KTA Governed Knowledge and Professional Learning

A reference architecture for personalized education, distributed knowledge and responsible AI in rural healthcare

Lanre Ogunkunle, DSc | Pleyverse Labs
Research architecture paper | Version 1.1 | November 2 2026

Abstract

Professional education frequently separates instruction from the later application of knowledge. AI assistance can bridge that separation, but personalization alone does not establish the reliability, ownership or permitted reuse of what a system learns. This paper proposes a learning and professional-practice embodiment of Knowledge Twin Architecture (KTA). It connects a versioned knowledge foundation, an evidence-linked professional Twin, AI orchestration, governance records and controlled knowledge exchange. A glossary supplies semantic anchors; lessons, flashcards and scenarios produce distinguishable forms of learning evidence; and PLEYCUP provides a compact interface to relevant knowledge and continuing practice support. Reviewed contributions may improve retrieval resources or enter a separately authorized open-weight model refinement pipeline. Individual, organizational and shared knowledge remain distinct permission domains. Rural healthcare illustrates the potential value of flexible instruction, locally grounded knowledge and connectivity-aware deployment. The paper specifies data structures, operating boundaries, deployment alternatives and a prospective evaluation agenda. It presents an architecture and testable hypotheses, not empirical evidence of learning gains, clinical safety, cost savings or population-scale delivery.

Keywords: Knowledge Twin Architecture; governed knowledge; professional learning; rural healthcare; open-weight models; AI governance; PLEYCUP; MCML.

1 Purpose and research contribution

The research objective is continuity: knowledge encountered during education should remain available, attributable and usable during later professional activity. At the same time, useful experience should be able to improve shared educational knowledge without silently transferring private information or treating an AI-generated explanation as established fact.

The proposed contribution is an explicit connection among knowledge identity, evidence of understanding, permissioned professional context and governed AI improvement. KTA is domain-independent; nursing and physician education are initial embodiments. Its scope is the underlying knowledge and governance infrastructure shared by these embodiments.

2 Research lineage and conceptual boundaries

Collaborative Knowledge Twins A Reference Architecture for Enterprise AI describes governed knowledge exchange across sovereign organizations [1]. Healthcare Data Democratization A Manifesto for Patient Controlled Health Intelligence develops the related patient-control and health-intelligence research direction [10]. This paper develops a compatible professional-learning embodiment. The relationship below is an architectural interpretation of those manuscripts, rather than a claim that the combined architecture has been empirically evaluated.

Dimension	Healthcare democratization embodiment	Learning and practice embodiment
Starting material	Patient-controlled information and clinical expertise	Curriculum, sources, exercises and professional experience
Primary purpose	Make fragmented health information useful under appropriate control	Connect education to retention and responsible application
Twin context	Relevant health information linked to expert knowledge	Knowledge relationships, demonstrated capability and permitted practice context
Shared contribution	Governed abstractions across healthcare participants	Reviewed explanations, scenarios and workflow knowledge
AI role	Contextualize permitted information and expertise	Teach, research, rehearse and support knowledge application
Control boundary	Patient and institutional permissions	Individual, organizational and shared permissions

Relationship to established work

Knowledge tracing estimates aspects of student knowledge from educational interactions [2]. KTA incorporates that function but also represents content provenance, application context and reuse permissions. Retrieval-augmented generation connects model outputs with external memory [3]; KTA adds explicit authority, review and lifecycle controls around retrieved material. W3C PROV provides a vocabulary for entities, activities and agents that can inform provenance interchange [4]. These foundations are complementary; a knowledge graph or retrieval system alone does not constitute the full proposed Twin.

The architecture does not assert a digital replica of a person, a complete representation of expertise, or a reliable inference of competence from clicks. Its novelty proposition concerns the specified integration and its evaluation. Comparative evaluation must establish the incremental value of that integration.

3 Reference architecture

KTA is a governed connection between a professional, a versioned body of knowledge and evidence about learning and application. The Twin is a queryable representation with uncertainty, not a model weight file or a single proficiency score. Its records remain useful when the underlying language model changes.

