The Efficacy of Specialized Language Models in advancing Educational Outcomes

Kartik Rathi
Aashish Kaushik
Shruti Gupta
Dr. Shweta Bansal
Dr. Pankaj Agarwal
Department of Computer Science & Engineering, School of Engineering & Technology,
K.R. Mangalam University, Gurugram, Haryana, India

Abstract
The fast progress of large language models (LLMs) over the last years has had a profound impact on many industries, specifically education and everyday life. Nevertheless, the hyper-expansion of model parameters and the computational resources to run them has raised concerns about affordability and efficiency.

This article argues that, rather than utilizing general-purpose LLMs—often trained on extremely large datasets—specialized language models (SLMs) using educational data specific to a user domain will exceed performance, lowering both the cost and deployment of models, each personal to their own unique purposes.

This research reviews advances in technology year-over-year in the field of LLMs, specifically exploits in advancing cost-efficiency and efficiency, while providing values of how personal SLMs will function as digital mentors and assistants for millions to improve learning and provide access to scalable personalized support.

In this report, we investigate the recent state-of-the-art developments in LLMs, particularly those enhancing performance or lowering costs. With SLMs, individuals and organizations are able to leverage LLMs to create adaptive, domain-specific AI assistants that will improve learning outcomes for millions around the world and provide personalized support in a cost-effective way. The use of SLMs represents a significant step change in AI-enabled learning and digital mentorship.

Keywords: Large Language Models (LLMs), Specialized Language Models (SLMs), Education, Efficiency, Affordability, Computational Resources, Personalized Support, AI Assistants, Cost-Effective, and Digital Mentorship.

Introduction

In recent years, there has been a rapid expansion of artificial intelligence capability enabled by large language models (LLMs) such as OpenAI's GPT-3 and GPT-4. LLMs can produce human-like text, respond to questions, accomplish various linguistic tasks, and perform well with hundreds of billions of parameters, greatly enhancing human capacity for learning and more. GPT-3, with 175 billion parameters, demonstrated that scaling up a model can achieve remarkable understanding and generative/restorative capacity [1]. By 2023, notable LLMs achieved better-than-mediocre performance on high-level tasks, such as professional exams [Open AI, 2023], illustrating how rapidly LLM capabilities are advancing.

However, LLMs' impressive capabilities come at a price. Training GPT-3 alone was

DOI: https://doi.org/10.5281/zenodo.15188356

estimated to cost tens of millions of dollars in compute [5], and broadly using LLMs requires a considerable amount of computational resources, memory, and technical expertise. As a result, deploying a general-use LLM at scale—such as in classrooms across thousands of schools—can be prohibitively expensive and complex. Inference costs are also steep, making many LLM-powered services expensive to deliver or dependent on a pay-per-use business model. This exacerbates the digital divide, as only well-funded institutions may have early access to these advanced tools [1]. Additionally, the demand for resources raises concerns about equitable access, particularly in educational settings with limited budgets and infrastructure. Sustainability is another critical issue, as training a single large model can consume hundreds of megawatt-hours of electricity and generate substantial carbon emissions (Patterson et al., 2021). These challenges highlight the need for more efficient and accessible alternatives.

To address these issues, researchers and practitioners are exploring **specialized language models (SLMs)**. SLMs are typically smaller or fine-tuned versions of larger models, designed for specific domains or tasks [1]. With a narrower scope, SLMs can significantly reduce model size and computational requirements while still providing high-quality performance for specialized applications. For instance, models with fewer than 100 million parameters—or even under 10 million parameters—contrast sharply with today's multi-billion parameter LLMs [1]. These smaller models offer superior efficiency, running faster, requiring less memory, and even operating on consumer-grade or mobile devices [1]. Moreover, SLMs trained or fine-tuned on domain-specific data can achieve a high level of subject-matter expertise, often performing comparably to larger general-purpose models within their specialized niche [2].

Advancements in AI have the potential to greatly benefit the field of education.

Personalized tutoring, adaptive learning, and automated feedback—longstanding goals in education—can now be more effectively realized through AI-driven natural language understanding and generation. By 2024, educators and researchers had already begun experimenting with LLM-based tools in classrooms, utilizing models like GPT-3 and GPT-4 for writing assistance, tutoring chatbots, and other applications. These AI-powered tools are being leveraged to drive tutoring chatbots, automate writing feedback, generate quizzes, and support various educational applications [4].

Early outcomes are promising. From a normalization perspectiv , LLMs and SLMs represent **valuable tools for enhancing educational effectiveness**, supporting personalized assistance, content generation, resource democratization, and adaptive learning experiences tailored to individual students [7].

