

Creator Expertise as Competitive Moat: Why Custom AI Trained on Proprietary Methodology Outperforms Generic AI in Education

Humberto Inciarte

May 2026

Abstract

The widespread availability of customizable large language models (LLMs) — particularly OpenAI’s Custom GPT framework, with over three million custom assistants created since 2023 — has produced a prevailing assumption among independent course creators: that a free or low-cost custom assistant trained on uploaded course materials is functionally equivalent to a system engineered exclusively around the creator’s proprietary instructional methodology. This paper challenges that assumption. Drawing on peer-reviewed evidence from three converging research streams — context-memory conflict in retrieval-augmented generation (RAG), domain-specific versus general-purpose model performance benchmarks, and instructional alignment as the principal mediator of intelligent tutoring system effectiveness — the analysis demonstrates that generic AI tutors and methodology-aligned AI tutors are not two implementations of the same product but two distinct epistemic objects. Documented hallucination rates of 45–58% in professional knowledge tasks, persistent parametric bias even when ground-truth context is supplied, and meta-analytic effect sizes that depend critically on alignment between the tutoring system and the course’s specific objectives all indicate that the architectural choice is consequential for learner outcomes. The paper introduces the CursoVivo implementation model — a constraint-based design in which the generative behavior of the AI is bounded exclusively by the creator’s codified methodology — as a framework that operationalizes proprietary expertise into a defensible competitive moat in the creator economy. Implications for course creators, ed-tech researchers, and the broader debate over data as a strategic asset in the foundation-model era are discussed.

Author: Humberto Inciarte

Affiliation: Department of Research, Arquitectos de la Felicidad / NeuroFlow 30H™

ORCID: [0009-0006-0379-7328](https://orcid.org/0009-0006-0379-7328)

Corresponding author: hola@humbertoinciarte.com

Date: May 2026

Keywords: custom AI education, proprietary methodology AI, ChatGPT vs specialized AI, course creator technology, AI competitive advantage, intellectual property AI, retrieval augmented generation, fine-tuning, LLM hallucination education, CursoVivo

1. Introduction

1.1 The Prevailing Narrative

Since the public release of OpenAI’s Custom GPT builder in late 2023, the dominant narrative among independent online educators has been one of disruptive equivalence. By 2025, OpenAI had reported more than three million Custom GPTs created on its platform, of which approximately 159,000 are publicly listed in the GPT Store (SEO.AI, 2025). Marketing literature, creator-economy newsletters, and informal advice

within course-creator communities have converged on a recurring claim: a creator can upload their course materials — videos, PDFs, transcripts, slide decks — into a Custom GPT, and the resulting assistant will function as a personalized tutor capable of replacing or supplementing the human instructor at near-zero marginal cost. The argument is intuitive. If the foundation model is already capable of fluent dialogue, and if retrieval-augmented generation (RAG) can ground responses in the creator’s own documents, then any specialized tutoring system carrying a four-figure price tag must, by implication, be charging for a feature the market already offers for free.

This narrative has practical consequences for the broader e-learning sector. The global e-learning services market was valued at USD 299.67 billion in 2024 and is projected to reach USD 1.31 trillion by 2032 (Fortune Business Insights, 2025), with Latin America growing at a compound annual rate that consultancies place between 14.5% and 16.8% (Future Market Insights, 2025; Market Data Forecast, 2025). Within this market, single creators on platforms such as Hotmart and Teachable have collectively distributed more than ten billion dollars in earnings across over 200,000 creators (Hotmart, 2024). For each of these creators, the question of whether to invest in a methodology-specific AI system or to assemble a Custom GPT is no longer hypothetical — it is a recurring procurement decision made under uncertainty, often resolved by intuition rather than evidence.

1.2 The Problem

The assumption of equivalence rests on a category error. It treats the AI tutor as an information-retrieval interface — where success is measured by whether the system can locate relevant material — when, in the educational context, the system functions as an instructional surrogate, where success depends on whether its outputs are aligned with the specific pedagogical model the learner originally enrolled to acquire. These are not the same evaluation criterion. A retrieval interface that occasionally surfaces consensus knowledge from its training corpus instead of the creator’s specific method may be acceptable in a general assistant. In an educational setting, the same behavior introduces a documented psychological mechanism: cognitive dissonance between the instructor’s stated method and the AI’s response, which the learner must resolve through one of three pathways — reconciliation, dismissal of one source, or disengagement (Festinger, 1957; Limón, 2001). The third pathway is, in operational terms, course abandonment.