Component	Proposed responsibility	Representative output
Knowledge foundation	Link terms, sources, lessons, scenarios and requirements using stable identifiers	Versioned concept and evidence graph
Professional Twin	Represent observations, demonstrated skills, misconceptions and permitted context	Evidence-linked state with uncertainty
AI orchestration	Select context and tools for explanation, research, rehearsal and feedback	Grounded response and action trace
Verification and governance	Enforce permissions, review states and release conditions	Policy decisions and audit records
Model infrastructure	Route tasks to approved commercial or open-weight models	Versioned execution manifest
Improvement pipeline	Curate contributions for retrieval, evaluation or authorized training	Reviewed knowledge and model releases

Interfaces and execution path

The web application, mobile application, PLEYCUP interface and physical learning artifacts are presentation surfaces over common services. They reference the same content identifiers and synchronize permitted learning events. A physical object can link to a concept or exercise without itself storing sensitive learner information.

A typical request passes through authentication and purpose authorization, Twin-context selection, permission-filtered retrieval, tool and model execution, verification and response delivery. A separate event records what occurred. Only eligible observations update the Twin; candidate contributions enter review before any shared publication. External writes or clinical-system actions require separately defined permissions and confirmation appropriate to the task.

A lightweight initial system can use a content registry, relational event store, retrieval index and policy service. Specialized graph infrastructure and learned mastery estimation can follow measured requirements. The architecture does not depend on training a foundation model from scratch.

4 Knowledge representation and governance

An operational definition of the professional Twin is T(u,t) = {K, E, S, C, P}: relevant knowledge references K; observed evidence E; estimated learning state S; permitted context C; and policies P for professional u at time t. Learning state is derived from evidence and remains contestable. The organization owns or licenses its resources; individuals and other rights holders retain applicable rights in their contributions.

Record	Minimum proposed fields
Knowledge object	Object ID; version; concept links; author or source; license; audience; jurisdiction or setting; review state; effective and review dates
Learning observation	Event ID; pseudonymous learner ID; activity and concept versions; response; rubric; assessor or model version; assistance used; timestamp
Twin assertion	Claim and context; supporting evidence IDs; uncertainty; inference method; last assessment; correction or supersession link
Governed contribution	Contributor and tenant; source lineage; disclosure classification; allowed audiences; permitted uses; reviewer; status; expiry
AI execution	Request and task IDs; approved context references; model and adapter versions; tool actions; validation results; response and review linkage

Separate authorization for separate uses

Access for personal learning, access to an employer resource, publication to a partner network and use for model training are different decisions. A sharing permission must specify audience, purpose and duration. Default sharing must not expose personal learning gaps to employers or sponsors. Assessment records should not become employment decisions without a separately governed and disclosed process.

Governance enforcement belongs in authenticated retrieval, APIs and tool execution, rather than only in prompts. Retrieval indexes, caches and exported packages must preserve tenant boundaries. Review status, source authority and age must remain visible; conflicting guidance is surfaced, not silently averaged. A clinical guideline, a local workflow and an anecdotal contribution have different authority.

MCML, the author-developed Model Connect Markup Language, supplies a proposed documentation and linking mechanism for model, dataset, interface and agent records [5]. The content and Twin schemas above are proposed extensions or linked application records, not assertions that every field already exists in MCML. Validation can check documentation completeness and referential integrity; it does not by itself establish factual correctness, safety or regulatory compliance.

5 Learning and research during professional activity

The reference nursing learning corpus comprises 650 foundational AI glossary terms and flashcards covering approximately 2,500 terms from AI by Practice (aibypractice.com), the public professional learning platform. Stable catalog records distinguish concepts, alternative labels, flashcards and questions, linking each to its source, version and learning objectives. These content counts describe the embodiment; educational effectiveness is assessed separately.

The interface has eight panels and seven shared functions: LEARN, ATTEMPT, PRACTICE, VERIFY, DEMONSTRATE, SAVE and SHARE. Panels and functions are separate design dimensions; a function may be available across multiple panels. Explain, Question, Rehearse, Critique and Reflect provide interaction modes. A panel is a navigation and learning unit; it is not assumed to correspond one-to-one with a concept or competency.

Learning or use event	What the system may infer	What it must not infer
Read a glossary definition	Exposure to a concept	Competence or retention
Answer a flashcard	Item-level retrieval evidence	General mastery from one answer
Complete a reviewed scenario	Performance against a defined rubric	Patient-care competence without further evidence
Research and verify an answer	Use of evidence-checking steps	That a cited source supports every claim
Save or share an artifact	A retained or submitted contribution	Permission for public reuse or model training

Research while an activity is underway

For a question arising during study or professional work, the agent identifies the task, retrieves authorized sources, checks source dates and applicability, resolves citation identifiers and assesses whether the cited passages support its material claims. It presents supporting evidence, limitations and unresolved conflicts. If evidence is inadequate or the task exceeds scope, it abstains or directs the user to an appropriate human resource. The execution record identifies what was retrieved and checked; it must not present a generated reasoning narrative as a faithful internal trace.