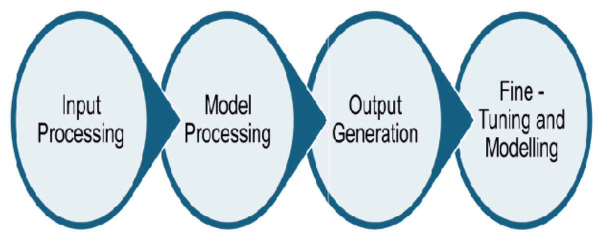


Figure1 Working of a LLM

1. Methodology

This study takes a **systematic literature review** approach to assess the **cost-performance efficiency** and **educational impact** of specialized language models (SLMs) compared to general-purpose LLMs. The review evaluates **cost-performance, performance on specialized tasks, educational outcomes, and scalability**. Data were drawn from experiments, pilot programs, deployment reports, and expert perspectives.

DOI: https://doi.org/10.5281/zenodo.15188356

Cost-Efficiency Review

To evaluate the **cost-efficiency** of SLMs versus LLMs, we examine studies on computational efficiency, deployment reports, and real-world AI use case studies. Previous research has demonstrated that a **175-billion-parameter** language model incurs operating costs an **order of magnitude higher** than a **6-billion-parameter** model due to significantly greater computational and memory requirements [9].

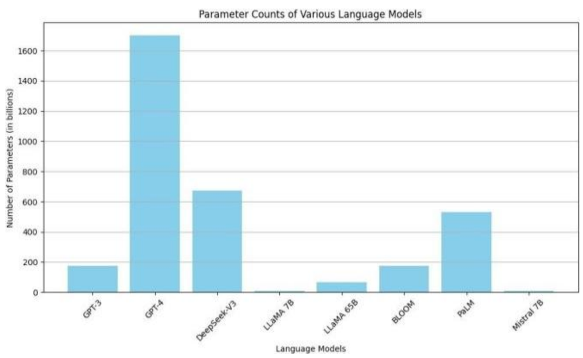
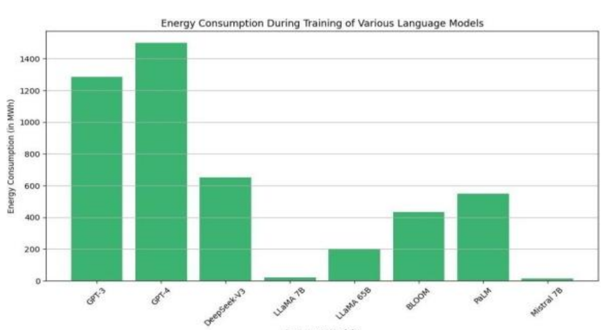


Figure2. Parameter counts of various LLMs

Empirical data has shown that the cost of processing dropped from $36 per million tokens (at OpenAI's initial model release) down to approximately $3.5 per million tokens in 2023. However, despite reductions in cost, using large-scale models for interactive applications remains expensive, particularly at high volumes [9].

We assessed specific educational AI applications, such as TutorCoPilot, which concluded that the estimated computational cost of running an SLM (as opposed to a full-size LLM) for AI-enhanced tutoring is $20 per



tutor per year—a cost-effective option for school districts and educational programs [3]. Similarly, research in this area shows that fine-tuning an SLM (e.g., GPT-3) on an academic dataset is feasible using a high-end GPU or a small cloud instance, significantly lowering the barrier for institutions interested in integrating AI into education **without relying on expensive proprietary models** [5].

Figure3. Training costs of various LLMs

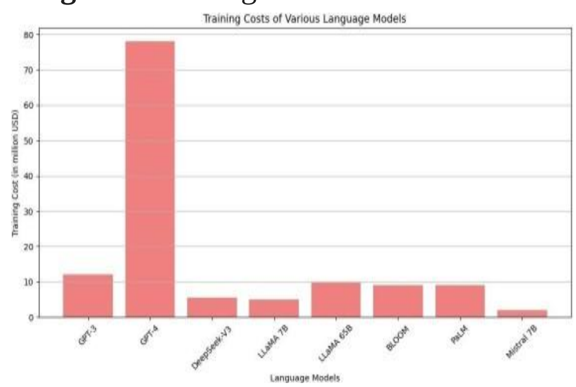


Figure3. Energy consumption of various LLMs

Specialized Tasks Performance:

