

# **A Generative Education Architecture for Planetary-Scale Personalized Learning**

Integrating AI Orchestration, Data Sovereignty,  
and Trustee Governance into a Unified System

**Matthias Pochmann**

*Independent Researcher, Winterthur, Switzerland*

m.pochmann@icloud.com  
ORCID: 0009-0001-7706-2504  
February 2026

DOI: 10.5281/zenodo.18759134

This work is licensed under a Creative Commons Attribution 4.0 International License (CC BY 4.0)

# Table of Contents

The Problem — Education is Being Digitized Wrong.....	3
Related Work and Why It Falls Short.....	5
Design Principles.....	9
Architecture.....	11
L1 — The Mentor.....	12
L2 — The Learner Profile.....	14
L3 — The Didactic Model.....	14
L4 — Data Sovereignty.....	15
L5 — The Federated Knowledge Graph.....	15
L6 — The Competency Evidence System.....	17
L7 — Autonomous Hub Infrastructure.....	18
L8 — Trustee Governance.....	18
Data Sovereignty as Architectural Principle.....	19
Economics of Generative Education.....	23
Open Questions and Research Agenda.....	25
Conclusion.....	27
Acknowledgments.....	29
References.....	30

## **Abstract**

In 1984, Benjamin Bloom demonstrated that students receiving one-on-one tutoring outperform conventionally taught students by up to two standard deviations. Subsequent meta-analyses have qualified the magnitude — VanLehn (2011) found human tutoring effect sizes closer to  $d = 0.79$  — but even the conservative estimate represents a transformation: moving the average student from the 50th to the 79th percentile. For forty years, even this lesser gap remained economically impossible to close. Generative AI removes the economic barrier. But AI is being deployed within an architecture designed for industrial-age standardization: learners interchangeable, knowledge static, assessment a gate. The result is digitized education — the old system made faster — not digital education, which requires a different architecture entirely.

This paper describes that architecture. It integrates six interdependent dimensions — deep learner profile, data sovereignty, generative learning experience, federated knowledge graph, process-based competency evidence, and trustee governance — into a unified system. No dimension works without the others. A deep profile requires sovereignty; sovereignty requires governance; a generative experience requires both a knowledge graph and a profile; process evidence requires all of the above. The dependency graph forces integration.

Every component technology exists in production. Intelligent tutoring systems, privacy-preserving computation, federated knowledge graphs, edge AI, competency-based assessment, cooperative data governance — each has been validated independently. The original contribution of this paper is not any single piece but the demonstration that the pieces are architecturally interdependent and that their integration produces a system qualitatively different from any subset.

A generative education architecture is a system in which an AI Mentor — constrained by a federated knowledge graph, informed by a lifelong learner profile, governed by a trustee structure, and cryptographically prevented from exposing its data — generates the learning experience itself, for each individual, at near-zero marginal cost.

Several challenges remain genuinely open: privacy-preserving credential verification, cultural curation governance, organizational bootstrapping. None is a fundamental blocker. Each has a defined scope and a plausible resolution path. The paper names them honestly.

The question is not whether AI-powered education at planetary scale will arrive. It is whether it arrives as trustee infrastructure — where the learner owns the data and democratic bodies curate knowledge — or as a platform product,

where data is an asset and knowledge definition is proprietary. The architecture described here is designed for the former.

**Keywords:** *generative AI, education architecture, data sovereignty, personalized learning, knowledge graph, trustee governance, competency-based assessment, federated systems*

## **The Problem — Education is Being Digitized Wrong**

Digitizing education means making the old system faster — putting lectures online, automating feedback, making tests adaptive. Digital education means a different system, one where the learning experience is generated for the individual, not distributed to the mass. Current educational technology improves fragments of a broken system. The fragments get shinier. The system stays broken.

The distinction is not semantic. It is architectural. And the evidence that it matters has been accumulating for forty years.

In 1984, Benjamin Bloom demonstrated the scale of the opportunity: students receiving one-on-one tutoring outperform conventionally taught students by up to two standard deviations.<sup>1</sup> His finding was not controversial. It was replicated, discussed, and quietly shelved, because its implication was economically impossible: you cannot hire a personal tutor for every child on Earth. Generative AI changes this. For the first time, a technology can approximate the core mechanism that makes tutoring effective — responsive, adaptive, one-on-one interaction that meets a learner where they are and adjusts in real time.

How large is the gap to close? Bloom's original 1984 study reported an effect size of two sigma — the tutored student outperforming 98% of the conventionally taught class. VanLehn's 2011 meta-analysis, examining a broader set of studies with stricter controls, found human tutoring effect sizes closer to  $d = 0.79$ .<sup>2</sup> The difference matters: two sigma is extraordinary; 0.79 sigma is merely transformative. This architecture targets the full range. The conservative estimate — closing a 0.79-sigma gap for every learner on Earth — would be the largest improvement in educational outcomes in history. The aspiration is to exceed it, because the architecture enables mechanisms VanLehn's meta-analysis could not capture: lifelong learner profiles that

---

<sup>1</sup>Bloom, B. S. (1984). The 2 sigma problem: The search for methods of group instruction as effective as one-to-one tutoring. *Educational Researcher*, 13(6), 4-16. <https://doi.org/10.3102/0013189X013006004>.

<sup>2</sup>VanLehn, K. (2011). The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems. *Educational Psychologist*, 46(4), 197-221.

accumulate over years (not single-session studies), cross-domain transfer through a knowledge graph, and simulation environments that go beyond dialogue-based tutoring. Whether these mechanisms recover the full two-sigma effect is an empirical question this architecture is designed to answer.

But the technology is being deployed within the wrong architecture. The United Nations' Sustainable Development Goal 4 — quality education for all by 2030 — provides the scoreboard. As of 2023, 272 million children and young people are out of school globally, a figure that has risen since 2015.<sup>3</sup> Three hundred million students currently in school will leave without basic literacy and numeracy.<sup>4</sup> PISA scores dropped 15 points in mathematics and 10 points in reading between 2018 and 2022.<sup>5</sup> Achieving the goal would require 44 million additional teachers and \$100 billion per year in additional funding in low- and middle-income countries.<sup>6</sup> The world has spent a decade increasing investment in education and the numbers have gotten worse. The current model is not underfunded. It is structurally incapable of reaching the people it needs to reach.

Into this failure, AI is arriving at extraordinary speed. Google's LearnLM reached 170 million users across 230 countries within 18 months through integration into Google Workspace for Education.<sup>7</sup> Khan Academy's Khanmigo scaled from 40,000 to 700,000 K-12 students in a single school year.<sup>8</sup> The products are genuinely impressive. Google's Guided Learning implements real Socratic tutoring — asking questions instead of giving answers, scaffolding step by step, generating multimodal responses. In a controlled UK study, students with LearnLM tutoring were 5.5 percentage points more likely to solve novel problems than students working with human tutors alone.<sup>9</sup>

But look at what these systems actually do. They generate explanations *about* physics. They do not generate environments *in which* one experiences physics. The learner profile resets with each session. The data belongs to the platform. The knowledge definitions are proprietary. The system requires a stable internet connection and a commercial account. For the 272 million children without school — many without reliable internet — these tools do not exist.

---

<sup>3</sup>UNESCO Institute for Statistics & GEM Report (2025). 2025 SDG 4 Scorecard: Focus on the Out-of-School Rate.

<sup>4</sup>UNESCO (2022). Transforming Education Summit Discussion Paper.

<sup>5</sup>OECD (2023). PISA 2022 Results (Volume I): The State of Learning and Equity in Education.

<sup>6</sup>UNESCO (2023). Global Report on Teachers: Addressing Teacher Shortages.

<sup>7</sup>Google (2025). LearnLM: AI for Learning. Official product page.

<sup>8</sup>Khan Academy (2024). Khanmigo Year in Review.

<sup>9</sup>Google DeepMind (2025). LearnLM Tutoring Study: UK Controlled Trial Results.

The OECD has now measured this distinction. The *Digital Education Outlook 2026* found that students with access to general-purpose generative AI produce higher-quality outputs on assignments — but the advantage disappears when those students take exams without AI access.<sup>10</sup> The learning did not transfer. The OECD's language is precise: "Offloading cognitive tasks to general-purpose chatbots creates risks of metacognitive laziness and disengagement." In contrast, purpose-built educational AI with intentional pedagogical design shows sustained improvements that persist when the AI is removed.

The problem is not AI in education. The problem is AI without architecture. The distinction between digitizing education and digital education is not rhetorical — the OECD has measured it. If the current approach is architecturally broken, the question is whether the pieces for a different architecture already exist.