The cost of abandonment in online education is substantial and chronic. A meta-analysis of 221 massive open online courses found a median completion rate of 12.6%, with a range from 0.7% to 52.1% (Jordan, 2015). A six-year longitudinal analysis of HarvardX and MITx courses concluded that the majority of learners never returned after their first year, that completion rates had not improved across the period studied, and that the disruptive promise of MOOCs failed to materialize (Reich & Ruipérez-Valiente, 2019). These figures predate the integration of generative AI into educational platforms. The relevant question for this paper is not whether AI can lift completion rates — substantial evidence suggests it can — but whether the *type* of AI matters.

1.3 Research Question and Thesis

This paper examines whether the prevailing assumption — that a Custom GPT built on uploaded course materials is functionally equivalent to a system engineered around a creator’s proprietary methodology — holds up under empirical scrutiny. The thesis advanced is that it does not. Three converging streams of

peer-reviewed research indicate that generic foundation models and methodology-constrained AI tutors produce systematically different outputs in educational settings: (a) documented context-memory conflict in retrieval-augmented systems, in which the model’s parametric knowledge overrides the supplied context; (b) consistent benchmark superiority of domain-specific models in narrow, idiosyncratic knowledge domains; and (c) the centrality of instructional alignment to tutoring effectiveness. From these findings, the paper derives a framework — the *CursoVivo* implementation model — in which the AI tutor’s generative behavior is bounded exclusively by the creator’s codified methodology, and in which proprietary expertise functions as a defensible competitive moat against commoditized general-purpose AI. The remainder of the paper proceeds as follows: Section 2 reviews the relevant literature; Section 3 presents the analysis and proposed framework; Section 4 offers conclusions, limitations, and directions for future research.

2. Literature Review

2.0 Methodological Note

This review synthesizes peer-reviewed empirical studies, archival preprints from arXiv with established citation traction, organizational case reports, and institutional data published between 2011 and 2026, sourced from Google Scholar, Semantic Scholar, the ACL Anthology, the National Library of Medicine (PubMed/PMC), arXiv, the SAGE Journals catalog, and reputable industry research from established consultancies. Inclusion criteria prioritized studies that (a) measured factual accuracy or hallucination rates in LLM outputs against verifiable ground truth, (b) compared domain-specific against general-purpose models on benchmark tasks, or (c) reported effect sizes for personalized or AI-mediated tutoring on learning outcomes. Grey literature was included only where peer-reviewed evidence was unavailable and where the source was independently verifiable (e.g., directly from OpenAI documentation, vendor press releases, or industry consultancies with published methodology).

2.1 Context-Memory Conflict in Retrieval-Augmented Generation

The architectural premise of a Custom GPT trained on uploaded materials is that retrieval-augmented generation — the practice of injecting relevant document fragments into the model’s context window at inference time — will cause the model to ground its responses in the supplied source. A growing body of evidence indicates that this premise is unreliable in practice.

Kortukov et al. (2024) examined LLM behavior under context-memory conflict by presenting models with realistic documents that contradicted incorrect parametric (pre-training) knowledge. Their findings document a *parametric bias* in which the model retains incorrect prior beliefs even when supplied with the correct factual evidence in context. The conflict is not resolved automatically by the architecture; it is mediated by factors the system designer does not fully control.

Gao et al. (2025) extended this analysis through hidden-state probing, showing that LLMs disproportionately amplify contextual signals consistent with their parametric knowledge while attenuating signals that contradict it. The result is a class of failure mode in which the system produces confidently worded, retrieval-grounded responses that are nevertheless unfaithful to the source — a phenomenon difficult to detect through surface-

level review.

Joren et al. (2024) introduced the construct of *sufficient context* as a diagnostic lens for RAG systems, demonstrating that hallucinations occur not only when retrieved context is insufficient but also when the context is technically sufficient yet stylistically or structurally divergent from the patterns learned during pre-training. Wood et al. (2024), in “RAGged Edges,” document a similar effect: when retrieved content departs from canonical templates, additional context does not improve accuracy and may, in some configurations, degrade it.