In the physician embodiment, human-led AI fundamentals establish an initial baseline. Subsequent AI-guided instruction is personalized, with human verification of educational content and escalation when needed. A 10 percent human and 90 percent AI-supported allocation is a delivery design target to be measured, not a proven optimum. Group instruction is optional when schedules permit; individual participation does not depend on cohort formation.

6 Distributed knowledge and open weight refinement

The system has three distinct improvement processes. Conflating them creates misleading promises that every interaction trains a model or that a fine-tuned model is the authoritative knowledge base.

Improvement process	What changes	Control required
Personalization	Twin context, learning sequence and selected exercises	Evidence quality, user visibility and correction
Knowledge improvement	Reviewed sources, explanations and retrieval resources	Rights, subject review, versioning and withdrawal
Model refinement	A model or adapter trained for a defined task	Training authorization, dataset lineage and release evaluation

A distributed deployment exchanges approved knowledge packages rather than routinely pooling raw conversations or patient records. Packages include concept mappings, sources, permitted uses, review records and version dependencies. Recipient organizations validate authority and local applicability before adoption. Local procedures remain locally scoped. Revocation or expiry invalidates retrieval eligibility and dependent caches; previously downloaded copies require explicit contractual and technical controls.

Controlled model refinement

Candidate examples pass through permission checks, privacy and rights review, subject-matter verification, deduplication and dataset versioning. Separate training, validation and held-out evaluation sets must be split at learner, organization or scenario level where needed to avoid leakage. Model selection follows a measured task requirement. Parameter-efficient techniques such as LoRA are possible methods, not assumptions about the best approach [6].

A candidate release is compared with the existing model and a retrieval-only baseline for task quality, citation support, privacy leakage, safety behavior, subgroup performance, latency and total cost. Approved releases retain dataset, model, adapter and evaluation identifiers and a rollback path. Newly generated answers cannot recursively become trusted training examples without independent review.

Distributed knowledge exchange is not synonymous with federated learning. Federated optimization or adapter exchange would require separate evaluation of information leakage, poisoning and cross-site compatibility. Deleting a source from a retrieval index does not remove its influence from already trained weights; training consent and withdrawal terms must address that limitation. Open-weight licensing, deployment rights and permitted uses must be checked for each selected model.

7 Rural healthcare embodiment and anticipated benefits

HRSA identifies persistent rural workforce and access constraints [7]. KTA cannot substitute for clinicians, infrastructure or specialist services. Its research opportunity is to make educational support more accessible and locally useful while preserving professional accountability. Benefits below are hypotheses to test, not demonstrated outcomes.

Rural practice need	Proposed response	Evaluation indicator
Limited training time	Resumable short sessions and individual learning paths	Participation, interruption recovery and learning time
Limited access to educators	AI explanations and rehearsal with reviewed content and escalation	Rubric performance and escalation resolution
Difficulty applying training later	Evidence-linked glossary, PLEYCUP and continuing toolkit	Delayed retention and scenario transfer
Expertise concentrated in few staff	Capture approved explanations and local workflow knowledge	Reuse quality and author review burden
Different local resources	Retrieve current local policies and service information	Local relevance and out-of-scope recommendations
Intermittent connectivity	Downloadable reviewed resources and optional local inference	Offline task success and synchronization failures

Illustrative scenario

A rural nurse studies AI-generated patient education using synthetic cases. The nurse correctly defines hallucination but accepts an unsupported reference during a scenario. The Twin records the discrepancy, the rubric and the assistance used, rather than declaring general competence. The agent offers targeted practice and retrieves the current approved patient-education policy. The nurse then demonstrates verification on a different case; the assessment becomes additional evidence, not a replacement for the earlier record.

With authorization, an educator converts a useful explanation from that interaction into a generalized learning contribution, removes local identifiers and submits it for review. A partner practice can adopt the explanation while keeping its own procedures and learner records private. A rural physician follows the same architecture with specialty-appropriate examples and separately reviewed content. No real patient data are needed for these learning pathways.