## Related Work and Why It Falls Short

The pieces exist. Over fifty years, independent research communities have each solved a genuine fragment of the educational challenge. What follows is not a competitive landscape analysis. It is an inventory of puzzle pieces — fair acknowledgment of real achievements and proof that the integration gap is structural, not accidental.

**Intelligent Tutoring Systems** stretch from PLATO in the 1970s to modern LLM-powered tutors. VanLehn's 2011 meta-analysis found that ITS perform roughly as well as human tutors within their domain (effect size  $d \approx 0.79$ ).<sup>11</sup> But each system covers a single narrow subject. There is no cross-curricular integration, no lifelong learner model.

**Adaptive Learning Platforms** — Coursera, Khan Academy, Duolingo — dynamically adjust content based on learner data. A 2025 systematic review of 142 studies by Hariyanto and colleagues found consistent gains in engagement and performance.<sup>12</sup> Yet they adapt *within* existing content: selecting and sequencing pre-authored materials rather than generating new experiences. Learning data stays siloed within each platform.

**MOOCs and Open Educational Resources** democratized access at unprecedented scale. But access to content turned out not to be the bottleneck. Completion rates remain between 5 and 15 percent.<sup>13</sup> A 2023 meta-analysis by Tlili, Garzón, Salha, and colleagues found an effect size of  $g = 0.07$  for open

---

<sup>10</sup>OECD (2026). OECD Digital Education Outlook 2026: Exploring Effective Uses of Generative AI in Education. OECD Publishing, Paris. <https://doi.org/10.1787/062a7394-en>.

<sup>11</sup>VanLehn (2011), see note 2.

<sup>12</sup>Hariyanto, Kristianingsih, F.X.D. & Maharani, R. (2025). Artificial intelligence in adaptive education: A systematic review of techniques for personalized learning. *Discover Education*, 4, 458. <https://doi.org/10.1007/s44217-025-00908-6>

educational resources — the content layer on which most MOOCs build — statistically significant but negligible in practice.<sup>14</sup>

**Learning Analytics** brought dashboards, dropout prediction, and knowledge tracing into institutional use. These tools measure what happened — clicks, time-on-task, scores — but not how the learner thinks. Data stays within each institution, offering no longitudinal view.

**Knowledge Graphs in Education** have demonstrated concept mapping and prerequisite tracking in narrow domains. Khan Academy once had exactly the right pattern — a dependency graph for mathematics — but discontinued it. Wikidata holds over 110 million entities but encodes no didactic dependencies. Recent work on federated knowledge graph completion (DFedKG, Zhao et al. 2025) has shown that joint learning across multiple graphs while preserving privacy is technically feasible.<sup>15</sup> But no existing system implements even a basic two-layer model separating universal scientific dependencies from culturally sovereign content — let alone the richer internal structure such a graph requires in practice.

**Competency-Based Assessment** has been validated in healthcare and vocational training through instruments like the OSCE.<sup>16</sup> An integrative review of programmatic assessment in healthcare education found that continuous low-stakes assessment supports both learning and robust decision-making more effectively than one-time high-stakes examinations.<sup>17</sup> But for K-12 academic subjects, continuous process-based assessment remains largely uncharted.

**Privacy-Preserving Computation** has reached production scale outside education. Apple's Private Cloud Compute demonstrated stateless processing where the infrastructure provider cannot access user data — not as a policy promise but as a cryptographic guarantee.<sup>18</sup> In parallel, data cooperatives have proven that citizens can govern sensitive personal data through democratic structures — most notably in healthcare, where cooperative models give

---

<sup>13</sup>Jordan, K. (2015). Massive open online course completion rates revisited. *International Review of Research in Open and Distributed Learning*, 16(3).

<sup>14</sup>Tlili, A., et al. (2023). Are open educational resources (OER) and practices (OEP) effective in improving learning achievement? *International Journal of Educational Technology in Higher Education*, 20(1), 54. <https://doi.org/10.1186/s41239-023-00424-3>

<sup>15</sup>Zhao, C., et al. (2025). DFedKG: Diffusion-Based Federated Knowledge Graph Completion. *Data Science and Engineering*, 10, 639–652. <https://doi.org/10.1007/s41019-025-00292-z>

<sup>16</sup>Harden, R. M. (1988). What is an OSCE? *Medical Teacher*, 10(1), 19–22.

<sup>17</sup>Schut, S., et al. (2021). Where the rubber meets the road — An integrative review of programmatic assessment in health care professions education. *Perspectives on Medical Education*, 10(1), 6–13. <https://doi.org/10.1007/s40037-020-00625-w>

<sup>18</sup>Apple (2024). Private Cloud Compute: A new frontier for AI privacy in the cloud. Apple Security Research.

individuals control over their health records while preserving research utility.<sup>19</sup> Neither mechanism has been applied to education.

**EdTech Frameworks** — SAMR, TPACK, and their variants — guide teachers in adopting technology within existing classrooms. Cherner and Mitchell's 2021 analysis examined nine such frameworks; the EdTech Hub identified seventeen; the UK Department for Education reviewed seventy-four.<sup>20</sup> Every single one focuses on technology adoption within the classroom. None addresses system-level architecture.

The natural instinct is to propose integration: take the best tutoring system, connect it to the best adaptive platform, add a knowledge graph, wrap it in privacy-preserving computation. This cannot work through incremental assembly, because the dimensions form a dependency graph.

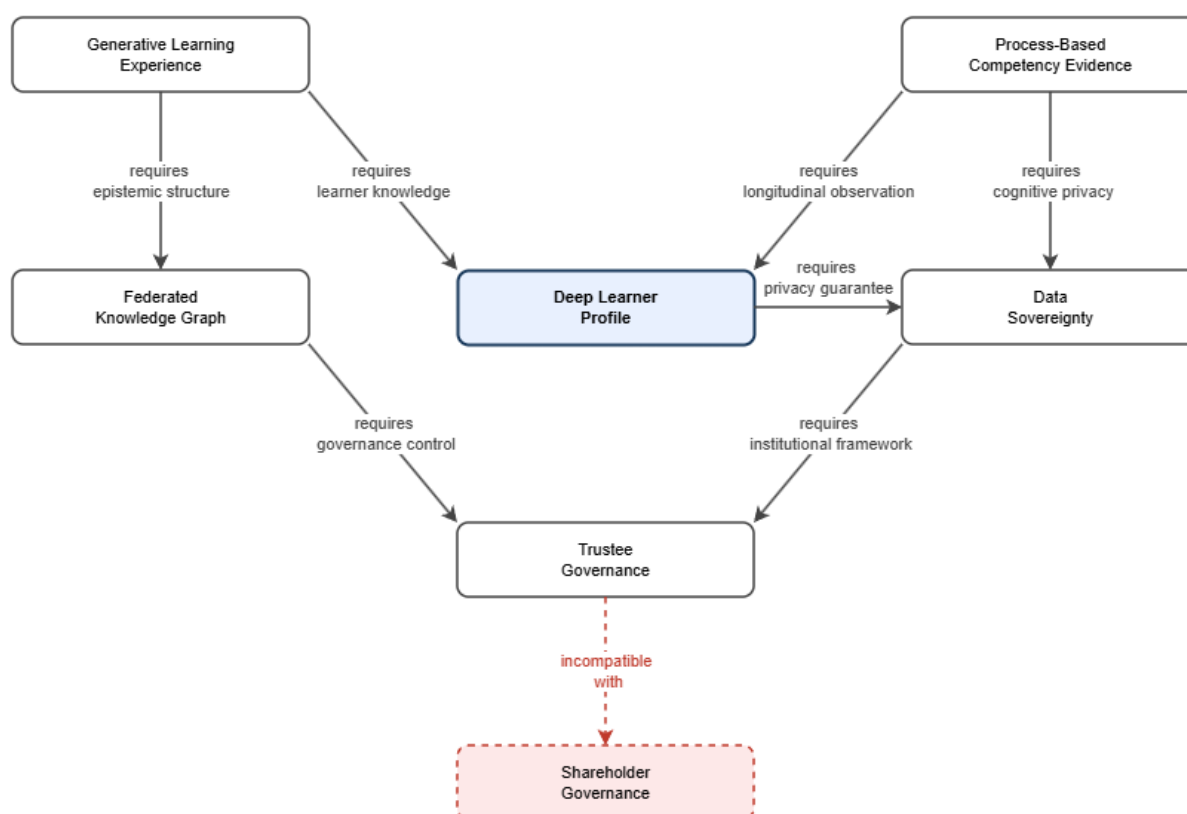
A deep learner profile — one that captures how a person thinks, not just what they answered — requires data sovereignty. Collecting intimate cognitive data without the architectural guarantee that no unauthorized party can access it is irresponsible. A generative learning experience requires both a knowledge graph (to ensure what is generated is epistemically sound) and a deep profile (to ensure it fits this learner). Process evidence requires the deep profile for longitudinal tracking and requires sovereignty because process evidence is cognitive fingerprinting — it reveals how you approached a problem, where you hesitated, how long it took you to correct course. Data sovereignty requires governance: a technical privacy architecture without an institutional framework is a lock without a key policy. A federated knowledge graph requires governance for its culturally sovereign content: who decides what counts as valid knowledge? And trustee governance is structurally incompatible with shareholder-driven governance — the obligations point in opposite directions the moment data becomes commercially valuable.