Hicks and Sinha (2025) provide a complementary upper bound from a detection perspective, establishing certified theoretical limits on embedding-based hallucination detection in RAG systems. Their result indicates that a non-trivial proportion of hallucinations are formally invisible to automated semantic-similarity audits, regardless of calibration. Taken together, this body of work establishes that the act of uploading proprietary content to a generic foundation model does not, by itself, constrain the model to that content.

2.2 Hallucination Rates in Professional Knowledge Domains

Empirical measurements of hallucination rates in domains adjacent to specialized education — law, medicine, public health — provide quantitative anchors for the foregoing theoretical analysis. Dahl et al. (2024) profiled legal hallucinations in GPT-3.5, ChatGPT, Llama-2, and PaLM-2 across verifiable legal tasks and reported hallucination rates of at least 58%. The same study found that the models accepted incorrect premises supplied by the user without challenge and were poorly calibrated to predict their own errors.

Yao et al. (2024) developed the FactChecker framework and reported factual error rates of up to 45% across GPT-3.5, GPT-4, Vicuna, and LLaMA-2. Critically, the same study demonstrated that domain-specific fine-tuning improved Llama-2-13B-chat accuracy from 35.3% to 68.5% — a near-doubling that was not attainable through prompt engineering alone. In biomedical and clinical contexts, evaluations of cancer-information chatbots have reported hallucination rates ranging from 19% (GPT-4 with web search) to 40% in conventional implementations, with curated retrieval pipelines reducing — but not eliminating — the rate (Han et al., 2025). These empirical magnitudes are relevant to educational settings: a tutor system that produces incorrect or off-method guidance in 30–50% of substantive interactions presents a non-trivial pedagogical risk.

2.3 Domain-Specific Versus General-Purpose Models

A separate but converging research stream compares the performance of models trained or fine-tuned on a specialized corpus against general-purpose foundation models on tasks within the specialized domain. Wu et al. (2023) introduced BloombergGPT, a 50-billion-parameter model trained on 363 billion tokens of financial data alongside 345 billion tokens of general text, and demonstrated significant performance gains on financial benchmarks without degradation on general tasks. Singhal et al. (2023) reported that Med-PaLM 2 surpassed human physicians on eight of nine axes of clinical utility — including factuality, medical reasoning, and low likelihood of harm — and was preferred over the human physician in 72.9% of pairwise consensus comparisons.

Wang et al. (2024) evaluated GPT-3.5 and GPT-4 on osteoarthritis-specific medical questions and found generalist models materially less effective than DocOA, a fine-tuned domain assistant, particularly for

personalized recommendations. The pattern is consistent across legal, financial, and biomedical evaluations: when the domain is specialized and the evaluation rewards faithful reproduction of expert reasoning rather than fluent prose, models trained on the specialized corpus outperform general models by margins large enough to be operationally meaningful.

This pattern is, however, not absolute and must be qualified. Nori et al. (2023) demonstrated that GPT-4, when paired with a structured prompting strategy (“Medprompt”), surpassed Med-PaLM 2 on the MedQA benchmark and generalized to other professional examinations. The implication is consequential for the present argument: the gap between general and specialized models is not fixed, and clever prompting can close it in domains whose knowledge is already represented in the public training corpus. The advantage of specialization persists most strongly in domains that are *idiosyncratic* — that is, where the content of interest is not present in public training data, regardless of how sophisticated the prompting. The methodology of an individual course creator, by definition, satisfies this idiosyncrasy condition: the creator’s specific framework, sequence, terminology, and calibrated examples are not representative of public corpora and therefore are not recoverable through prompt engineering applied to a generalist model.

2.4 Instructional Alignment and Tutoring Effectiveness

The pedagogical literature offers a third relevant line of evidence. Kulik and Fletcher (2016), in a meta-analysis of 50 controlled evaluations of intelligent tutoring systems, reported a median effect size of 0.66 standard deviations — an improvement equivalent to moving the average student from the 50th to the 75th percentile. The most consequential finding for present purposes was not the magnitude of the effect but its dependence on instructional alignment: the effect was substantially larger when the assessment was aligned with the tutoring system’s specific instructional objectives than when assessed against externally standardized tests. The authors concluded that the degree of improvement attributable to a tutoring system depends primarily on whether the assessment is congruent with the system’s stated instructional goals.

