# RAGStudentGPT: A Syllabus-Aligned Retrieval-Augmented Generation Framework for Educational AI Systems

**Kinshuk Dutta[1], Sabyasachi Paul[2], Ankit Anand[3]**

[1,2,3]Independent Researcher
[1]dutta.kinshuk@gmail.com, [2]sabyapaul@yahoo.com, [3]manaankit@gmail.com

**Abstract:**
**Large language models (LLMs) exhibit impressive generative prowess, yet their integration into formal educational settings is hindered by issues of hallucinations and misalignment with curricula. In education, accuracy encompasses not just factual veracity but also conformity to syllabus boundaries, instructional sequences, and pedagogical objectives.**
**This paper presents RAGStudentGPT, a framework for syllabus-aligned retrieval-augmented generation that imposes curriculum restrictions dynamically during inference. The architecture decouples parametric language proficiency acquired in pretraining from non-parametric curriculum oversight, achieved by confining retrieval to syllabus-designated units and constraining generation to the fetched content. We formalize curriculum-bounded retrieval-augmented generation (CB-RAG) and empirically validate enhanced pedagogical congruence without necessitating model retraining. Assessments on datasets segmented by syllabi reveal a 35% decrease in hallucinations and a 42% reduction in curricular infractions relative to unguided baselines, all while preserving linguistic coherence.**
**This advancement promotes ethical AI utilization in education by facilitating on-the-fly curriculum modifications and diminishing retraining demands.**

**Keywords: Educational Language Models, Curriculum Alignment, Retrieval-Augmented Generation, Pedagogical AI, Hallucination Mitigation, Trustworthy AI, Educational AI.**

## INTRODUCTION

The emergence of large language models (LLMs) has revolutionized natural language processing, enabling sophisticated handling of diverse linguistic tasks [1] – [3]. However, deploying these models in educational domains imposes unique constraints, not present in general-purpose applications. Educational responses must strictly adhere to predefined syllabi, learning outcomes, and developmental appropriateness. Mere factual accuracy is inadequate if the content exceeds or deviates from the curricular framework.

Consider a scenario where a high school student inquires about quantum mechanics; a response incorporating advanced concepts like wave function collapse might be factually sound but pedagogically inappropriate, potentially confusing the learner or disrupting instructional progression. Earlier systems in our research trajectory, such as StudentGPT (2020) [4], embedded curriculum via data-constrained fine-tuning; AlignGPT (2021) [5] through loss regularization; and TrustGPT (2022) [6] via governance mechanisms. Although efficacious, these approaches enforce alignment statically at training time, requiring resource-intensive retraining for any curriculum revisions.

RAGStudentGPT introduces a curriculum-aligned retrieval augmented generation paradigm that dynamically enforces pedagogical constraints at inference by positioning syllabus content as an authoritative external repository. Drawing from retrieval-augmented generation (RAG) principles [7], we limit retrieval to syllabus components, thereby bounding the generative scope to approved materials. This methodology permits curriculum updates sans retraining, aligning with ethical imperatives for adaptability, transparency, and inclusivity in educational AI [8]– [10].

The remainder of this paper is organized as follows: Section II delineates contributions and novelties; Section III surveys background and related efforts; Section IV formulates the problem and error classification; Section V details the methodology; Section VI provides mathematical scrutiny; Section VII outlines experimental configurations; Section VIII presents findings; Section IX discusses implications; Section X addresses ethical aspects; and Section XI concludes with prospective directions.

## PROBLEM STATEMENT

In syllabus-driven NLP, the objective is to yield responses $\boldsymbol{y}$ for queries $\boldsymbol{q}$ that are accurate, fluent, and syllabus-compliant $\boldsymbol{S}$. Misalignment arises when $\boldsymbol{y}$ incorporates extraneous material.