---

<sup>19</sup>Hafen, E., Kossmann, D., & Brand, A. (2014). Health data cooperatives — Citizen empowerment. *Methods of Information in Medicine*, 53(2), 82–86. <https://doi.org/10.3414/ME13-02-0051>. See also the European Data Governance Act (2022) and the MyData Global initiative for broader policy convergence.

<sup>20</sup>Cherner, T. & Mitchell, C. (2021). Deconstructing EdTech Frameworks Based on their Creators, Features, and Usefulness. *Learning, Media and Technology*, 46(1), 91–116. <https://doi.org/10.1080/17439884.2020.1773852>





**Figure 1.** *Interdependence Graph — Why the Six Dimensions Cannot Be Solved Independently*

These dimensions are architecturally interdependent. They cannot be solved independently and bolted together. They must be designed as an integrated architecture. This is not an opinion — it is a structural observation about the dependency graph.

Every one of these puzzle pieces has existed for years — some for decades. What has changed is that generative AI makes them integrable. Before LLMs, a deep learner profile was inert data — useful for analytics dashboards but not for generating a learning experience. Before edge AI, privacy-preserving computation meant the learner could not be reached offline. Before federated learning, a knowledge graph could not be jointly maintained without centralizing the data. Generative AI is the catalyst that transforms these independent components from individually validated but inert pieces into an architecture that can function as a whole.

From here, the honest differentiation from the most capable actor in the space. Google's LearnLM is pedagogically sound and genuinely helpful. The argument is not that Google builds bad educational technology. The argument operates on three levels. First, what LearnLM achieves: real Socratic tutoring, measurable learning gains, deployment at scale. Second, what it does not generate: persistent learner profiles, simulation environments, process-based competency evidence. Third, what it structurally cannot provide: data sovereignty (Google's

business model requires data to flow to the platform), federated knowledge governance (knowledge definition is proprietary), and trustee accountability (a publicly traded company has fiduciary duties to shareholders, not learners). A policy promise that "data will not be used for training" is a business decision reversible by a terms-of-service update. An architectural guarantee that the operating organization *cannot access* the data is a different kind of commitment — one that a company whose revenue depends on data aggregation cannot credibly make.

The pieces are on the table. Each works. None works alone, and they cannot be assembled by stacking. The question is what forces their integration into a single architecture.

## Design Principles

Three premises survive the analysis of the preceding sections. Education must be personalized to the individual — not as a luxury but as the baseline. The learner must own their data — not as a policy preference but as an architectural constraint. The system must serve a lifetime. The SDG 4 numbers rule out scaling the existing model. The OECD evidence rules out general-purpose AI without pedagogical architecture. The dependency graph rules out piecemeal assembly.

If you accept these premises, you cannot avoid the conclusions that follow.

What follows are nine requirements that emerge when these premises meet the constraints of the real world. They are organized in three clusters and connected by a dependency graph. The six dimensions identified in Section 3 describe *what* must be integrated. These nine principles describe *what the integrated system must satisfy*. The eight layers described in Section 5 describe *how* the architecture realizes them. Dimensions are the problem structure. Principles are the requirements. Layers are the solution.

The first cluster concerns what the system must do for the learner.

**Autonomous completeness.** The system must deliver a complete academic education autonomously — from first literacy to university-level competence — without requiring a human teacher at any point in the learning path. This is arithmetic: the teacher gap identified in Section 2 grows faster than teachers can be trained. A system that *requires* a teacher to function cannot reach the people who need it most. The claim is specific: the system replaces the instructional and assessment functions of a teacher. It does not claim to replace the socialization, identity formation, and emotional development that human relationships provide. The architecture actively facilitates human connection —

through collaborative learning (L3), community hubs (L7), and teacher integration where teachers exist — but it cannot guarantee these, and it does not pretend they are unnecessary. Where teachers are available, they enrich the experience profoundly. Where they are not, the learner still receives a complete academic education.

**Radical individualization.** The learner steers through intention, not curriculum. The system recognizes where the learner stands, adapts continuously, and generates content in whatever form serves that learner at that moment. This derives directly from the tutoring research: the only way to close the gap at scale is to make individual tutoring the default mode.

**Universal access.** Every human, regardless of location, income, disability, or life situation, must be able to use the system. Access cannot depend on owning personal hardware or having a stable internet connection. A system that excludes the population it most needs to reach has failed before it starts.

The second cluster concerns trust — the guarantees without which the learner-facing promises collapse.

**Competence over certificate.** Abilities are demonstrated through continuous, granular process evidence — not point-in-time examinations. The OECD finding that AI-assisted output quality does not transfer to unassisted performance confirms what the competency-based literature has long argued: if AI can produce any output on demand, only evidence of the thinking process itself remains valid.

**Protection of the learner.** The system must not harm, manipulate, create dependency, or stigmatize. It recognizes its own limits and hands over to humans when it reaches them. A system operating at the intimacy level that radical individualization demands holds dangerous power. Protection is architecturally mandatory from the first design decision.

**Data sovereignty.** Learning data belongs to the learner. No state, no company, no institution may use it for control, evaluation, or surveillance. The separation between educational data and any executive system is absolute. The deep learner profile — capturing how a person thinks — demands the strongest possible guarantee: not "it is forbidden to access the data" but "it is impossible to access the data." This principle is stated here; Section 6 describes the architecture that enforces it.

The third cluster concerns governance — who controls the system and how.

**No monopoly control.** No single actor controls the system. The operating organization is a trustee: responsible for infrastructure, access, and data sovereignty — not for content. The structural incompatibility between trustee

and shareholder governance, identified in Section 3, makes this a design requirement, not a preference.

**Cultural diversity with universal interoperability.** Scientific foundations are universal. Cultural content is sovereignly curated. Both coexist and interoperate. This two-layer model avoids two failure modes simultaneously: a global system that imposes cultural monoculture, and a fragmented system where competencies do not transfer across contexts.

**Education as binding right.** The system is built as if effective education were a human right — not because any treaty currently guarantees it in this form, but because the architecture does not wait for permission. When a person anywhere on earth can demonstrably move from illiteracy to professional competence without institutional gatekeeping, the absence of legal recognition becomes harder to sustain.

These nine principles form a dependency graph. Autonomous completeness requires universal access — a self-sufficient system that only works with broadband is not autonomous. Radical individualization requires a deep learner profile, which requires data sovereignty, which requires no monopoly control. Competence over certificate requires process evidence, which requires both the deep profile and protection of the learner. Cultural diversity requires federated governance, which requires the no-monopoly principle. Every principle connects to at least two others. Remove any one and the system collapses. Each partial solution — an adaptive platform that individualizes but does not guarantee sovereignty, a privacy system that protects data but does not generate learning, a MOOC that provides access but not individualization — satisfies some principles while violating others. This is why piecemeal approaches fail.

What architecture satisfies all nine simultaneously?