VanLehn (2011), in a complementary analysis of human tutoring, intelligent tutoring systems, and other tutoring formats, reported that human tutors produced effect sizes of 0.79 SD, step-based intelligent tutoring systems 0.76 SD (functionally equivalent to human tutoring), and answer-based computer-assisted instruction only 0.40 SD. The granularity of feedback and its alignment with the underlying instructional model accounted for the variance.

Studies of perceived credibility complement these effectiveness findings. Lee, Chen, and Liu (2024) found that students perceive AI tutors as having less control over response quality than human tutors, and that human tutors retain a credibility premium when learner involvement is high. The implication is that AI tutoring systems whose responses diverge from the human instructor’s stated method risk eroding the perceived authority of the instructor — an effect with downstream consequences for engagement and persistence.

2.5 Research Gap

The literature reviewed above documents three phenomena: (a) generic foundation models retain parametric biases that override supplied context in reproducible ways; (b) specialized models outperform generalists in idiosyncratic domains; and (c) instructional alignment is the principal mediator of tutoring effectiveness.

What the literature does not yet provide is a synthesizing framework that translates these findings into design principles for AI implementations in the creator economy, where the unit of value is not a published curriculum but a single creator's tacit, codified, proprietary methodology. The present paper contributes to this gap by proposing such a framework and articulating its implications.

3. Analysis and Discussion

3.1 The Architectural Distinction That Matters

The principal finding from the literature reviewed in Section 2 can be stated as follows: a Custom GPT built on a creator's uploaded materials and a system whose generative behavior is bounded exclusively by the creator's proprietary methodology are not two implementations of the same product. They are two distinct epistemic objects. The first answers from the consensus distribution of its public training corpus, modulated — incompletely and inconsistently — by retrieved fragments of the creator's content. The second answers from a constrained behavior space defined by the creator's method. The output difference is not stylistic; it is structural.

This distinction maps onto a documented technical reality. Kortukov et al. (2024), Gao et al. (2025), and Joren et al. (2024) collectively establish that retrieval into a foundation model is an *augmentation*, not a *replacement*, of the model's parametric knowledge. The model's pre-training continues to shape outputs even when authoritative source material is in context. In a general assistant, this characteristic is benign or even useful, because the model's broader knowledge supplies plausible defaults when the supplied context is incomplete. In an educational tutoring application, the same characteristic is the source of the failure mode this paper addresses: the AI silently substitutes general consensus for the creator's specific method.

The phenomenon has a clean reformulation. A generic AI tutor surfaces what the public corpus collectively believes about a topic. A methodology-aligned AI tutor surfaces what the creator empirically teaches. These are not equivalent descriptions of the same output.

3.2 Quantifying the Pedagogical Cost

The aggregate cost of off-method tutoring in online education can be estimated from two converging data streams. First, hallucination rates documented in adjacent professional domains range from 19% in best-case retrieval-augmented configurations to 58% in unaugmented general models (Dahl et al., 2024; Yao et al., 2024; Han et al., 2025). Even taking the lower bound, an AI tutor that produces off-method guidance in roughly one in five substantive interactions presents a meaningful pedagogical risk over the duration of a course.

Second, the persistence of off-method guidance interacts with the cognitive dissonance literature in a manner that is operationally consequential. When a learner receives an instruction from the course (the human creator's method) and a divergent instruction from the AI tutor embedded in the course experience, the resulting inconsistency requires cognitive resolution (Festinger, 1957). The learner can reconcile, dismiss one source, or disengage. Field evidence on the credibility differential between AI and human tutors (Lee et al., 2024) suggests that, under conditions of conflict, the human instructor retains a credibility advantage — but

only at the cost of the learner now distrusting the AI tutor that was deployed to scale the instructor's method. Either way, the value proposition of AI-mediated personalization is compromised.

Translated into the units that course creators tend to track — completion, satisfaction, and downstream conversion — the pattern compounds. Online courses already exhibit chronically low completion rates, with peer-reviewed analyses placing the median MOOC completion rate at 12.6% (Jordan, 2015) and longitudinal analyses showing that completion rates have not improved over multi-year periods (Reich & Ruipérez-Valiente, 2019). Each non-completion represents not only a loss of educational outcome but a missed opportunity to generate the verified testimonial evidence on which the creator's subsequent enrollment depends. The marketing literature within the creator economy routinely frames the problem as one of insufficient lead generation; the present analysis suggests that the binding constraint is, instead, an *implementation gap* further downstream — the interval between content delivery and method execution by the learner.