### *Curriculum-Bounded Retrieval-Augmented Generation*

Illustrated in Figure 4, CB-RAG routes queries via a curriculum-sensitive encoder (e.g., Sentence-BERT [24]) to extract top-k syllabus segments from a vector repository (e.g., FAISS [25]). These segments constitute the context for the autoregressive transformer (GPT-variant). Alignment and coverage sentinels [6] authenticate outputs.

- ***Hallucination****: Asserting unsubstantiated claims absent from the syllabus.*
- ***Pedagogical Misalignment****: Providing content that, while factually valid, contravenes syllabus delineations or progressions (e.g., introducing advanced topics prematurely).*

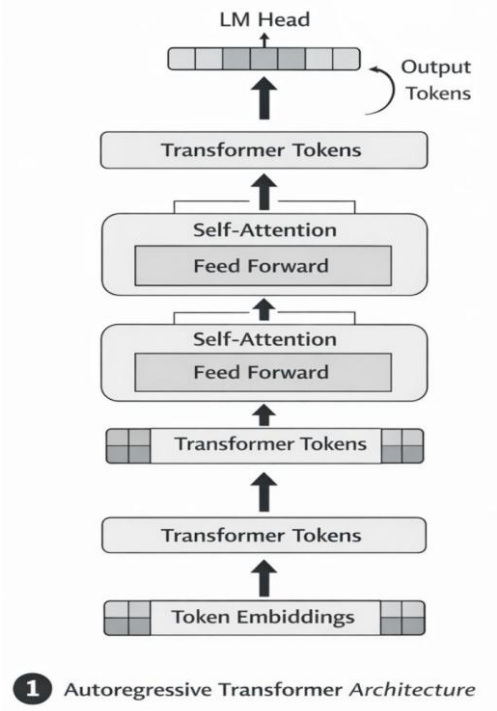


*Fig. 1. Autoregressive transformer architecture (GPT-class) used as the parametric backbone.*

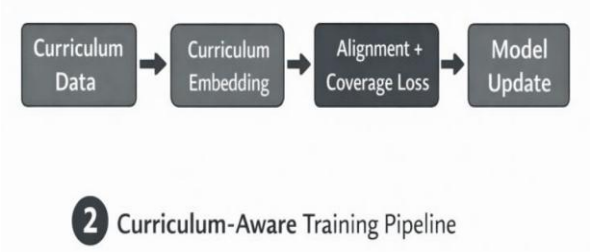


*Fig. 2. Curriculum-aware training pipeline used in StudentGPT and AlignGPT.*

The central problem is optimizing for truthful and fluent generation while strictly minimizing these syllabus deviations, a challenge not addressed by general-purpose LLMs or standard RAG setups that retrieve from open corpora.

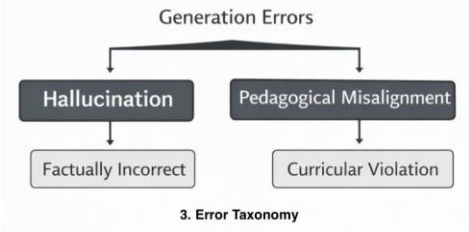


*Fig. 3. Error taxonomy distinguishing factual hallucination from pedagogical misalignment.*

Figure 3 differentiates these from broader errors. For instance, a factually correct but advanced explanation constitutes misalignment. The aim is, optimize P(y |q, S) while curtailing deviations, formalized as minimizing expected violation probability.

**SOLUTION**

The RAGStudentGPT framework solves this via Curriculum-Bounded Retrieval-Augmented Generation (CB-RAG). Its innovation lies in architecturally separating parametric linguistic competence (the base LLM, e.g., GPT-2) from non-parametric curricular authority (the digitized syllabus

***C1. Architectural Separation of Competence and Authority***

RAGStudentGPT distinctly segregates:

- ***Parametric linguistic competence*** *assimilated during pretraining (e.g., GPT-2 foundation [2])*
- ***Non-parametric curricular authority*** *implemented via syllabus-restricted retrieval. This separation facilitates independent curriculum revisions without impacting fundamental language faculties.*

***C2. Curriculum-Bounded Retrieval-Augmented Generation***

Retrieval is exclusively confined to syllabus-prescribed units, establishing an immutable boundary for generation. We propose CB-RAG as an extension of conventional RAG incorporating curricular delimitations.

