_sft directory; summarised in research/local/LOCAL_DOSSIER.md §1.1.
29. Boqiang Zhang et al., “VideoLLaMA 3: Frontier Multimodal Foundation Models for Image and Video Understanding”, arXiv 2501.13106.
30. Zeyang Zhang, Internship Summary Form, AI Research Intern, Institution Lab, 2026 (VideoLLaMA-3 fine-tune on five artworks, Hugging Face deployment, model sharding). Drive 1QVvJ2ZXp229MIAA5gP6n6GJKYvQIa4yF.
31. Institution AI Team, Art Content Samples with curator-written labelling texts (Kozyra, Ortega, Lacy, Robakowski), Drive folder AI_Team / Art Content / Samples; summarised in research/drive/DRIVE_DOSSIER.md (b) 11.
32. Xia Wang, Zeyang Zhang, Fabio Cavallucci, Tomislav Cebic, Jonathan Sprinkle, “ALAN: A Demonstration of an Emotionally-Aware AI Companion with Human Expert Guidance”, submitted to the ECAI demonstration track. Drive: AI_Team / Papers Submitted / ECAI ; local export
research/drive/raw/alan_demonstration_emotionally_aware_companion.txt.
33. Jeff Hawkins, A Thousand Brains: A New Theory of Intelligence, Basic Books, (cited by [32] as the source of the cortical-column voting mechanism).
34. Institution, ALAN VR service, ai_assistant/app/services/vr_alan.py (answer budgets lines 176 to 178; public-tier retrieval line 363; tables VRVisitorProfile, VRConversationTurn), and ai_assistant/docs/vr-alan-unity-integration.md. Code as of October 2026; summarised in research/local/LOCAL_DOSSIER.md and research/local/vr_world_notes.md.
35. Institution VR client, IVC/Assets/Scripts/Alan/AlanClient.cs (calls at lines 52, 76, 95, 118), AlanManager.cs (lines 153 to 260). Source of 2026; summarised in research/local/vr_world_notes.md.
36. Institution VR backend contract, Supabase tables (likes, playlists, profiles), SP/unity/VR_BACKEND_CONTRACT_2026-09-19.md; summarised in research/local/LOCAL_DOSSIER.md.
37. Institution, “AI-Guide Business Model” (2026 version) and “Institution AI Guide” slides (19 slides), phase table and timeline slide. Local exports
research/drive/raw/institution_ai_guide_slides_2025-04.txt and
research/drive/raw/business_model_2026-02.txt; phase table reproduced in research/drive/DRIVE_DOSSIER.md, section (d).

38. Institution, research/drive/DRIVE_DOSSIER.md, October 2026: finding 9 (affective and biometric layer: no implementation or data on Drive) and section (e) (the two papers with Zeyang Zhang).
39. Institution, Vision Deck v2, slide 13 (“Phygital integration: every node has its digital twin”), 2026. research/drive/images/visiondeck_v2_p13_brooklyn-13.png; quoted in research/drive/CLOUDS_AO_DESIGN.md.
40. Extended Reality and Institution Labs APS, Data Processing Agreement and “Details of Processing (GDPR Art. 28(3)), Annex A”. Drive 1JYfVfGA29Zajd6taRoM54GYnqFM6lDSx; local export research/drive/raw/nils_gdpr_details_of_processing_annexA.txt.
41. Institution, ATRIUM_EXHIBITION_WORKFLOW.md, walkthrough of the Atrium engine with fixes F1 to F10, 2026.
42. Institution, AGENT_GATEWAY_PLAN.md and ai_assistant/mpc_agent/agent_gateway/(email_gateway.py, slack_socket.py, engine_bridge.py, core/fast_loop.py, core/send.py), 2026.
43. Institution, translate_app/ on the atrium box (translate-institution.service, forum_ingest.py, dub_mixer.py, the /interpreter console), built in 2026; FORUM_INTERPRETER_RUNBOOK.md.
44. Institution, FORUM_HANDOFF.md (2026), FORUM_VENUE_AUDIO.md, institution_tv/site/config/streams.json; Culture Commons Forum, 2026.
45. World models, physics and robots: the technical frontier as of October 2026. Internal survey compiled for this lecture, with arXiv records checked through the arXiv API on that date. research/external/WORLD_MODELS_FRONTIER_2026-10.md.
46. Matrix-Game 4.0, Skywork AI. arXiv 2610.27999, October 2026.
47. PhysEditWorld, Hu et al., with Jiaya Jia and Xiu Li. arXiv 2606.26694, 2026.
48. PhysWeep, Khanbayov and Kurban. arXiv 2609.06207, 2026.
49. QuantiPhy, CVPR 2026 (secondary source in the survey; scores 56.1 to 15.4 under a counterfactual prior).
50. World Models’ Last Exam in Physics, Gao et al., with Qinhuai Na. arXiv 2610.08791, October 2026.
51. CausalCascade. arXiv 2610.15562, October 2026.
52. GAUGE, Weinan Zhang, Chunhua Shen et al. arXiv 2608.05948, 2026.
53. Counterfactually robust physical world models, Wang, Priye, Wei and Ling. arXiv 2610.00280, 2026.
54. WEAVER. arXiv 2606.13672, 2026.
55. SC3-Eval, Levine, Vuong et al. arXiv 2606.18610, 2026.
56. RoboWorld, Kimin Lee et al. arXiv 2607.01060, 2026.
57. Physical Canvas, Xing et al., with Qichao Zhang. arXiv 2610.28613, October 2026.