3.3 Why Specialization Holds in Idiosyncratic Domains

Section 2.3 noted the contrary evidence: Nori et al. (2023) demonstrated that prompt engineering applied to GPT-4 can match or exceed specialized models in publicly represented domains. This finding deserves direct treatment, because it would, if generalized, weaken the present argument.

The Medprompt result holds for medicine because medical knowledge is densely represented in the public training corpora of frontier models — peer-reviewed literature, clinical guidelines, board-examination preparation materials, and large volumes of secondary commentary. A sufficiently capable generalist, prompted carefully, can recover the consensus position. The same finding does not generalize to domains that are *idiosyncratic by construction*. An individual creator's instructional methodology — the specific sequence in which concepts are introduced, the calibrated examples used, the diagnostic questions deployed at each stage, the terminology adopted, the framework that distinguishes this creator from competitors — is not present in public training data. No prompt strategy can recover what is not represented.

The implication for the creator economy is that the *moat* against generic AI substitution is not the underlying domain (which may be highly represented in training data: nutrition, fitness, marketing, finance) but the creator's *specific method within the domain*. The defensibility of a course offering against generic AI substitution is therefore proportional to the degree to which the creator's method is methodologically distinctive, codified, and operationalized — not the domain's exoticism. A creator who teaches generic nutrition guidance is substitutable. A creator with a proprietary diagnostic-and-protocol method, codified into an instructional sequence with calibrated decision points, is not — provided the AI implementation preserves that specificity.

3.4 The CursoVivo Implementation Model

The findings reviewed here suggest that the design space for AI-mediated personalization in the creator economy is bounded by two principal constraints: (a) the AI's generative behavior must be confined to the creator's proprietary methodology rather than augmented by public consensus, and (b) the implementation must preserve instructional alignment between the AI's outputs and the course's specific objectives. The CursoVivo model is proposed as a framework that operationalizes these constraints.

The model rests on three design principles. *Methodology-grounded constraint*: the corpus from which the AI generates is the creator’s codified method — videos, written materials, and the explicit decision logic the creator uses with learners — and the system is engineered to prevent substitution of this corpus by the foundation model’s parametric defaults. The objective is not to enrich the model with the creator’s content but to *bound* the model’s outputs to that content. This is an architectural commitment, not a content-management feature. *Implementation alignment*: the system’s outputs are designed to advance the learner along the course’s specific instructional sequence, including weekly action plans, scaffolded check-ins, and produced deliverables aligned with the method’s defined milestones. *Instructor authority preservation*: the system is positioned as an extension of the creator’s voice rather than as an independent advisor, with explicit constraints that prevent the AI from substituting general-domain advice for the creator’s method.

The contrast with adjacent commercial alternatives is concrete and relevant. Coachvox AI (2026) offers conversational chat interfaces at USD 99 per month per creator, with creator-club tiers at USD 3,000 per month for resale rights. Kajabi (2026), following a 2025 pricing revision, offers tiered course-platform plans ranging from USD 179 to USD 499 per month with embedded AI features focused on content authoring and Q&A. OpenAI’s Custom GPT, while frequently described as free, requires a USD 20 per month ChatGPT Plus subscription for the creator and, per OpenAI documentation, the same subscription for end-users to access the resulting GPT (OpenAI, 2025). The functional comparison is therefore not between a paid system and a free alternative but between four distinct architectures, each making different trade-offs between cost, control, and methodological fidelity. The CursoVivo implementation model is differentiated on the dimension of methodological fidelity — the dimension that, per the analysis above, is the principal determinant of pedagogical outcome.

3.5 Practical Implications

For independent course creators, the analysis carries three implications. First, the procurement decision between a Custom GPT and a methodology-aligned implementation is not a price comparison between functionally equivalent options; it is a choice between architectures that produce systematically different learner experiences. Second, the value of a creator’s instructional methodology — historically treated as background know-how rather than as a strategic asset — increases under conditions of generalist AI commoditization, because methodological specificity is precisely the property that generalist models cannot recover. Third, the relevant defense against generic AI substitution is not to compete with public foundation models on the dimension of fluent dialogue (a contest the foundation models will continue to win) but to deliver, at scale, the creator’s proprietary method with a fidelity that the foundation models cannot reproduce.