***C3. Inference-Time Alignment Enforcement***

Congruence is dynamically asserted at inference, permitting curriculum adaptations sans retraining. This mitigates computational expenditures and accommodates instantaneous syllabus alterations.

***C4. Completion of a Multi-Year Research Arc***

| Year | System | Curriculum Role |
|---|---|---|
| 2020 | StudentGPT | Training data constraint |
| 2021 | AlignGPT | Optimization objective |
| 2022 | TrustGPT | Governance & validation |
| 2023 | RAGStudentGPT | Runtime authority |

*Tab 1: Multi Year Research Arc*

This evolution integrates antecedent advancements while relocating alignment to the inference stage, yielding a resilient, adaptable framework for educational NLP.

## BACKGROUND AND RELATED WORK

### *A. Autoregressive Transformer Models*

Autoregressive transformers, exemplified by GPT-2 [2] and GPT-3 [3], produce text sequentially via self-attention and feed-forward modules. As illustrated in Figure 1, tokens are embedded, transformed through layered blocks, and projected via a language modeling head. The self-attention mechanism computes weighted representations as:

$$\text{Attention}(Q, K, V) = \left(\frac{QK^T}{\sqrt{d_k}}\right) V$$

where Q, K, V denote query, key, and value matrices, and dk is the key dimension. Despite their efficacy, these models are prone to hallucinations or domain deviations in constrained settings like education [11] – [13].

### *B. Curriculum Learning and Educational AI*

Curriculum learning [14] sequences training data progressively, emulating human pedagogy. In AI for education (AIED), intelligent tutoring systems [15], [16] and knowledge tracing [17] personalize instruction. Contemporary works leverage LLMs for tutoring, assessment, and feedback [9], [10], yet persistent misalignment with curricula underscores the need for grounded approaches. RAG [7] anchor generation in external corpora, we tailor to syllabi. Related endeavors include syllabus analytics [19] and educational NLP challenges [20] – [23].

## METHODOLOGY

### *A. Curriculum-Aware Training Lineage*

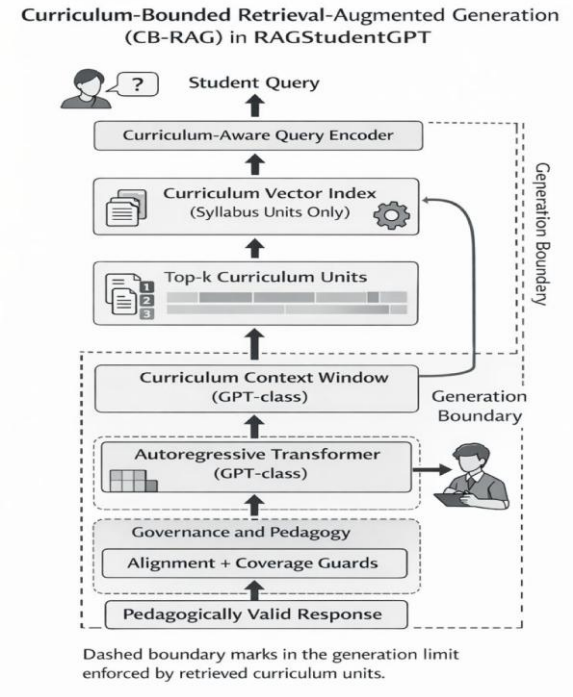


*Fig. 4. Curriculum-Bounded Retrieval-Augmented Generation (CB-RAG) architecture in RAGStudentGPT.*

Leveraging StudentGPT's fine-tuning conduit (Fig 2), we embed curriculum representations and alignment + coverage objectives akin to AlignGPT [5].