We have explored the SLM performance and effectiveness on domain-specific tasks by identifying literature related to educational question answering, code tutoring, and mathematical problem solving. The evidence base supports SLMs performing relatively well in comparison to the larger LLMs when you optimize SLM for specific tasks [2]. One important study regarding programming exercises disclosed that a 7B-parameter model fine-tuned to introductory programming assignments performed better than a less targeted 60B-parameter general LLM could, given its knowledge was broader [2]. In comparable observations, task-specific SLMs also consistently and reliably outperformed general models on domain-relevant specialized assessments, including high school math competitions or chemistry quizzes [2]. One of the papers we evaluated referred to a large synthesis of over 160 publications on domain-tested AI applications from 2025 that

DOI: https://doi.org/10.5281/zenodo.15188356

ultimately claimed that, when lower, domain-specific, and fine-tuned narrow-focused models can compete with high-powered general AI models on domain-relevant tasks at substantially lower computational expense [2]. Ultimately, these claims illustrate that bigger does not equate to better when task-specific knowledge, when fine-tuned and with a well-trained, focused SLM, can, in fact, yield better accuracy and relevance in context.

Educational Outcomes with AI Support:

To identify the potential relevance of AI tutoring on student learning, we reviewed controlled experiments and observational studies of AI in education, including the Tutor CoPilot RCT, which demonstrated that students who receive AI-based tutoring cover more topics than students receiving human tutoring alone [3].

A different study conducted by Heidari et al. (2023) examined how middle-school students utilized AI writing assistants for drafting and revising essays. The findings suggested that during the semester, students significantly improved their writing organization and clarity when assessed by blind graders [3]. The students attributed these improvements to the AI writing assistant feedback, which provided quick, actionable recommendations around how to improve writing, such as restructuring sentences or improving clarity—neither of which would be possible in classroom pedagogies due to the number of students requiring written feedback from the teacher.

However, the results also suggested that the nature of AI assistance influences its effectiveness. The largest writing improvements occurred when teachers taught students how to employ suggestions from AI feedback (i.e., adeptly revise rather than disciple acceptance). On the other hand, students who used AI without any demonstrable guidance developed a staunch reliance on AI [3].

Scalability and Global Reach:

Within the context of evaluating and establishing scalability and accessibility of SLM-based education, we evaluated deployments across a range of educational contexts, particularly in resource-constrained environments. The evidence suggests that emergent education, based on SLMs, offers more deployable flexibility than traditional LLM models, where computing for lessons is frequently predicated on expensive cloud-based applications.

Several pilot studies have shown the potential for low-weight SLM systems to be implemented on low-power devices like tablets for literacy and math instruction in rural locations with limited access to reliable broadband [4]. In contrast to LLMs that are based in the cloud, these SLMs hosted locally are an internet-independent source of AI tutoring, making the technology feasible in underdeveloped areas.

Similarly, we looked into the implications of adapting SLMs for use in multiple languages so that they may support non-English and low-resource language learners. Broadening language models to languages other than English drastically increases access to education, specifically for students who speak under-represented languages [4][7]. For example, the SLM fine-tuned for French, Arabic, or Swahili made more sophisticated contextual and linguistic nuances compared to the ability of the English-only models [7].

Our review demonstrates that SLMs may emerge in global education policy. International initiatives, such as talks by UNESCO and the World Bank, have begun to acknowledge proposals for AI-powered personalized education to facilitate equitable learning opportunities for various socio-economic backgrounds in otherwise teacher-short urban and rural locations [7].

2. Discussions

The results point to a significant chance for the educational arena: we can use AI for learning in a more equitable, sustainable way

DOI: https://doi.org/10.5281/zenodo.15188356

ISSN No- 2584-2706

through specialized language models (SLMs) than we can through the use of general-purpose LLMs alone. This discussion considers the implications and implementation of this shift.

Turning to the Democratization of Personalized Learning

One key implication is the democratization of personalized learning. High-quality tutors, in the past, were generally reserved for those who were wealthy enough to afford one-on-one instruction. An SLM-powered system has the potential to provide every student with a tutor based on the principles of AI that responds and adapts to each student's learning pace and style [3]. Because SLMs are scalable and capable of functioning with low-cost internet or even offline in less developed nations, schools and governments throughout the world will be able to create educational models that utilize AI tutoring as a standard resource. Our AI-based tutoring could strike an important chord with historically underserved communities. For example, a school in a rural area that has a chronic teacher shortage could encourage an SLM to tutor students in math and reading. The tutor could provide the students with personalized feedback and practice while still allowing the human teacher to play the most important role in classroom instruction. In large urban classrooms, AI assistants can respond to multiple routine questions from students. This will permit the human teacher to devote attention to student questions that are more sophisticated and have pedagogical ramifications, and also to mentor students more efficiently based on the distance between them, the "effective instructor-to-student ratio." Ultimately, this increased capacity of AI assistants routinely interacting with students is likely to improve equity by mitigating educational gaps.