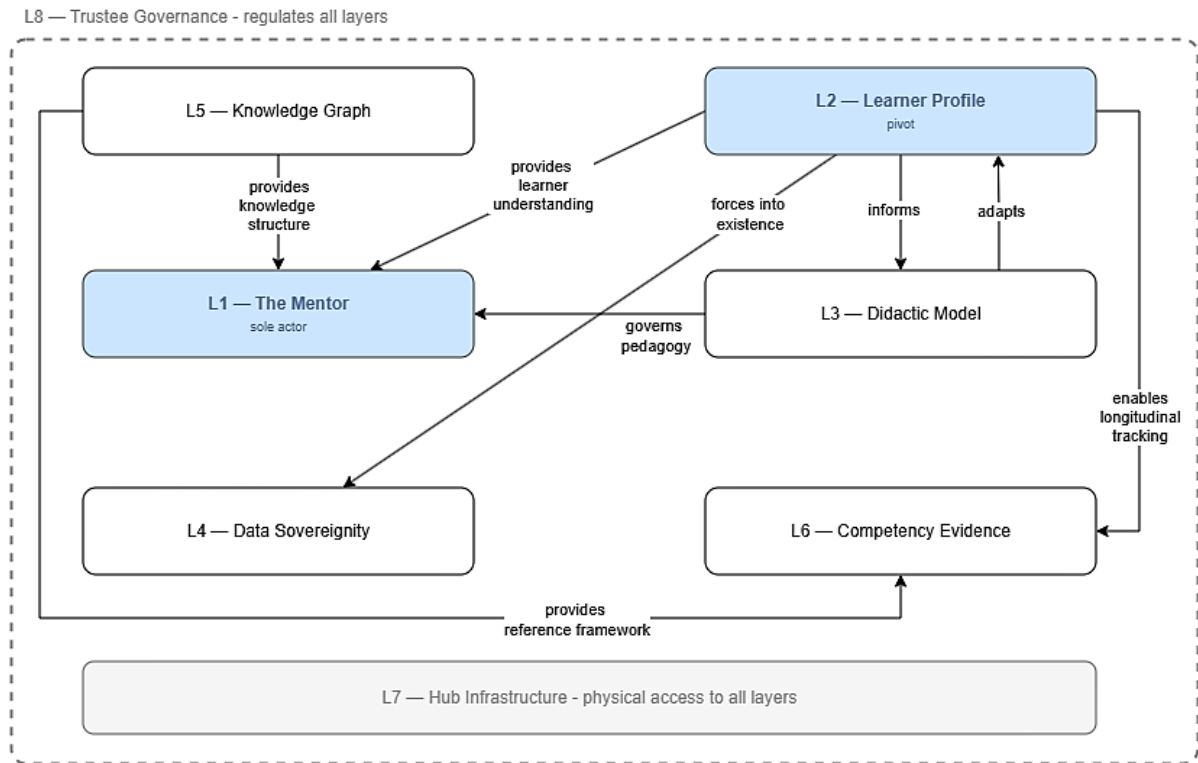
## Architecture

The architecture has eight layers. They form a dependency graph in which each layer requires, feeds, or constrains others, and removing any single layer causes cascading failures.

The topology is clear. The Knowledge Graph (L5) and the Learner Profile (L2) both feed the Mentor (L1) — one provides the structure of what can be learned, the other the understanding of who is learning. The Didactic Model (L3) connects L1 and L2: it defines the principles by which the Mentor translates knowledge of the learner into a learning experience. The depth of L2 forces Data Sovereignty (L4) into existence: without architectural privacy guarantees, the profile becomes a surveillance instrument. L2 and L5 together enable the

Competency Evidence System (L6): evidence requires both a reference framework and longitudinal observation. The Hub Infrastructure (L7) physically carries everything. Trustee Governance (L8) regulates everything.

L2 is the pivot. The deep profile is the prerequisite for individualization (L1), didactic adaptation (L3), and competency evidence (L6) — and the reason data sovereignty (L4) is non-negotiable. L1 is the only actor; everything else is a resource, a principle, or a constraint.



**Figure 2.** Dependency Graph of the Eight Architectural Layers

**L1 — The Mentor.** The system has one actor: the Mentor. Everything else serves it. The Mentor is not a chatbot that answers questions. It is a counterpart that knows the learner and accompanies them — through explanation and silence, encouragement and provocation, patience and demand.

What the Mentor generates is the learning experience itself — content, tone, timing, provocation, context, the sense of being accompanied. When a learner says "I want to understand why my bridge collapsed," no material exists that answers this question for this person in this state of mind. The Mentor creates it.

Three tiers of generation operate at different timescales. The first is real-time dialogue: Socratic exchange, tasks, feedback, generated in the moment. The second is simulation environments: a specialized Forge-AI generates interactive physics labs, chemistry environments, and engineering scenarios

asynchronously and places them in a shared repository. Google DeepMind's Genie 3 (2025) has demonstrated that generating interactive 3D environments with real-time physics is technically feasible — a user moves through a generated world and the physics responds dynamically.<sup>21</sup> What Genie 3 does not yet guarantee is the physical precision that education demands: a lever arm that is five percent off suffices for a game but not for physics instruction. The architecture addresses this by combining tiers: Tier-2 Forge-AI generates the environment and its pedagogical framing; Tier-3 deterministic engines (physics engines, chemistry engines) compute the actual natural laws within that environment. The generation is AI; the physics is exact. Validating the pedagogical soundness of generated environments — ensuring they teach what they claim to teach — is an engineering challenge, not a research problem, and is handled through automated test suites against the Knowledge Graph's dependency structure. The third tier is these long-lived deterministic engines themselves — computing natural laws from lightweight instruction sets. The distinction matters: this architecture generates environments *in which* one experiences physics, not explanations *about* it. The child changes the lever arm, increases the load, sees the forces respond in real time.

The Mentor is an ensemble of six specialized agents operating on different time horizons. A Content-Agent generates in seconds. A Didactics-Agent adjusts difficulty over minutes to hours. A Navigation-Agent tracks progress over weeks to months. A Perception-Agent reads behavioral signals in real time. A Protection-Agent acts as a circuit breaker when frustration crosses from productive to destructive. An Audit-Agent monitors the Mentor itself for manipulative rhetoric, discriminatory bias, or ideological drift. Not everything in this ensemble is a large language model — perception and protection are fast classifiers; navigation can be classical graph algorithms. The expensive foundation model is reserved for the expensive task: generating the nuanced, contextual learning experience.

The ensemble coordinates through a shared session state and a priority hierarchy: Protection overrides Didactics, which overrides Content. The Audit-Agent is architecturally independent — it runs in a separate execution context, cannot be overridden or disabled by the system operator, and its logs are append-only and tamper-evident. If coordination mechanisms prove insufficient at scale, the specific failure modes are defined problems for iterative engineering, not architectural unknowns.

The Mentor works from the first interaction. No onboarding gate, no questionnaire. The learner says "I want to learn to read" and it starts — broad at first, narrowing with every interaction. Like any new relationship.

---

<sup>21</sup>Google DeepMind (2025). Genie 3: Generating Interactive 3D Environments. Technical Report.

**L2 — The Learner Profile.** Without L2, the Mentor generates generic experiences. With L2, it generates *this* experience for *this* human.

The profile is organized as a compression pyramid. At the base: raw data — clicks, hesitations, typing speed, answer patterns — captured continuously. One level up: real-time interpretation — fatigue, frustration, attention, inferred in seconds. Next: daily synthesis — thousands of micro-cognitive signals compressed into a sharpened understanding. At the top: the long-term profile — learning type, cognitive strengths, persistent blockades, deep interests — synthesized over months and years.

Separate from the pyramid, the system maintains an episodic memory — curated key moments the Mentor recognizes as significant. "November 12th — cracked vector calculus." "The crisis in March and how it was overcome." A good mentor remembers.

The profile has no expiration date. It accompanies the person from childhood literacy through adult reskilling through retirement language learning. Within it lives the personal competency graph — a projection onto the Knowledge Graph (L5). In the old system, a student has one position: grade 7, GPA 2.3. In this system, the learner has thousands of positions in a multi-dimensional space — algebra but not fractions, Python but not the logic underneath, physics concepts but not formula rearrangement. Granular, not linear.

From this, the system derives a living effort prognosis. It knows the learner (L2), the graph (L5), and the aggregated experience of millions of trajectories. "Route A takes an estimated 24 months. Route B uses your strength in visual-spatial thinking and takes 14." The estimate updates with every step.

**L3 — The Didactic Model.** L3 is the score; L1 is the performance. L3 defines the principles by which learning happens. L1 executes them in the moment.

Seven principles govern the model. Intent-based navigation: the process begins with an intention, not a curriculum. The Socratic method: the Mentor asks back rather than answering. Dynamic difficulty: maintaining the learner in the zone of proximal development. Functional embedding: fundamentals are woven into the learning path at the moment they become necessary. Content polymorphism: the musician learns fractions as rhythm intervals, the architect as geometric ratios — the logical understanding is non-negotiable, but the form adapts. Progressive de-adaptation: with growing competence, the system deliberately and systematically reduces individualization, building the learner's capacity to function without it. Collaborative learning: the system connects learners — co-located at a hub, synchronously across distance with real-time translation, or asynchronously across time — matched by complementary strengths, shared interests, and compatible levels.