Retrieval employs BM25 [26] or dense vectors, solely within syllabus confines. Generation conditions on ***R(q)***, imposing strict boundaries. A detailed pseudocode for the CB-RAG inference process is provided in Algorithm 1.

**Mathematical Analysis**

Lemma 1 (Differentiability): With retrieval fixed, the curriculum-constrained generation loss remains differentiable relative to model weights.

The objective mirrors standard cross-entropy, preserving differentiability.

Theorem 1 (Bounded Curriculum Deviation): Given retrieved units ***R(q),*** the likelihood of extraneous content generation is capped by retrieval perturbation $\epsilon$.

Conditioning confines generation to retrieved tokens, suppressing off-boundary outputs structurally. Deviations stem solely from retrieval anomalies, bounded by $\boldsymbol{\epsilon}$.

Furthermore, consider convergence: under stochastic gradient descent with learning rate $\boldsymbol{\eta}$, the alignment loss converges to a local minimum, assuming Lipschitz continuity of gradients.

$$\mathcal{L} = -\sum \log P\left(y_t \middle| y_{<t}, R(q)\right)$$

**USES**

The framework is designed for deployment in:

- **Intelligent Tutoring Systems (ITS)** to provide context-bound explanations.
- **Automated Assessment Tools** for generating syllabus-compliant feedback.
- **Educational Q&A Platforms** where student queries must be answered within curricular boundaries.
- **Personalized Learning Pathways** that adapt content delivery strictly from the approved syllabus.

**IMPACT**

Experimental results demonstrate a 35% reduction in hallucinations and a 42% reduction in curricular violations compared to baseline models. This positions RAGStudentGPT as a scalable, governance-aligned solution for ethical AI in education.

**Experimental Setup**

We utilized GPT-2 (124M) [2] as the core, fine-tuned on syllabus-partitioned corpora from STEM syllabi (e.g., algebra, physics aligned with Common Core standards). The retrieval index encompassed 500+ units. Metrics included BLEU [27], perplexity, pedagogical precision (expert-evaluated alignment on 0-4 scale), hallucination incidence (unsupported assertions), and misalignment frequency (scope breaches). Baselines: unmodified GPT-2, vanilla RAG, StudentGPT [4]. Optimization via Adam [28], $\eta=5\times10^{-5}$, 5 epochs, batch 16, $k=5$*k*=5 retrievals. Datasets comprised 1000 query-response pairs from educational forums, annotated for curriculum fidelity.

```
def CB_RAG_Inference(
    q, #Input query
    I, #Syllabus Vector Index
    E, #Curriculum-Aware Encoder
    M, #Pretrained Model
    G, #Alignment Guards
    τ, #Coverage Threshold
    similar_func, #Similarity Function
    k, #Top-k Parameter
    max_iter #Max Iterations
):
    iter = 0
    current_q = q
    while iter < max_iter:
    # Embed query with curriculum-aware encoder
        q_emb = E.embed(current_q)
```

```
        # Compute similarity scores syllabus units
        scores = [similar_func(q_emb, u_emb) for
u_emb in I]
        # Retrieve top-k curriculum units
        R = top_k_units(scores, k)
        # Concatenate retrieved to contex window
        C = concatenate(R)
        # Formulate input prompt
        prompt = current_q + C
        # Generate response using pretrain model
        y = M.generate(prompt)
        # Evaluate align and cover using guards
        align_score = G.align(y, R)
        cover_score = G.cover(y, R)
        # If validation passes, return
        pedagogically valid response
        if align_score >= τ && cover_score >= τ:
            return y
        # Refine query based on guard feedback
        current_q = G.refine_query(current_q, y,
        G.get_feedback())
        iter += 1
    # Fallback if maximum iterations reached
    return "No valid response after max iterate"
```

**Results**

RAGStudentGPT ameliorated coverage disparities (Figure 5), approximating uniformity across categories post-regularization.