58. T2-z, Fei-Fei Li, Jiajun Wu, Leonidas Guibas, Yejin Choi, David Held et al. arXiv 2610.22364, October 2026.
59. SCOUT. arXiv 2609.36107, 2026; CoRL 2026.
60. PhysiCalWorld, Chai, Fazeli and J. J. Park. arXiv 2610.30125, October 2026.
61. FADA, Guanya Shi, Max Simchowitz et al. arXiv 2606.28476, 2026.
62. Russo et al., Journal of Neurophysiology; Delle Monache et al., Frontiers in Physiology; review in Frontiers in Human Neuroscience (secondary sources in the survey: the human Earth-gravity prior in VR interception).
63. ISO/TS 15066, Robots and robotic devices, collaborative robots; ISO 10218-1 and 10218-2.
64. Code2Worlds-2, Zhang et al., with Hao Tang. arXiv 2610.17651, October 2026.
65. Text2Sim, Xiong, T.-H. Wang, Qiao, Tao Du and Minchen Li. arXiv 2609.36593, 2026.
66. GS-Agent, Chuang Gan group with Zhou Xian. arXiv 2607.21522, 2026.
67. DiscoverPhysics, 2026 (arXiv record checked in the survey; identifier not recorded there).
68. Hy-Nomas, Tencent Hunyuan, Hugging Face model card and demo, 2026, Apache-2.0; Unifolm-WLA-1.0, Unitree, GitHub, 2026; Genie Envisioner 2.0, AgiBot, 2026.
69. The Shed, New York: a telescoping outer shell on rail wheels that deploys over the plaza (public record).
70. Perelman Performing Arts Center, New York: three theatres whose walls, floors and seating reconfigure into dozens of arrangements behind a translucent marble skin (public record).

Appendix A. The rule program schema, the equation of motion and the benchmark data

A.1 Schema

The rule sheet as the lab executes it in October 2026 [5]. Each field has a range the validator clamps to, a unit, and a neutral value that together make the Earth sheet. The last column says whether the rule has a physical twin in the building (section 8.3): a tilting floor, a turntable, a foam or membrane floor, a tank, a pressure floor; or whether it exists only in simulation.

FIELD	RANGE	UNIT	EARTH	MEANING	TWIN
gravity	0 to 20	m/s ²	9.81	gravity strength (Moon 1.62)	sim only
upsideDown	true or false		false	gravity points to the ceiling	sim only
tilt	0 to 30	°	0	gravity leans this many degrees from vertical	physical
tiltDir	0 to 359	°	0	compass direction the tilt leans toward	physical
grip	0 to 1.2	μ	0.6	surface friction; how hard the robot's wheels can push	physical
bounce	0 to 0.95	e	0.35	restitution of every surface and collision	physical
viscosity	0 to 3	1/s	0.05	drag of the medium (water about 2)	physical
pulse	0 to 0.95	fraction	0	gravity oscillates by this fraction	sim only
pulseHz	0.1 to 1.5	Hz	0.5	frequency of the gravity pulse	sim only

FIELD	RANGE	UNIT	EARTH	MEANING	TWIN
spin	0 to 1.2	rad/s	0	the room rotates: Coriolis and centrifugal forces	physical
cohesion	0 to 5		0	works attract each other by shared meaning	sim only
lift (per theme)	-1 to 2		0	per-theme weight: 0 normal, 1 weightless, 2 falls upward, -1 twice as heavy; themes water, stone, light, body, machine, memory	sim only
touch	0 to 1		0	surfaces record pressure and show it	physical
touchMemory	0.5 to 20	s	6	how long surfaces remember pressure	physical
softness	0 to 1		0	surfaces dent under pressure; objects slide into dents	physical
ripple	0 to 1		0	impacts send ripples across the surfaces	physical

FIELD	RANGE	UNIT	EARTH	MEANING	TWIN
timeBubble	0 to 0.9	fraction	0	time slows near the visitor by up to this fraction	sim only
timeScale	0.25 to 2	×	1	playback speed of the whole room (live view only)	sim only

The bodies are twelve works, each with a six-dimensional theme vector over water, stone, light, body, machine and memory, a ball, a wheeled robot and a visitor, with masses of 6 kg, 1 kg and 3 kg for work, ball and robot. The room is a disc with floor and ceiling as 72×72 grids that hold pressure and ripple fields. In the lab the twelve works are invented samples; in a pavilion they are the works in the slots (section 3.6).