Progressive de-adaptation deserves elaboration because it addresses the strongest pedagogical objection to the entire architecture: that a system this responsive creates dependency. The objection is valid — a learner who has never navigated a poorly structured resource, endured ambiguity, or decided what to study next has never developed the self-regulation the real world demands. De-adaptation is the architectural answer. It begins at intermediate competency levels and escalates through defined stages: first, the Mentor reduces content polymorphism (the musician must see fractions as geometry, not just rhythm). Then it reduces scaffolding (hints become less specific, wait times before intervention grow longer). Then it introduces deliberate friction (ambiguous problems, resources with gaps, tasks requiring the learner to identify what they need to learn). Finally, it withdraws the adaptive profile entirely for defined periods — the learner works with generic materials and self-directs. The system measures de-adaptation readiness through a specific signal: can the learner achieve comparable outcomes with reduced support? If performance collapses, the system re-engages. The goal is not a learner who needs the Mentor forever. It is a learner who outgrows it.

**L4 — Data Sovereignty.** The depth of L2 — a profile that captures how a person thinks, not just what they answered — makes L4 non-negotiable. Without architectural privacy guarantees, the deep profile becomes the most dangerous surveillance instrument ever built for education. Section 6 describes the full sovereignty architecture: stateless processing, learner-held encryption keys, custodial key management for children, and the threat model that makes violation not merely forbidden but physically impossible.

**L5 — The Federated Knowledge Graph.** A knowledge graph is a structured map of concepts and their relationships — what depends on what, what enables what. It is the system's understanding of the structure of knowledge itself.

The graph models dependencies, not sequences. "Fractions require basic arithmetic" is a dependency — it belongs in L5. "First fractions, then decimals" is a didactic decision — it belongs in L3. This distinction is essential: if the graph prescribed sequences, it would become the curriculum.

The graph has a two-layer architecture. Layer 1 is the universal STEM core — globally shared, versioned, maintained by independent rotating expert panels. Every change is recorded in an open ledger. Layer 2 is the sovereignly curated cultural layer — history, society, values, ethics — maintained by legitimated curatorial authorities. A national education ministry opens the governance cockpit, sees the universal graph, marks where its narrative diverges, and sees the consequences. The system stores both perspectives as attributes, not errors.



When Layer 1 and Layer 2 curators disagree — for instance, when a cultural authority contests a dependency that the STEM panel considers universal — the conflict resolution principle is explicit: Layer 1 governs empirically falsifiable claims (the boiling point of water, the prerequisites of calculus); Layer 2 governs interpretation, narrative, and values. Boundary disputes — where a claim is contested as to whether it is empirical or interpretive — are escalated to a standing boundary committee with rotating membership drawn from both panels, whose rulings are published with reasoning and are revisable.

Within these two governance layers, the graph operates with richer internal structure. A concept layer maps abstract knowledge and dependencies. A manifestation layer maps the same concept across domains — "ratios" as fractions in mathematics, rhythm intervals in music, recipe scaling in cooking — enabling the Mentor's polymorphic delivery. A competency layer connects to L6: what one must be able to do. A cross-cutting competency layer maps skills that grow through everything — creativity, critical thinking, frustration tolerance — observed bidirectionally across learning situations. These internal layers were not foreshadowed in the Related Work inventory because they do not exist in any current system. They are part of the original architectural contribution.

The graph is alive. Initial construction is AI-assisted — automated methods extract prerequisite relationships with usable precision (F1 scores of 70–90% depending on domain),<sup>22</sup> providing a starting point for human validation. Continuous refinement happens through expert review, implicit validation through millions of learning trajectories, and ingestion of new findings. The system discovers relationship types never pre-modeled: after millions of trajectories, it might find that mastering concept X systematically blocks concept Y — not a "requires" relationship but an "interferes with" edge that no one anticipated.

Adoption follows a specific dynamic. The graph becomes irreversible when it is referenced — by Mentors generating experiences, by evidence anchored to its nodes, by universities recognizing its structure, by employers using its competency definitions. De-facto standardization through utility, not regulation. But this dynamic has a cold-start problem: the graph needs users to become valuable, and users need the graph to be valuable. The natural first vertical is mathematics from primary school through university-level calculus — a domain with clear dependencies, minimal cultural contestation, and existing prerequisite research to validate against. The cold-start problem is the single hardest practical challenge in the architecture and receives full treatment in Section 8.

---

<sup>22</sup>Liang, C., et al. (2018). Investigating active learning for concept prerequisite learning. Proceedings of AAAI.

**L6 — The Competency Evidence System.** Competence is not proven by separate exams. It emerges from the accompanied learning process. The Mentor witnesses the cognitive journey, and from this continuous observation, evidence accumulates.

Three dimensions define the evidence. Autonomy: did the learner solve it independently, or with heavy guidance? Transfer: can the learner apply the concept in a new context — fractions learned through recipes used in an architecture problem? Reflection: can the learner explain *why* something works, not just *that* it works?

This evidence is inherently fraud-resistant against unassisted cheating. A traditional exam requires the right answer at the right moment; cheating is a point event. Process evidence over months would require consistently faking every Socratic dialogue, simulating transfer across contexts, and maintaining a coherent profile over hundreds of interactions. The threat model must also account for AI-assisted fraud: a learner using a second AI to feed answers into the Mentor's dialogues could maintain a coherent fake profile without the sudden capability jumps or stylistic inconsistencies the Audit-Agent checks for. The architecture addresses this through layered defenses: behavioral biometrics (typing dynamics, hesitation patterns, response latency distributions) that are harder to simulate than content; periodic challenge probes — unexpected, timed tasks that require real-time reasoning with latency constraints incompatible with an intermediary; and the progressive de-adaptation mechanism in L3, which strips away scaffolding and forces unassisted performance at critical thresholds. No single defense is foolproof. The combination raises the cost of sustained fraud to a level where maintaining the deception over months requires more cognitive effort than learning the material. If a future AI can perfectly simulate months of Socratic dialogue including hesitation, confusion, gradual insight, and authentic transfer — the distinction between simulating learning and actually learning becomes philosophically interesting rather than practically concerning.

The architecture enforces separation of powers. The Mentor generates evidence through accompaniment — it is the teacher, not the examiner. An independent Audit-Agent — performing the function of a notary — verifies whether the performance is genuine: does it fit the profile? Are there sudden capability jumps? Stylistic inconsistencies? This is the same Audit-Agent introduced in L1, operating in a different mode: in L1 it monitors the Mentor for bias and drift; in L6 it validates the evidence the Mentor produces. Systemic calibration operates across the entire system through statistical anomaly detection.

Evidence gains weight through graph anchoring. Fractions used successfully in algebra, then in integral calculus, then in physics problems, are implicitly confirmed hundreds of times. The higher floors validate the lower ones.

**L7 — Autonomous Hub Infrastructure.** Containerized data centers are an established market — estimated at \$29 billion in 2025.<sup>23</sup> Edge AI has reached a threshold: quantized 70-billion-parameter models run on a single high-end GPU; models under 9 billion parameters run on smartphones.<sup>24</sup> Satellite internet provides connectivity practically everywhere.<sup>25</sup>

The hub is a standardized, transportable core module: GPU server, graph database, encrypted storage, simulation engine, satellite uplink, solar power, robust tablets. Deliverable by truck, ship, or helicopter.

Three operating modes define resilience. Cloud mode: full connectivity, heavy tasks run remotely. Hybrid mode: cloud handles heavy tasks, the hub runs core operations locally. Island mode: degraded but functional — local models, local graph, local profiles, no synchronization until connectivity returns. Island mode is the safety net, not the norm. But when the internet goes dark, the hub continues.

Three hub types serve different scales. A Basis-Hub serves dozens of learners. A Standard-Hub serves hundreds. A Regional-Hub serves thousands and adds physical laboratories and workshops.

One technical fact deserves precision. The architecture's simulation tier — deterministic engines computing physics from lightweight instruction sets — requires less bandwidth than video streaming. A projectile-motion simulation needs kilobytes of code transmitted to a local engine; a video lecture on the same topic streams hundreds of megabytes. This advantage applies specifically to tier-3 deterministic simulations. The Mentor's real-time conversational AI (tier 1) requires cloud connectivity for full capability, which is why hybrid mode is the expected norm and island mode runs on smaller local models with reduced capacity. The bandwidth advantage is real but specific — it means the richest form of interactive simulation works precisely where infrastructure is weakest.