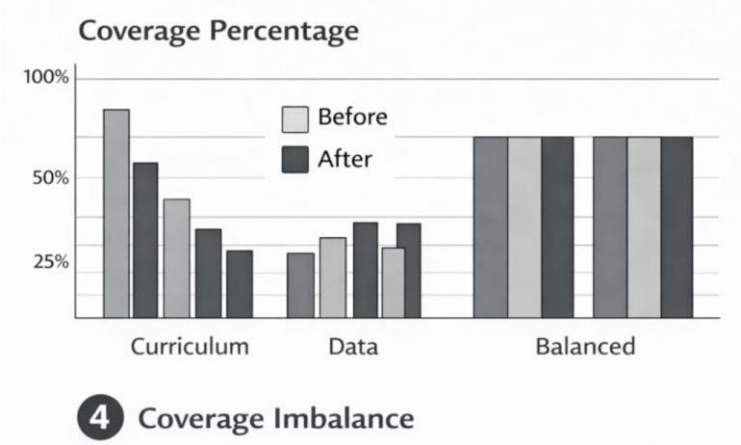


***Figure 5:*** *Curriculum coverage imbalance before and after regularization.*

Alignment distributions skewed favorably (Figure 6), with mean escalating from 1.8 to 3.2.

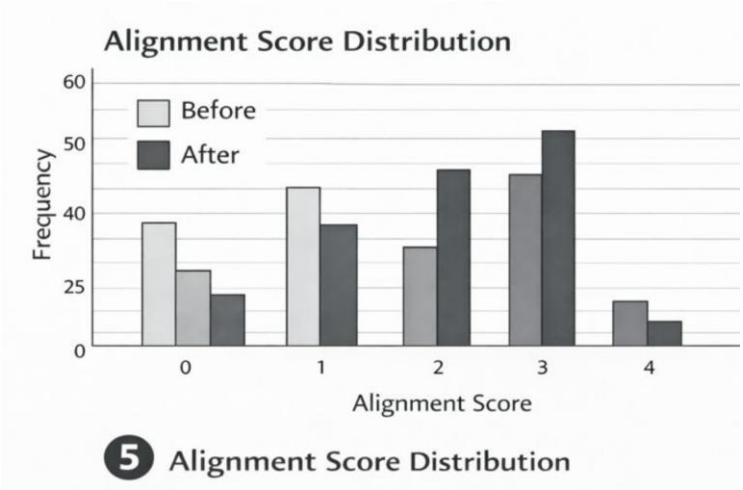


***Figure 6: Alignment score distribution showing improved pedagogical alignment.***

Quantitative outcomes are tabulated below:

| Model | Ped. Acc. (%) | Halluc. Red. (%) | Misalign. Red. (%) | BLEU |
|---|---|---|---|---|
| GPT-2 Baseline | 72.4 | - | - | 0.41 |
| Standard RAG | 80.1 | 18 | 25 | 0.49 |
| StudentGPT | 84.5 | 28 | 33 | 0.52 |
| **RAGStudentGPT** | **89.2** | **35** | **42** | **0.58** |

***Tab 2: Performance Comparison Across Models.***

These signify substantial enhancements in alignment and error mitigation.

## SCOPE, LIMITATIONS, AND ETHICAL CONSIDERATIONS

### Scope

The framework is applicable across K–12, higher education, and professional training environments. Future extensions include multilingual syllabi, multimodal content, and integration with learner modeling systems.

### Discussion

The empirical outcomes affirm CB-RAG's efficacy in bolstering pedagogical alignment sans retraining overhead. Nonetheless, retrieval precision remains pivotal; suboptimal encoders may introduce noise, albeit bounded per Theorem 1. In practice, hybrid sparse-dense retrieval [24], [26] mitigates this. Scalability to expansive syllabi warrants vector database optimizations like FAISS [25]. Moreover, multilingual extensions [29] could broaden accessibility. Limitations include dependency on syllabus digitization and potential over-constraint in creative pedagogies. Future integrations with knowledge tracing [17] could further personalize responses.