For ed-tech researchers and platform designers, the analysis suggests that the dominant frame of “AI-augmented learning” — in which a generalist model is grounded with course-specific materials — is insufficient as a basis for evaluating system effectiveness. Instructional alignment, treated as a measurable property of the system rather than as a marketing claim, deserves explicit benchmarking. Future evaluations of AI-mediated educational systems should report, alongside engagement and completion metrics, the rate at which the system’s outputs are aligned with the course’s stated instructional method.

4. Conclusions

4.1 Summary of Findings

The prevailing narrative within the creator economy holds that a Custom GPT built on uploaded course materials is functionally equivalent to a system engineered exclusively around a creator's proprietary methodology. The evidence reviewed here does not support that equivalence. Three converging lines of peer-reviewed research — context-memory conflict in retrieval-augmented generation (Kortukov et al., 2024; Gao et al., 2025; Joren et al., 2024; Wood et al., 2024; Hicks & Sinha, 2025), domain-specialization performance gaps (Wu et al., 2023; Singhal et al., 2023; Wang et al., 2024), and the centrality of instructional alignment to tutoring effectiveness (Kulik & Fletcher, 2016; VanLehn, 2011) — indicate that generic and methodology-aligned AI tutors produce systematically different outputs in ways that matter for learner outcomes.

A Custom GPT, by architectural necessity, will surface what its public training corpus collectively believes about a topic, modulated incompletely by the retrieved fragments of the creator's content. A system whose generative behavior is bounded exclusively by the creator's codified methodology surfaces what the creator empirically teaches. These are not two configurations of the same product; they are two different epistemic objects. In the creator economy, where the unit of value is the codified expertise of an individual practitioner rather than a generic body of public knowledge, the difference is consequential. The creator's proprietary method is, under these conditions, a defensible competitive moat against commoditized AI — but only when the implementation preserves the methodological specificity that constitutes the moat.

4.2 Limitations

This analysis is subject to limitations inherent in narrative literature reviews. The body of evidence cited is drawn from rapidly evolving research streams in which findings may be revised within twelve to eighteen months. Direct empirical comparisons between Custom GPT implementations and methodology-aligned systems on the specific outcome of course completion in independent online courses do not yet exist in the peer-reviewed literature; the present argument is therefore a synthesis of adjacent evidence rather than a report of direct experimental data. Hallucination-rate estimates drawn from legal, medical, and public-health domains may not transfer cleanly to educational tutoring contexts, where the consequences of error differ in kind. Industry market data, while sourced from established consultancies, exhibits non-trivial variance across providers and should be treated as approximate. The proposed *CursoVivo* framework has not been validated through controlled experimental studies, and the comparative effect sizes reported in the literature on intelligent tutoring systems (Kulik & Fletcher, 2016; VanLehn, 2011) cannot be assumed to transfer directly to the specific configuration described.

4.3 Future Research Directions

Three directions for future research are indicated. First, controlled comparative studies measuring completion rates, instructional alignment, and learner satisfaction across Custom GPT and methodology-aligned implementations in matched course contexts would provide the direct evidence that this synthesis cannot supply. Second, longitudinal evaluation of methodology-aligned implementations in Latin American and Spanish-language creator markets — currently underrepresented in the global ed-tech research literature

— would test the generalizability of findings derived primarily from English-language settings. Third, the broader question of intellectual property as a strategic asset in the foundation-model era invites systematic empirical investigation. The argument advanced here is that proprietary expertise functions as a moat against commoditized AI in proportion to the idiosyncrasy of the underlying method; this proposition is testable through cross-domain comparative studies and merits attention from researchers in technology strategy, instructional design, and the economics of the creator economy.