A.2 Equation of motion

For body i with position x_i , velocity v_i and theme vector s_i :

$$\begin{aligned} \mathbf{a}_i = & f_i \mathbf{g}(t) - \nu \mathbf{v}_i - \\ & 2\boldsymbol{\Omega} \times \mathbf{v}_i - \\ & \boldsymbol{\Omega} \times (\boldsymbol{\Omega} \times \mathbf{x}_i) + \\ & \mathbf{a}_{\text{slope}} + \mathbf{a}_{\text{cohesion}} + \mathbf{a}_{\text{query}} + \\ & \mathbf{a}_{\text{policy}} \end{aligned}$$

with $g(t) = g_0 (1 + \text{pulse} \cdot \sin 2\pi f t)$ rotated by the tilt and its direction and inverted when upsideDown is set; $f_i = 1 - \Sigma_c s_{ic} \cdot \text{lift}_c / \Sigma_c s_{ic}$ for works and 1 for every other body; Ω = spin about the vertical; the slope term the gradient of the dent field under the body scaled by softness; the cohesion term $-\text{cohesion} \cdot \Sigma_j \cos(s_i, s_j) \cdot (x_i - x_j)$ over the other works; the query term a pull toward the visitor proportional to $\cos(s_i, q)$ when a query vector q is

set; and the policy term the robot's own control, a Coulomb-limited wheel force toward the point behind the ball that lines up with the goal, with an exclusion radius around the visitor. Integration is semi-implicit (symplectic) Euler at 120 Hz in the live view and 60 Hz in the benchmark; contacts are sphere contacts with restitution; the ball has rolling resistance. Time near the visitor runs at $1 - \text{timeBubble} \cdot w(d)$, with w falling from 1 at the visitor to 0 at the bubble radius. The Earth model is the same engine fixed at the Earth sheet; it forecasts the ball 2 s ahead and its error is the three-dimensional distance after 2 s [3, 5].

A.3 Benchmark data

Twelve seeds (1000 to 1011), 25 s each at 60 Hz, one scripted robot controller, measured October 2026 and exported as `demo/benchmark_presets_2026-10-26.jsonl` [5]. The summary table is in section 7.3. Per trial the export records the seed, whether the ball was delivered, the time to delivery, the close-call time and the Earth-model error, and the full rule sheet. The dose-response sweeps step one field through nine levels: gravity from $-2\times$ to $4\times$ Earth (negative meaning upside down), tilt 0° to 30° , grip 0.02 to 1.2, viscosity 0 to 3, spin 0 to 1.2 rad/s, softness 0 to 1, time bubble 0 to 0.9, with twelve seeds per level.

A.4 ALAN's output contract

ALAN returns JSON only: a `changes` object holding only the fields that change, a one-sentence `summary` in the present tense, and one to three `predictions`, each with a `metric` (delivery_rate, median_time, close_calls or earth_model_error), a `direction` (up, down or same), an `expected` number and a `because`. The validator clamps every number to the schema range, coerces upsideDown to a boolean, keeps only the six theme keys under lift, and lists any other field as ignored. A prediction is judged right in direction when the measured change exceeds a small tolerance in the stated direction, or stays within it for "same", and its absolute error is the distance between `expected` and the measured value. The offline parser implements the same contract with pattern rules for the example requests, so that the demo runs with no model at all.

Appendix B. Sheet-by-sheet map

The section that supports each sheet of the deck.

SHEET	TITLE	SECTION
1	Art as Simulation, Exhibition as Path	1
2	Institution: Italy, New York and now China; five public systems	2
3	The system: one exhibition, two spaces, one curator, one operations layer	1, 5.6
4	The VR world runs today: 23 pavilions, 16 open, 37 works, three languages	3.1
5	The pavilion catalog defines nine pavilion types	3.2
6	Works in their pavilions	3.3
7	Lectures, articles and the Education Pavilion	3.3
8	How a visitor moves	3.4
9	Platform architecture	3.5
10	A pavilion is data: template, row, content	3.6
11	ALAN is a system of six layers	4
12	The model layer: Sonnet 4.6 at launch, Sonnet 5.5 now	4.1
13	The corpus layer	4.2
14	The grounding layer	4.3, 4.4
15	Fine-tuning work	4.7
16	ALAN in a pavilion	4.8
17	Semantic and affective readings are separate layers	5.1

SHEET	TITLE	SECTION
18	The exhibit already produces data	5.2
19	Sensing	5.3
20	One work, two sensors: one event schema	5.4
21	Consent and data rules	5.5
22	Atrium is the operations layer	6.1
23	Atrium services in use	6.2, 6.3
24	A pavilion is a simulation; its rules go on the sheet	7.1
25	ALAN writes the rules; the simulator checks the prediction	7.2
26	Benchmark: one task, eleven rule sheets	7.3
27	Four commissions written as sheets	7.4
28	The building concept	8.1
29	The building as mechanism	8.2, 8.3
30	World models in October 2026	9.1 to 9.5
31	Exhibition as path: the pavilion as test-bed	9.6, 9.7
32	What the future projects need, and where the work stands	10

The deck is at <https://claude.ai/artifact/MGLGgUsRRXWb6JiwBLxcRy>; the standalone demo and its test log are in [demo/standalone/](#).