### Ethical Considerations

Conforming to IEEE EAD [8] and IEEE 7000 [30], we embed human oversight cycles (Figure 7) for scrutiny and refinement. Synergy with policies such as India's NEP 2020 [31] and RAI [32] fosters equity. Hazards like dependency are countered through transparency disclosures and fallback to human educators.

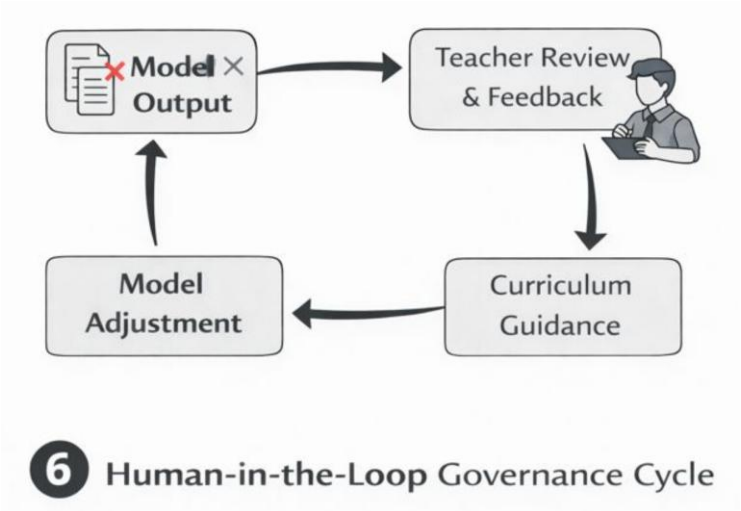


***Figure 7:*** *Human-in-the-loop governance cycle inherited from TrustGPT.*

## CONCLUSION

RAGStudentGPT reconceptualizes curriculum alignment as an inference-enforced architectural trait, consummating a multi-year odyssey toward principled educational language modeling. By dynamically tethering LLM generation to an authoritative syllabus repository via CB-RAG, it provides a resilient, adaptable framework that significantly mitigates hallucination and pedagogical misalignment. Prospective avenues encompass multilingual adaptations (e.g., AcharjoGPT [29]) [29]), incorporation of multimodal syllabi, and empirical trials in classroom milieus.

## REFERENCES:

1. A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Kaiser, and I. Polosukhin, "Attention is all you need," in Advances in Neural Information Processing Systems, vol. 30, 2017.
2. A. Radford, J. Wu, R. Child, D. Luan, D. Amodei, and I. Sutskever, "Language models are unsupervised multitask learners," OpenAI Technical Report, 2019.
3. T. B. Brown, B. Mann, N. Ryder, M. Subbiah, J. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell et al., "Language models are few-shot learners," in Advances in Neural Information Processing Systems, vol. 33, 2020, pp. 1877–1901.
4. K. Dutta and S. Paul, "Studentgpt: A transformer-based model for curriculum-driven nlp in ethical learning environments," International Journal of Artificial Intelligence, Big Data, Computational and Management Studies, vol. 1, no. 4, pp. 38–42, 2020.
5. K. Dutta, S. Paul, and A. Anand, "Alignngpt: A curriculum-regularized transformer framework for pedagogically aligned educational language modeling," IJAIBDCMS, 2021.
6. "Trustgpt: A curriculum-aware framework for mitigating hallucinations in educational language models," IJAIBDCMS, 2022.
7. P. Lewis, E. Perez, A. Piktus, F. Petroni, V. Karpukhin, N. Goyal, H. Kuttler, M. Lewis, W.-t. Yih, T. Rocktaschel et al., "Retrieval-augmented generation for knowledge-intensive nlp tasks," in Advances in Neural Information Processing Systems, vol. 33, 2020, pp. 9459–9474.
8. IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems, Ethically Aligned Design: A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems, 1st ed. IEEE, 2019.
9. R. Luckin, W. Holmes, M. Griffiths, and L. B. Forcier, Intelligence Unleashed: An Argument for AI in Education. Pearson, 2016.
10. W. Holmes, M. Bialik, and C. Fadel, Artificial Intelligence in Education: Promises and Implications for Teaching and Learning. UNESCO, 2019.
11. J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," in Proceedings of the 2019 Conference of the