**L8 — Trustee Governance.** A Global Education Alliance (GEA) operates as trustee of the infrastructure — analogous to CERN or the internet's governance bodies. The GEA is the guardian of the cable, not the author of the message.

---

<sup>23</sup>MarketsandMarkets (2025). Modular Data Center Market — Global Forecast.

<sup>24</sup>Dettmers, T., et al. (2022). LLM.int8(): 8-bit matrix multiplication for transformers at scale. NeurIPS 2022.

<sup>25</sup>SpaceX (2024). Starlink Specifications and Coverage Map.

The GEA is responsible for infrastructure operation, data sovereignty enforcement, access guarantees, and the universal STEM core. It is *not* responsible for cultural content, didactic decisions, or learner evaluation. Whoever operates the infrastructure must not control the knowledge.

The GEA emerges gradually. Phase one: a foundation — a private educational offering, funded philanthropically, requiring no political legitimacy. Phase two: a consortium — first states join because results convince, governance becomes formal. Phase three: an international organization with statutes and multipolar stewardship.

No gatekeeper stands between the learner and the system. Access cannot be used as leverage — enforced through architecture, not policy. And if the GEA fails, decentralized hubs continue operating. The governance steers the standard, not the daily operation.

These eight layers are not a menu. Remove L4 and the deep profile becomes a surveillance tool. Remove L5 and the Mentor generates without epistemic structure. Remove L8 and whoever operates the infrastructure controls the knowledge. Remove L7 and 272 million children without school remain unreachable. Remove L6 and the system teaches but cannot prove what was learned.

The architecture is the integration that emerges when the nine principles are taken to their logical conclusion. Of all these layers, one is most consistently underestimated — not as a feature, but as the principle that makes everything else trustworthy.

## Data Sovereignty as Architectural Principle

The preceding sections established *why* data sovereignty is required and *what* it means in principle. This section describes *how* the architecture enforces it — the specific mechanisms that make violation not merely forbidden but physically impossible.

The principle that individuals should control their own data has moved from advocacy to implementation. Data cooperatives, self-sovereign identity frameworks, and the European Data Governance Act converge on the same structural insight: personal data must be governed by the person it describes, not by the organization that processes it.<sup>26</sup> In healthcare, cooperative governance models have demonstrated that citizens can control sensitive records through democratic structures — encrypted so that neither

---

<sup>26</sup>Hafen, E. (2019). Personal data cooperatives — a new data governance framework. In *The Ethics of Medical Data Donation*, Springer. See also European Parliament (2022), Regulation on European Data Governance (Data Governance Act), and the MyData Declaration (2017).

administrators nor management can access them — while preserving the data's utility for research. This paper applies the same structural logic to education, where the data is arguably more intimate: a health record captures what happened to your body; a lifelong learning profile captures how your mind works.

Apple's Private Cloud Compute, launched in 2024, demonstrated the core mechanism at global scale: stateless computation (data exists only in volatile memory during processing), enforceable guarantees (only cryptographically signed code runs), no privileged access (no employee can access user data), non-targetability (requests cannot be routed to specific servers), and verifiable transparency (the codebase is publicly auditable).

Apple can be stateless — process and forget. A lifelong learning system cannot. The profile must persist over years and follow the learner between hubs, cities, countries. The solution separates two concerns Apple could collapse into one: stateless processing combined with stateful encrypted storage under the learner's control.

During a learning session, the Mentor operates on decrypted profile data in volatile memory — the same ephemeral processing Apple demonstrated. When the session ends, the data is re-encrypted. The encryption key belongs to the learner, managed through Self-Sovereign Identity. Without the key, the data is mathematically unreadable. The hub is an access point, not a storage location. If the operating organization is compromised, acquired, or pressured by a government, what it holds is ciphertext.

Key management for a lifelong system that begins in early childhood is a serious architectural challenge, not a solved problem. A six-year-old cannot manage cryptographic keys. The architecture addresses this through a custodial model with graduated transfer: parents hold the child's keys initially, managed through a threshold cryptography scheme where multiple guardians (parents, the hub, a recovery trustee) each hold key shares — no single party can access the data alone, but a defined quorum can recover access if a key is lost. As the learner matures, key shares transfer progressively: the adolescent gains independent shares that create private zones parents cannot access, while parents retain emergency recovery shares until full legal majority. Key loss triggers a recovery protocol — not data destruction — using the threshold scheme. Social recovery (designating trusted contacts who together can reconstruct access) provides an additional safety net. These mechanisms draw on established building blocks — Shamir's Secret Sharing, social recovery wallets, custodial bridges — but their application to a lifelong educational profile with graduated autonomy is novel and requires careful implementation.

No single actor sees the complete picture. The Mentor sees the learner's profile deeply but cannot issue credentials. The Audit-Agent verifies evidence but sees only the evidence stream and statistical signatures, not the full profile. The governance layer sees system health but not individual data. No human at any level of the operating organization sees raw data or psychometric profiles.

The learner sees an interpreted cockpit: progress, strengths, trajectories. Parents see what the child shares — milestones, areas needing attention — never emotional patterns or psychometric details. Teachers receive discrete action impulses: "have a conversation," "this learner needs encouragement." Never diagnoses. Employers see only verified competency evidence for exactly the skills requested. Nothing about the process, the profile, or the person beyond what they chose to share.

A state that demands access gets nothing — not because a contract forbids it, but because no interface exists. The separation between educational data and executive systems is architectural: no API endpoint connects them. Seizing a hub yields encrypted storage; the keys are with the learners.

**Threat Model.** The sovereignty claims above require a concrete threat model. The following table maps specific threats to architectural mitigations:

#	Threat	Mitigation
1	Operating organization compromised or acquired	Hub stores only ciphertext; keys are with learners. Acquiring the organization yields no usable data.
2	State demands bulk access to learner data	No API endpoint connects educational data to external systems. No technical interface exists to comply, even under legal compulsion.
3	State bans the system entirely	See deployment model below. Hubs are portable, operable by NGOs, and functional in island mode without central coordination. The architecture is designed to operate as a supplement within existing state education systems (minimally threatening) or as an independent offering by non-state actors where necessary.
4	Insider at hub level attempts data exfiltration	Stateless processing: data exists decrypted only in volatile memory during sessions. No persistent plaintext on disk. Hardware attestation ensures only signed code runs.
5	Rogue Mentor instance profiles learners for external use	Audit-Agent monitors all Mentor outputs in a separate execution context. Mentor has no network egress path except through the encrypted storage and evidence pipelines.
6	Side-channel attack on volatile memory during session	Hardware-level mitigation: trusted execution environments (TEEs) with memory encryption, as demonstrated in Apple's Secure Enclave and AMD SEV. Residual risk exists and is bounded by session duration.

7	Key loss by learner (accidental)	Threshold recovery via guardian quorum (see key management above). Data is recoverable; no child loses their educational history due to a lost device.
8	Parent coercion — accessing adolescent's private zones	Threshold scheme requires the adolescent's own key share for private zones. Parent shares alone are insufficient.
9	Re-identification from anonymized research data	Research AI operates in secure enclaves. Differential privacy with conservative epsilon bounds. No human sees aggregated data. Publication threshold enforced inside the enclave.
10	Coordinated attack: state + insider + physical seizure	Defense in depth: even with physical hub access and a compromised insider, data remains encrypted with keys distributed across learner-held shares. The attack must also compromise the learner's personal key shares — at which point the threat model is equivalent to physical coercion of the individual, which no technical system can fully prevent.

This threat model does not claim invulnerability. It claims that the architecture reduces the attack surface to the individual learner's key custody — a fundamentally different security posture than any system where a central operator holds the keys.

The political threat — a state that bans the system rather than trying to break it — deserves direct attention. The states where Lost Einsteins are most concentrated are often the states most likely to reject an education system they cannot inspect or control. The architecture addresses this through deployment flexibility, not political confrontation. In cooperative states, the system operates as a supplement within existing education systems — the least threatening integration model. In hostile environments, the hub design enables operation by NGOs, religious organizations, or community groups, with island mode ensuring functionality without central coordination. The hubs are portable, deniable in their specificity (a shipping container with solar panels and tablets), and operationally independent. This does not solve the political problem — no architecture can force a state to permit education it opposes — but it ensures that the technical architecture does not *add* political dependencies beyond those inherent in any educational intervention.

Data sovereignty grows with the learner. For young children, parents co-determine sharing through the custodial key model described above. As the learner matures, control shifts — the adolescent defines private zones parents cannot see, enforced by the threshold scheme. In adulthood, sovereignty is complete. The kill-switch — the right to be forgotten — is an accompanied process: immediate functional lock, mandatory human conversation, waiting period, then irreversible cryptographic destruction if the learner still chooses it.