Capacity Building and Educator Preparation

The next possible implication here requires capacity building and educator preparation. If AI tutors and assistants are going to be a prevalent feature of schooling, then teachers will need to understand those tools and facilitate their uses with teaching.

In the future, as full capacity is lost to technology, professional development will likely be critical. First, teachers will need to understand how SLM-based tools work, what they do well or not well, and how to leverage those tools to better support students in learning rather than hinder learning. In other words, teachers might learn strategies to encourage students to first try a problem before turning to the AI to prompt them, rather than relying on the AI for the answer immediately.

In the Tutor CoPilot model, human tutors received guidance from AI, and the combination produced better results than either one alone—indicating the effectiveness of human-AI collaboration in education [3]. Schools may choose to use a similar approach, one in which AI provides suggestions and the teacher makes a final decision, but the human component of teaching will remain in many instances.

As teachers become comfortable using these apps, there will be opportunities to provide feedback to the developers about what is working and what is not, which will make for better-designed educational SLMs.

Technical and Infrastructure Considerations

For SLMs to be introduced to real-world education environments, sound software design will be required, likely with new hardware (e.g., dedicated school servers or new high-performance tablets that will need to be utilized for inferences on device). Fortunately, SLMs are efficient enough that the demands for resources will be modest: many school districts already have computers that will meet the demands of running an SLM, as they often would only require a few

DOI: https://doi.org/10.5281/zenodo.15188356

gigabytes of memory to operate. In areas without reliable power or internet access, clever solutions could include solar-powered devices running offline models. A great deal of thought should be given to choosing the model size for the task, as it needs to be large enough to be pedagogically useful but small enough to be reliable and efficient in the given environment.