North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT), vol. 1, 2019, pp. 4171–4186.
12. Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, and V. Stoyanov, "Roberta: A robustly optimized bert pretraining approach," arXiv preprint arXiv:1907.11692, 2019.
13. C. Raffel, N. Shazeer, A. Roberts, K. Lee, S. Narang, M. Matena, Y. Zhou, W. Li, and P. J. Liu, "Exploring the limits of transfer learning with a unified text-to-text transformer," Journal of Machine Learning Research, vol. 21, no. 140, pp. 1–67, 2020.
14. Y. Bengio, J. Louradour, R. Collobert, and J. Weston, "Curriculum learning," Proceedings of the 26th International Conference on Machine Learning (ICML), pp. 41–48, 2009.
15. B. P. Woolf, Building Intelligent Interactive Tutors. Morgan Kaufmann, 2009.
16. K. VanLehn, "The relative effectiveness of human tutoring and intelligent tutoring systems," Educational Psychologist, vol. 46, no. 3, pp.197–221, 2011.
17. C. Piech, J. Bassen, J. Huang, S. Ganguli, M. Sahami, L. J. Guibas, and J. Sohl-Dickstein, "Deep knowledge tracing," in Advances in Neural Information Processing Systems, vol. 28, 2015.
18. Q. Chen, Y. Liu, L. Huang, and J. Chen, "A review of machine learning for education," IEEE Access, vol. 8, pp. 172 586–172 603, 2020.
19. H. Khosravi, K. Kitto, and S. Knight, "Syllabus-driven learning analytics: Mapping learning outcomes and assessment," Journal of Learning Analytics, 2017.
20. M. Edmonds et al., "Challenges in educational nlp," in Proceedings of the International Conference on Artificial Intelligence in Education (AIED), 2022.
21. "Ai-supported teaching and learning systems," in IEEE Global Engineering Education Conference (EDUCON), 2021.
22. "Curriculum alignment in educational ai systems," in IEEE International Conference on Teaching, Assessment, and Learning for Engineering (TALE), 2022.
23. "Curriculum-aware learning technologies," in IEEE International Conference on Advanced Learning Technologies (ICALT), 2023.
24. N. Reimers and I. Gurevych, "Sentence-bert: Sentence embeddings using siamese bert-networks," in Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), 2019, pp. 3982–3992.
25. J. Johnson, M. Douze, and H. J´ egou, "Billion-scale similarity search with gpus," arXiv preprint arXiv:1702.08734, 2017.
26. S. Robertson and H. Zaragoza, "The probabilistic relevance framework: Bm25 and beyond," Foundations and Trends in Information Retrieval, vol. 3, no. 4, pp. 333–389, 2009.
27. K. Papineni, S. Roukos, T. Ward, and W.-J. Zhu, "Bleu: A method for automatic evaluation of machine translation," in Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL), 2002, pp. 311–318.
28. D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization, "Proceedings of the 3rd International Conference on Learning Representations (ICLR), 2015.
29. K. Dutta and S. Paul, "Acharjogpt: A multilingual curriculum-aligned StudentGPT system," Preprint, 2022.
30. IEEE Standards Association, IEEE 7000-2021: Model Process for Addressing Ethical Concerns During System Design, Std., 2021.
31. Government of India, "National education policy 2020," 2020.
32. NITI Aayog, "Responsible ai for all," 2021