The same architecture that protects individuals enables learning research at unprecedented scale. Millions of Mentors accompany millions of learners. Anonymized, aggregated data flows into a research layer: which explanations work for which cognitive patterns? Where do learners systematically fail? Which paths through the Knowledge Graph are efficient? This is learning research at a scale that has never existed — not 100 subjects in a study but millions of real learning trajectories. The research AI operates in secure enclaves under the same guarantees as the individual Mentor. No human sees the aggregated data. Insights flow back as abstract patterns — "learners with pattern X benefit from approach Y" — never as individual profiles.

Same architecture, same protection, two levels: individual (full profile, only this learner's Mentor) and collective (anonymized, only the research AI). The system does not just get better at teaching. It gets better at understanding what knowledge is and how humans acquire it.

Sovereignty is not the enemy of collective learning. It is the precondition. The architecture is trustworthy. Can it be financed at global scale?

## **Economics of Generative Education**

Excellent education for every human has been utopian for a simple reason: costs grew linearly with enrollment. Every additional student required an additional teacher, an additional classroom, an additional hour. A lecture delivered in São Paulo did nothing for students in Nairobi. The entire economic history of education is a history of duplication costs. And once the hour was over, the money was spent — education was consumptive.

This architecture breaks that logic. An improved model, a new simulation, a refinement of the Knowledge Graph — all are available to every learner afterward. The marginal cost for an additional learner approaches zero. Whether the system accompanies one learner in Zurich or a billion learners in Lagos costs almost the same on the software side. Every expenditure is investment: it accumulates instead of evaporating. The capital serves permanent asset building, not maintenance of the status quo. This is the structural difference between Baumol's cost disease — the observation that human labor in service sectors does not become more productive over time, so education gets perpetually more expensive<sup>27</sup> — and software economics, where code scales at near-zero marginal cost. The SDG 4 arithmetic established in Section 2 is the cost disease made visible. This architecture escapes it.

---

<sup>27</sup>Baumol, W. J. (1967). Macroeconomics of unbalanced growth: The anatomy of urban crisis. *American Economic Review*, 57(3), 415–426.



The economic case rests on the largest misallocation of resources in human history.

Bell, Chetty, Jaravel, Petkova, and Van Reenen tracked 1.2 million inventors in the United States.<sup>28</sup> Their core finding: if women, minorities, and children from low-income families invented at the same rate as white men from high-income families, the rate of innovation would quadruple. The mechanism is not ability — it is exposure. Children who excelled in third-grade mathematics but came from the top twenty percent of the income distribution were five times more likely to become inventors than equally talented children from the bottom eighty percent. Hsieh and colleagues estimated that improved talent allocation alone explained twenty to forty percent of aggregate U.S. output growth between 1960 and 2010.<sup>29</sup>

They called these missing innovators "Lost Einsteins." The research is specific to the United States — it tracks patent rates in a particular economy with particular institutions. Extrapolating to a child born in rural Bangladesh requires acknowledging that the barriers there extend beyond education: nutrition, political stability, market access, intellectual property infrastructure, and capital markets all mediate whether talent converts to innovation. Education is necessary but not sufficient. What the Lost Einsteins research establishes is the *scale* of the waste: even in the world's richest economy with the world's best universities, the majority of potential innovators never reach the starting line. In contexts with fewer institutional supports, the waste is necessarily greater, even if the precise multiplier is unknown. An education architecture that provides equal exposure to innovation pathways addresses the largest single barrier — but it does not claim to address all of them.

The populations with the greatest concentration of Lost Einsteins are those with the weakest infrastructure. The bandwidth characteristics described in Section 5 address this directly: the architecture's simulation tier works where infrastructure is weakest, because deterministic simulations computed from lightweight instruction sets require less bandwidth than video streaming. A village with a solar-powered hub and a satellite uplink can run physics simulations that a video platform cannot deliver there.

The economics contain a self-reinforcing dynamic. When Mentors reference the Knowledge Graph, when evidence is anchored to its nodes, when universities recognize its structure, when employers use its competency definitions — a de-facto standard emerges through adoption, not regulation. The network effect

---

<sup>28</sup>Bell, A., et al. (2019). Who becomes an inventor in America? The importance of exposure to innovation. *Quarterly Journal of Economics*, 134(2), 647-713.

<sup>29</sup>Hsieh, C.-T., et al. (2019). The allocation of talent and U.S. economic growth. *Econometrica*, 87(5), 1439-1474.

compounds: each university that maps its curriculum to the graph makes it more useful for every employer; each employer that accepts graph-based evidence makes it more valuable for every learner. The same dynamic that made GSM or TCP/IP irreversible. The more actors use the graph, the more valuable it becomes for all — and the harder it becomes to ignore.

The system creates value that should, over time, generate the revenue to sustain it: millions of newly skilled workers generate tax revenue, economic growth, and solutions to problems whose costs currently run into the trillions. A full cost-benefit analysis — hub unit costs, operating expenses, projected economic returns over defined timelines — requires empirical data from pilot deployments and is beyond the scope of this architecture document. What can be stated with confidence is the direction: the downstream costs of global educational failure (estimated by the World Bank at \$21 trillion in lost lifetime earnings for the current generation)<sup>30</sup> dwarf plausible infrastructure investments by orders of magnitude. The economic case is directional and overwhelming, even if the precise return timeline remains to be established through implementation.

The question is not whether the world can afford this system. The question is whether it can afford not to build it. The economics work. But intellectual honesty demands naming what remains unsolved.

## Open Questions and Research Agenda

An architecture document that claims everything is solved is not credible. Every component technology exists in production. But integration at this scale has never been attempted, and several challenges remain genuinely open. None is a fundamental blocker — but none is trivial. Each is an invitation to the research community: here is where your work is needed.

**Privacy-Preserving Credential Verification.** The architecture claims competency evidence is verifiable without exposing raw data. W3C Verifiable Credentials and zero-knowledge proofs provide the technical foundation. A minimal viable credential system works as follows: the Audit-Agent (L6) signs a competency attestation anchored to specific Knowledge Graph nodes, including the evidence dimensions (autonomy, transfer, reflection) as scored attributes. The learner holds this attestation in their SSI wallet as a W3C Verifiable Credential. When a university or employer requests proof, the learner uses selective disclosure — revealing, for example, that they demonstrated calculus competency with high autonomy and transfer scores, without revealing the underlying Socratic dialogues, emotional patterns, or learning trajectory. Zero-

---

<sup>30</sup>World Bank, UNESCO, UNICEF (2022). The State of Global Learning Poverty: 2022 Update.

knowledge proofs can further attest to threshold claims ("this learner's autonomy score exceeds X") without revealing the score itself. The gap that remains is specific: every existing self-sovereign identity system focuses on point-in-time credentials. No system has implemented continuous process-based evidence with privacy preservation. Translating a living, accumulating record of how someone thinks into a verifiable credential that proves competence without revealing process is a hard engineering challenge. Without it, the evidence system cannot interface with universities, employers, or professional bodies. *Criticality: high. Path: clear. Implementation: not yet done.*

**Cultural Curation in Layer 2.** The Knowledge Graph's universal STEM core launches independently. Layer 2 — history, society, values, ethics — requires legitimated curatorial authorities, and no precedent exists for this governance model at global scale. The political negotiation over who counts as a legitimated authority, how mandates are granted, and how disputes at the Layer 1/Layer 2 boundary are resolved will be difficult. The architecture supports this process rather than depending on it being resolved before anything begins. *Criticality: high for global deployment, zero for initial launch.*

**Knowledge Graph Cold Start.** The graph becomes irreversible once widely referenced, but crossing the initial adoption threshold is arguably the single hardest practical problem in the entire architecture. The bootstrap strategy outlined in Section 5 — a funded expert team building the K-12 mathematics graph as the first vertical — defines the minimum viable artifact. The critical unknowns are: what funding model sustains multi-year expert labor before network effects engage? What is the minimum graph density at which a Mentor can generate a useful learning experience? How many institutional adopters constitute the tipping point? Khan Academy's dependency graph demonstrated both the value and the fragility of this approach — it was discontinued not because the pattern failed but because it was embedded in a monolithic platform rather than federated as shared infrastructure. The structural difference — an open, federated graph that multiple systems can build on — is the hypothesis. Validating it requires a funded pilot with a defined success criterion: a cohort of learners completing K-12 mathematics using graph-guided Mentors, with competency evidence that external evaluators accept. *Criticality: high. This must be the first engineering milestone.*