For professional associations and health systems, this connects initial education with continuing support and a contribution pathway. Reach, learning and sustained application must be reported separately. A national ambition such as educating one million nurses is a dissemination target, not evidence of current enrollment, proficiency or benefit.

8 Deployment and responsible operation

Deployment	Potential advantage	Requirement or tradeoff
Managed cloud	Central updates and shared operations	Connectivity, vendor controls and usage cost
Local open weight	Local inference and selectable data boundary	Hardware, maintenance, updates and local evaluation
Hybrid	Route suitable tasks to local or remote resources	Explicit routing, egress policy and consistent controls
Offline learning package	Reviewed resources without continuous connectivity	Expiry handling, limited features and safe synchronization

Local inference is a deployment property, not a guarantee of privacy or lower cost. Total cost includes equipment, energy, staffing, inference, training, updates, security and availability. On reconnection, offline events retain timestamps and package versions; duplicate events are reconciled, and superseded content is flagged. Time-sensitive professional questions must not be answered from expired material as if it were current.

Operating boundaries

The initial intended use is professional education, rehearsal, evidence navigation and AI-governance assistance. Automated diagnosis, treatment selection, medication dosing and autonomous entry into clinical systems are outside this embodiment. Any later clinical use requires its own intended-use definition, evidence, institutional authorization and applicable review. Educational capability does not establish clinical validity.

The principal threats include fabricated citations, outdated guidance, prompt injection in retrieved material, cross-tenant access, sensitive-data disclosure, poisoned contributions and false confidence in inferred competence. Controls include permission-filtered retrieval, isolated tool execution, independent source checks, contributor review, version constraints and monitored escalation. These reduce specified risks; they do not guarantee that generative outputs are correct [8,9].

Model, content and policy changes have different release paths. A wording correction should not force a model retrain; a model update should not silently alter an assessment rubric. Each deployment identifies its model and adapter, knowledge snapshot, policy version, assessment version and verification configuration. The operational record links those versions to governance documentation and review decisions.

Accessibility includes low-bandwidth text, screen-reader support and alternatives to voice, animation or physical objects. Language adaptation requires meaning and source-equivalence checks. No user should need a physical cup, a headset or merchandise to access the core learning pathway.

9 Evaluation agenda and research implications

H7 is the primary hypothesis: Twin-guided learning with continuing governed support improves delayed application performance relative to the same content without Twin-based personalization. H7 is a prospective hypothesis, not an observed effect. Further hypotheses address evidence verification, governance reliability, rural feasibility and the incremental value of model refinement. Study design and analysis should be specified before outcome inspection.

Domain	Proposed measure	Interpretation boundary
Reach and equity	Unique participants and active learners by setting; rural definition reported	Registrations are not completions or learning gains
Learning and retention	Equivalent pre/post and delayed scenarios scored with a reviewed rubric	Self-confidence is a secondary outcome
Continued application	Verification performance on new scenarios and appropriately scoped toolkit use	Usage frequency does not prove benefit
Research reliability	Source validity, claim support, freshness and appropriate abstention	Citation presence does not establish correctness
Governance	Unauthorized retrieval tests, lineage completeness and withdrawal propagation	Schema validity is not clinical safety
Operational feasibility	Latency, uptime, offline recovery, human review time and total cost	Local hosting is not presumed cheaper

Prospective study design

Use expert-reviewed synthetic scenarios and a frozen baseline. Compare equal-content groups with and without Twin-guided personalization where feasible, using blinded assessors. Observational studies must report design limitations. Report denominators, attrition, assistance use, effect estimates with uncertainty and results by setting. Select sample size from the primary outcome and anticipated variability.

Appropriate ethics review or determination, consent and data-use agreements precede studies involving participants or sensitive records. Participation rewards should remain distinct from proficiency scoring and research enrollment. Initial educational outcomes cannot establish reduced clinical errors, improved patient outcomes or national-scale effectiveness.

10 Research program and implementation alignment

H7 belongs to the Foundations hypothesis register as the delayed-application hypothesis. It is distinct from the healthcare manifesto hypotheses H2 on epistemic stability and H3 on patient understanding. A learning benefit under H7 would not establish either of those outcomes. Each requires its own population, measures and analysis.