Resumen en Español

La amplia disponibilidad de modelos de lenguaje personalizables — particularmente la herramienta Custom GPT de OpenAI, con más de tres millones de asistentes creados desde 2023 — ha producido un supuesto generalizado entre creadores independientes de cursos: que un asistente gratuito o de bajo costo entrenado con los materiales del curso es funcionalmente equivalente a un sistema diseñado exclusivamente alrededor de la metodología propietaria del creador. Este artículo cuestiona dicho supuesto. A partir de evidencia revisada por pares en tres líneas de investigación convergentes — el conflicto entre memoria paramétrica y contexto en sistemas de generación aumentada por recuperación, las diferencias de rendimiento entre modelos de dominio específico y modelos generalistas, y el alineamiento instructivo como mediador principal de la efectividad de los sistemas de tutoría inteligente — el análisis demuestra que los tutores de IA genéricos y los tutores alineados a una metodología no son dos implementaciones del mismo producto sino dos objetos epistémicos distintos. Tasas documentadas de alucinación de 45% al 58% en tareas de conocimiento profesional, sesgo paramétrico persistente incluso cuando se proporciona el contexto correcto, y tamaños de efecto meta-analíticos que dependen críticamente del alineamiento entre el sistema y los objetivos del curso, indican en conjunto que la elección arquitectónica es consecuente para los resultados de aprendizaje. El artículo introduce el modelo de implementación CursoVivo — un diseño basado en restricciones donde el comportamiento generativo de la IA está acotado exclusivamente por la metodología codificada del creador — como un marco que operacionaliza la pericia propietaria en una ventaja competitiva defendible dentro de la economía de creadores.

Palabras clave: IA personalizada en educación, metodología propietaria con IA, ChatGPT vs IA especializada, tecnología para creadores de cursos, ventaja competitiva en IA, propiedad intelectual e inteligencia artificial, generación aumentada por recuperación, fine-tuning, alucinación de modelos de lenguaje en educación, CursoVivo

References

- Coachvox AI. (2026). *Pricing*. Retrieved May 2026 from <https://coachvox.ai/>
- Dahl, M., Magesh, V., Suzgun, M., & Ho, D. E. (2024). Large legal fictions: Profiling legal hallucinations in large language models. *Journal of Legal Analysis*. arXiv:2401.01301. <https://arxiv.org/abs/2401.01301>
- Festinger, L. (1957). *A theory of cognitive dissonance*. Stanford University Press.
- Fortune Business Insights. (2025). *E-learning services market size, share and industry analysis, 2025–2032*.

- <https://www.fortunebusinessinsights.com/industry-reports/e-learning-services-market-100757>
- Future Market Insights. (2025). *EduTech industry analysis in Latin America: Global market analysis report 2025–2035*. <https://www.futuremarketinsights.com/reports/edutech-industry-analysis-in-latin-america>
- Gao, L., Liu, Z., Zhao, X., Li, Y., Yu, P. S., & Wang, S. (2025). Probing latent knowledge conflict for faithful retrieval-augmented generation. arXiv:2510.12460. <https://arxiv.org/abs/2510.12460>
- Han, T., Adams, L. C., Bressemer, K. K., Truhn, D., & Lett, E. (2025). MEGA-RAG: A retrieval-augmented generation framework with multi-evidence guided answer refinement for mitigating hallucinations of LLMs in public health. *Frontiers / NIH PMC12540348*. <https://pmc.ncbi.nlm.nih.gov/articles/PMC12540348/>
- Hicks, R., & Sinha, D. (2025). The semantic illusion: Certified limits of embedding-based hallucination detection in RAG systems. arXiv:2512.15068. <https://arxiv.org/abs/2512.15068>
- Hotmart. (2024, March). *Hotmart Company, home to Teachable, announces record-breaking \$10 billion in global creator earnings* [Press release]. <https://press.hotmart.com/hotmart-company-announces-record-breaking-10-billion-in-global-creator-earnings>
- Jordan, K. (2015). Massive open online course completion rates revisited: Assessment, length and attrition. *International Review of Research in Open and Distributed Learning*, 16(3), 341–358. <https://doi.org/10.19173/irrodl.v16i3.2112>
- Joren, H., Zhang, J., Ferng, C.-S., Veit, A., & Cheng, F. (2024). Sufficient context: A new lens on retrieval augmented generation systems. arXiv:2411.06037. <https://arxiv.org/abs/2411.06037>
- Kajabi. (2026). *Pricing*. Retrieved May 2026 from <https://www.kajabi.com/pricing>
- Kortukov, E., Rubinstein, A., Nguyen, E., & Oh, S. J. (2024). Studying large language model behaviors under context-memory conflicts with real documents. arXiv:2404.16032. <https://arxiv.org/abs/2404.16032>
- Kulik, J. A., & Fletcher, J. D. (2016). Effectiveness of intelligent tutoring systems: A meta-analytic review. *Review of Educational Research*, 86(1), 42–78. <https://doi.org/10.3102/0034654315581420>
- Lee, Y.-H., Chen, M., & Liu, F. (2024). My tutor is an AI: The effects of involvement and tutor type on perceived quality, perceived credibility, and use intention. University of Florida College of Journalism and Communications. <https://www.jou.ufl.edu/insights/my-tutor-is-an-ai-the-effects-of-involvement-and-tutor-type-on-perceived-quality-perceived-credibility-and-use-intention/>
- Limón, M. (2001). On the cognitive conflict as an instructional strategy for conceptual change: A critical appraisal. *Learning and Instruction*, 11(4–5), 357–380. [https://doi.org/10.1016/S0959-4752\(00\)00037-2](https://doi.org/10.1016/S0959-4752(00)00037-2)
- Market Data Forecast. (2025). *Latin America e-learning market size, share, growth and forecast, 2025–2033*. <https://www.marketdataforecast.com/market-reports/latin-america-e-learning-market>
- Nori, H., Lee, Y. T., Zhang, S., Carignan, D., Edgar, R., Fusi, N., King, N., Larson, J., Li, Y., Liu, W., Luo, R., McKinney, S. M., Ness, R. O., Poon, H., Qin, T., Usuyama, N., White, C., & Horvitz,