**Governance Bootstrapping.** The GEA needs legitimacy to operate at scale, but legitimacy comes from demonstrated results. CERN, ICANN, and ESA each solved this same chicken-and-egg problem: start small, demonstrate value, earn the mandate for the next phase. Phase one requires no political legitimacy — it is a private educational offering. Results create the political case for phase two. *Criticality: medium. Does not block technology development; blocks scaling.*

**Offline AI Capability.** Island mode requires the Mentor to function without connectivity. Current edge AI makes this partially feasible for core tasks — reasoning support, conversational tutoring, basic assessment. Complex multi-step reasoning and novel simulation generation may exceed edge hardware. The architecture is designed to grow with the technology: what requires a data center today runs on a phone in five years. *Criticality: medium.*

**Effort Prognosis Validation.** Living effort estimates require validated models for learning time prediction. For K-12 academic subjects, the evidence base is thin. The resolution path is built into the architecture: every learner who follows a path provides calibration data. Initial estimates can be conservative and transparent about uncertainty. *Criticality: medium.*

**The Learning Research Loop.** Anonymized aggregated learning data at this scale is unprecedented. The risk of re-identification from rich longitudinal educational data is non-trivial — a learner's cognitive trajectory over years may be as distinctive as a fingerprint. The architecture's answer is structural: the research AI operates in secure enclaves, no human sees aggregated data, and the publication threshold is set conservatively inside the enclave. Federated learning and differential privacy are mature in principle; their application to educational data at this scale is unexplored. *Criticality: medium-high.*

None of these challenges threatens the architecture's foundations. Credential verification, cultural curation, and the knowledge graph cold start are the most critical — all have clear paths. Offline capability, effort prediction, and the research loop improve with scale and time. The honest assessment: this system can be built and deployed with today's technology in a useful form. It will become excellent as these questions are resolved. Each is a defined problem with a plausible approach and a clear criterion for success — the kind of problem research communities are built to solve.

## Conclusion

The technology converges. Generative AI, privacy-preserving computation, federated data systems, edge computing, satellite connectivity — each developed independently, each maturing on its own trajectory. The architecture emerges from the requirements: nine principles that are logically interdependent and that force, through their dependency graph, an integrated design. The pieces were not designed to fit together, but they do, because the problem has a shape, and that shape constrains the solution. This system will come, in some form, because the forces driving it are structural, not political.

Two paths lead to AI-powered education at planetary scale. One builds it as trustee infrastructure: the data belongs to the learner, democratic bodies curate

knowledge, the operating organization cannot access individual profiles, and no single actor controls the system. The other builds it as a platform product: the data belongs to the company, knowledge definition is proprietary, governance answers to shareholders, and the terms of service can be rewritten with a quarterly earnings call.

Both paths produce personalized learning. Both scale. They lead to fundamentally different futures. In one, a fifteen-year-old's cognitive struggles are protected by architecture — no API endpoint connects her learning data to any external system. In the other, that same data is an asset on a balance sheet, protected by a policy promise that lasts until the next acquisition. In one, the question of what counts as valid knowledge is answered by legitimated communities through transparent governance. In the other, it is answered by a model optimizing for engagement.

One path produces citizens who own their learning. The other produces users who rent it.

The architecture is described. The principles are stated. The open questions are named. In 1984, Bloom posed the challenge of making individual tutoring available to all. Forty years later, the architecture exists. The question is no longer how. The question is who builds it — and for whom.

## **Acknowledgments**

The author thanks Ernst Hafen (ETH Zürich, MIDATA.coop) for discussions on data sovereignty and cooperative governance that informed the approach to learner data protection described in this paper.

## References

- Apple (2024). Private Cloud Compute: A new frontier for AI privacy in the cloud. Apple Security Research.
- Baumol, W. J. (1967). Macroeconomics of unbalanced growth: The anatomy of urban crisis. *American Economic Review*, 57(3), 415–426.
- Bloom, B. S. (1984). The 2 sigma problem: The search for methods of group instruction as effective as one-to-one tutoring. *Educational Researcher*, 13(6), 4–16. <https://doi.org/10.3102/0013189X013006004>.
- Bell, A., Chetty, R., Jaravel, X., Petkova, N., & Van Reenen, J. (2019). Who becomes an inventor in America? The importance of exposure to innovation. *Quarterly Journal of Economics*, 134(2), 647–713.
- Cherner, T. & Mitchell, C. (2021). Deconstructing EdTech Frameworks Based on their Creators, Features, and Usefulness. *Learning, Media and Technology*, 46(1), 91–116. <https://doi.org/10.1080/17439884.2020.1773852>
- Dettmers, T., Lewis, M., Belkada, Y., & Zettlemoyer, L. (2022). LLM.int8(): 8-bit matrix multiplication for transformers at scale. *NeurIPS 2022*.
- European Parliament (2022). Regulation (EU) 2022/868 on European Data Governance (Data Governance Act).
- Google (2025). LearnLM: AI for Learning. Official product page.
- Google DeepMind (2025). Genie 3: Generating Interactive 3D Environments. Technical Report.
- Google DeepMind (2025). LearnLM Tutoring Study: UK Controlled Trial Results.
- Hafen, E., Kossmann, D., & Brand, A. (2014). Health data cooperatives — Citizen empowerment. *Methods of Information in Medicine*, 53(2), 82–86. <https://doi.org/10.3414/ME13-02-0051>.
- Hafen, E. (2019). Personal data cooperatives — a new data governance framework. In *The Ethics of Medical Data Donation*, Springer.
- Harden, R. M. (1988). What is an OSCE? *Medical Teacher*, 10(1), 19–22.
- Hariyanto, Kristianingsih, F.X.D. & Maharani, R. (2025). Artificial intelligence in adaptive education: A systematic review of techniques for personalized learning. *Discover Education*, 4, 458. <https://doi.org/10.1007/s44217-025-00908-6>
- Hsieh, C.-T., Hurst, E., Jones, C. I., & Klenow, P. J. (2019). The allocation of talent and U.S. economic growth. *Econometrica*, 87(5), 1439–1474.
- Jordan, K. (2015). Massive open online course completion rates revisited. *International Review of Research in Open and Distributed Learning*, 16(3).
- Khan Academy (2024). Khanmigo Year in Review.
- Liang, C., Ye, J., Wang, S., Pursel, B., & Giles, C. L. (2018). Investigating active learning for concept prerequisite learning. *Proceedings of AAAI*.
- MarketsandMarkets (2025). Modular Data Center Market — Global Forecast.
- MyData Global (2017). MyData Declaration.
- OECD (2023). PISA 2022 Results (Volume I): The State of Learning and Equity in Education.
- OECD (2026). OECD Digital Education Outlook 2026: Exploring Effective Uses of Generative AI in Education. OECD Publishing, Paris. <https://doi.org/10.1787/062a7394-en>.

- Schut, S., Maggio, L.A., Heeneman, S., van Tartwijk, J., van der Vleuten, C., & Driessen, E. (2021). Where the rubber meets the road — An integrative review of programmatic assessment in health care professions education. *Perspectives on Medical Education*, 10(1), 6–13. <https://doi.org/10.1007/s40037-020-00625-w>
- SpaceX (2024). Starlink Specifications and Coverage Map.
- Tlili, A., Garzón, J., Salha, S., Huang, R., Xu, L., Burgos, D., ... & Wiley, D. (2023). Are open educational resources (OER) and practices (OEP) effective in improving learning achievement? A meta-analysis and research synthesis. *International Journal of Educational Technology in Higher Education*, 20(1), 54. <https://doi.org/10.1186/s41239-023-00424-3>
- UNESCO (2022). Transforming Education Summit Discussion Paper.
- UNESCO (2023). Global Report on Teachers: Addressing Teacher Shortages.
- UNESCO Institute for Statistics & GEM Report (2025). 2025 SDG 4 Scorecard: Focus on the Out-of-School Rate.
- VanLehn, K. (2011). The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems. *Educational Psychologist*, 46(4), 197–221.
- World Bank, UNESCO, UNICEF (2022). The State of Global Learning Poverty: 2022 Update.
- Zhao, C., Cheng, Y., Li, B., Yang, Y. & Wang, X. (2025). DFedKG: Diffusion-Based Federated Knowledge Graph Completion. *Data Science and Engineering*, 10, 639–652. <https://doi.org/10.1007/s41019-025-00292-z>