Research identifier	Role and method	Interpretation boundary
Register H7	Test delayed application using reviewed synthetic scenarios, a frozen baseline, equal-content groups and blinded assessors	No clinical or population-scale outcome follows from learning performance alone
Program B experiment H2	Use the H7 comparison method within Sovereign Knowledge and Data	A local experiment ID; distinct from manifesto H2
Program C study	Use the same comparison method for PLEYCUP as a presentation surface over common services	Keep interface exposure equivalent or isolate it as a separate experimental factor

The shared protocol should pre-specify the delay interval, primary rubric score, assessor masking, assistance permitted, content snapshot, model versions, allocation method, attrition handling and statistical analysis. The comparison holds content, access period and assessment conditions constant, varying Twin-based support. Any evaluation of the physical board or compact PLEYCUP interface should separately identify the interface effect. The study protocol is forthcoming; this paper is not a preregistration.

Specification first implementation

The implementation order aligns with the Foundations plan: first specify and implement stable content identifiers, evidence records and policy enforcement; second add permission-filtered retrieval; third enable reviewed contribution workflows; and fourth consider open-weight refinement only when evaluation demonstrates a need and an authorized dataset is available. Content versions, evidence schemas and policy contracts connect each research protocol to the corresponding platform release, preventing product and lab definitions from drifting.

Shared services across professional domains

The same Twin representation T(u,t), evidence schema, provenance contracts and authorization services can support clinicians and engineers. Domain-specific curricula, competency rubrics, source authorities and policies remain separately versioned. AI by Practice and the Systems Lab engineering bootcamp can therefore use a common KTA service layer while presenting different learning experiences. This is a reference architecture for reuse, not a claim that migration or platform consolidation has occurred. Public branding does not determine whether the backend services must be duplicated.

Conclusion

KTA offers a research framework for preserving continuity among learning, governed knowledge and professional application. Its potential rural value lies in flexible support, local relevance and controlled knowledge exchange. The test is whether these mechanisms produce measurable benefit while preserving boundaries, provenance and professional accountability. PLEYCUP is an interface to that architecture; the architecture is the durable research contribution.

References and research provenance

Primary sources and research context consulted November 2 2026. Numbered references support the surrounding factual or methodological statements; proposed system behavior and benefits are the author's architectural synthesis.

[1] Ogunkunle L. Collaborative Knowledge Twins A Reference Architecture for Enterprise AI. Pleyverse Labs research manuscript. 2026.

[2] Piech C, Bassen J, Huang J, et al. Deep Knowledge Tracing. Advances in Neural Information Processing Systems 28. 2015.
https://proceedings.neurips.cc/paper/2015/hash/bac9162b47c56fc8a4d2a519803d51b3-Abstract.html

[3] Lewis P, Perez E, Piktus A, et al. Retrieval Augmented Generation for Knowledge Intensive NLP Tasks. NeurIPS. 2020. arXiv 2005.11401.
https://arxiv.org/abs/2005.11401

[4] Groth P, Moreau L, editors. PROV Overview An Overview of the PROV Family of Documents. W3C Working Group Note. April 30 2013.
https://www.w3.org/TR/prov-overview/

[5] MCML Standard. Model Connect Markup Language for AI Governance. Author-developed specification and documentation.
https://mcml-standard.org/

[6] Hu EJ, Shen Y, Wallis P, et al. LoRA Low Rank Adaptation of Large Language Models. 2021. arXiv 2106.09685.
https://arxiv.org/abs/2106.09685

[7] Health Resources and Services Administration. Rural Access to Health Care Services Request for Information. Background on rural workforce and access challenges.
https://www.hrsa.gov/rural-health/rfi-rural-health-care-access

[8] World Health Organization. WHO releases AI ethics and governance guidance for large multi modal models. January 18 2024.
https://www.who.int/news/item/18-01-2024-who-releases-ai-ethics-and-governance-guidance-for-large-multi-modal-models

[9] Autio C, Schwartz R, Dunietz J, et al. Artificial Intelligence Risk Management Framework Generative Artificial Intelligence Profile. NIST AI 600-1. July 2024.
https://doi.org/10.6028/NIST.AI.600-1

[10] Ogunkunle L. Healthcare Data Democratization A Manifesto for Patient Controlled Health Intelligence. Publication edition v0.9. August 2026.

Authorship and interests

The author is associated with the Pleyverse research and product ecosystem and is the developer of MCML. The author reports a provisional patent application relating to KTA and has a commercial interest in its implementation. These relationships are relevant to interpretation of the proposed architecture and its evaluation. A provisional application is not an issued patent. No endorsement by a professional association, healthcare institution or cited standards organization is implied.