- E. (2023). Can generalist foundation models outcompete special-purpose tuning? Case study in medicine. arXiv:2311.16452. <https://arxiv.org/abs/2311.16452>
- OpenAI. (2025). *What is ChatGPT Plus? Custom GPT access requirements*. OpenAI Help Center. <https://help.openai.com/en/articles/6950777-what-is-chatgpt-plus>
- Reich, J., & Ruipérez-Valiente, J. A. (2019). The MOOC pivot. *Science*, 363(6423), 130–131. <https://doi.org/10.1126/science.aav7958>
- SEO.AI. (2025). *GPT Store statistics and facts: Inside OpenAI's marketplace of three million custom GPTs*. <https://seo.ai/blog/gpt-store-statistics-facts>
- Singhal, K., Tu, T., Gottweis, J., Sayres, R., Wulczyn, E., Hou, L., Clark, K., Pfohl, S., Cole-Lewis, H., Neal, D., Schaekermann, M., Wang, A., Amin, M., Lachgar, S., Mansfield, P., Prakash, S., Green, B., Dominowska, E., Aguera y Arcas, B., . . . Natarajan, V. (2023). Towards expert-level medical question answering with large language models. arXiv:2305.09617. <https://arxiv.org/abs/2305.09617>
- VanLehn, K. (2011). The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems. *Educational Psychologist*, 46(4), 197–221. <https://doi.org/10.1080/00461520.2011.611369>
- Wang, X., Sanders, H. M., Liu, Y., Seang, K., Tran, B. X., Atanasov, A. G., Qiu, Y., Tang, S., Car, J., Wang, Y. X., Wong, T. Y., Tham, Y.-C., & Chung, K. C. (2024). Evaluating and enhancing large language models' performance in domain-specific medicine: Development and usability study with DocOA. *Journal of Medical Internet Research*. <https://pmc.ncbi.nlm.nih.gov/articles/PMC11301122/>
- Wood, A., Rogers, A., & Korhonen, A. (2024). RAGged edges: The double-edged sword of retrieval-augmented chatbots. arXiv:2403.01193. <https://arxiv.org/abs/2403.01193>
- Wu, S., Irsoy, O., Lu, S., Dabrovolski, V., Dredze, M., Gehrmann, S., Kambadur, P., Rosenberg, D., & Mann, G. (2023). BloombergGPT: A large language model for finance. arXiv:2303.17564. <https://arxiv.org/abs/2303.17564>
- Yao, J.-Y., Ning, K.-P., Liu, Z.-H., Ning, M.-N., & Yuan, L. (2024). The Earth is flat? Unveiling factual errors in large language models. arXiv:2401.00761. <https://arxiv.org/abs/2401.00761>