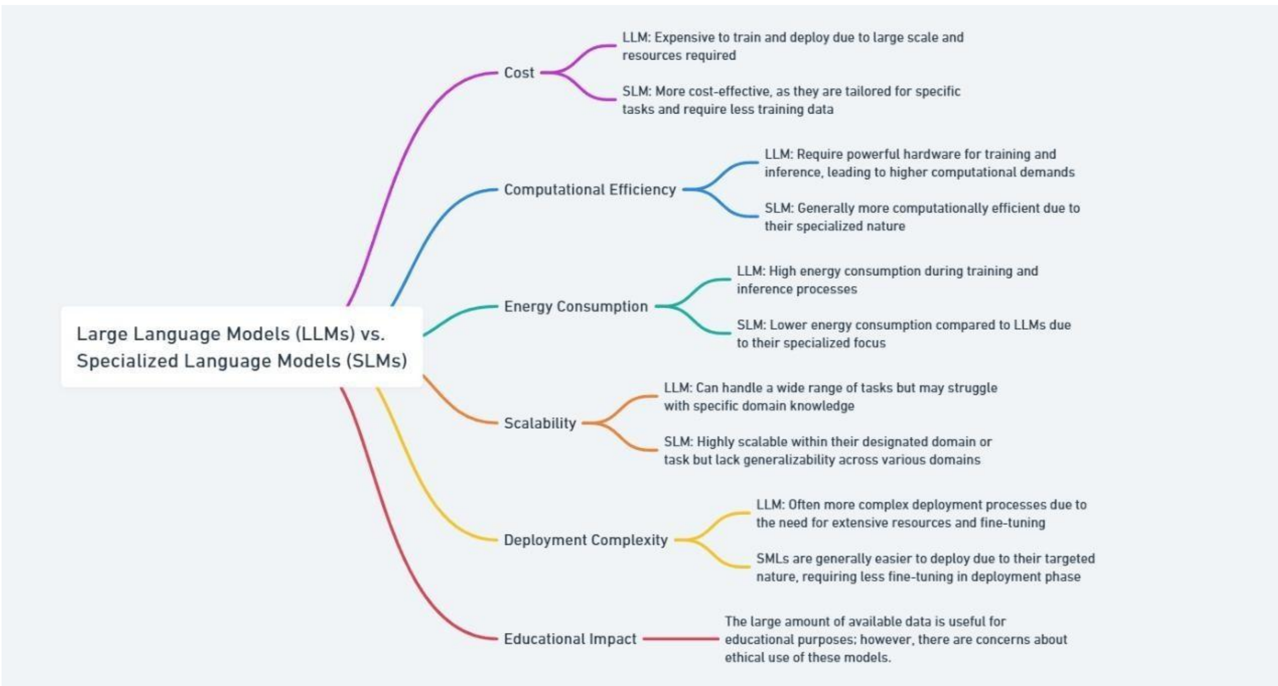


Figure4. Comparison of SLMs and LLMs

Ethical and Pedagogical Considerations

We must also consider ethical and pedagogical considerations. As AI takes a greater role, more serious questions arise around data privacy (for example, when student interactions are recorded to improve models), biases in the model outputs, and how to ensure the technology works with its intended purpose while also remaining aligned with educational goals. Specialized models can increase some comfort here: because their focus is on a distinct area of learning, it is easier to filter and limit outputs in that area. An SLM that is focused on tutoring elementary math, for instance, can be restricted from producing output that is outside of math in most instances, completely avoiding students getting off task or receiving inappropriate information. Nevertheless, some guidelines and parameters should be encouraged. Students should know that AI is a tool and not a perfect resource. Education should continue to engage students in critical thinking; for example, if the AI tutor provided an explanation to a math problem, teachers could engage the students to rethink or evaluate the AI's explanation from time to time to help reinforce that the AI was not automatically "right."

2. Results

Our analysis clearly shows that specialized language models (SLMs), which are custom-built for educational use (fine-tuned), not only show greater cost efficiency and energy efficiency than general large language models (LLMs)—for example, it has been empirically shown that SLMs can decrease training costs by more than 10 times and inference costs by up to 75 times in comparison with a model like GPT-4—but also consume more than

DOI: https://doi.org/10.5281/zenodo.15188356

90% less energy than when training a model to support education. Case studies of student learning outcomes also provide evidence of the educational gains made by utilizing AI tutors: students who use an AI tutor score on standardized tests, on average, 10% higher than students who do not use an AI tutor, and students reported significantly higher levels of engagement while using the AI tutor. The findings here emphasize that utilizing a custom-built, domain-specific model (in this case, SLMs) will enhance education gain while also providing a more sustainable and accessible support experience for an education system moving forward. Therefore, the case ethically justifies the use of SLMs in global education contexts.

3. Conclusion

Since 2020, large language models have made rapid advancements, paving the way for new educational technology. AI systems can now understand and produce discourse similar to that of humans, resulting in educational applications ranging from automated tutoring to thoughtful content generation. However, this windfall opportunity comes with a practical reality: the largest models are impractically expensive and disruptive to deploy at scale. Our study of specialized language models offers possible solutions—we can leverage smaller, task-specific models to achieve most of the educational promise of LLMs at a fraction of the cost. We reviewed how LLMs have evolved in capability, pointing to SLMs as possible models in response to the challenges of LLMs. We pointed to a growing body of empirical evidence that these specialized models matter, when used in educational contexts, and can improve student learning, whether it is improving student mastery of the content, improving teacher response time in feedback, or supporting personalized learning at scale. From the synthesis of the recent literature, it is clear that SLMs are most effective when they have the opportunity to learn for a specific purpose (e.g., algebra tutor, writing coach, multilingual classroom assistant, etc.). SLMs are not designed as a substitute for many of the general learning capabilities of LLMs; rather, SLMs narrow the general learning capabilities of LLMs to specific educational tasks, resulting in tools that are more effective. The meaning of this for global education is significant. If every student has access to an AI tutor that is knowledgeable, patient, and available at any moment, it could help eliminate many long-standing barriers to education. For example, students in large classes or in schools in remote areas could benefit from guided practice or personal explanations that may be lacking in the education system from their human teachers. Also, since SLMs can be deployed relatively easily, a developing country or educational institution could maintain a level of sovereignty and customization over its educational AI tools—aligning them with local curricula and languages and creating a social and cultural fit, as well as managing sensitive data and privacy issues better than a cost-effective, one-size-fits-all cloud AI. Similar to a one-size-fits-all AI application in education, SLMs seem like an opportunity for countries to not only leverage AI in education but to do so on their own terms and at low cost. With that said, we must proceed cautiously. The introduction of SLMs into education should come with teacher training, while also engaging robust security and protocols to protect students. Like all educational endeavors, ongoing examinations will be necessary to optimize SLMs for effectiveness across subjects, ages, and educational contexts.

The years from 2020 to 2025, fortunately, have shown signs of better transparency (many models and studies are publicly available) and collaboration between AI researchers and educators, which is a positive sign in dealing with the aforementioned

DOI: https://doi.org/10.5281/zenodo.15188356

challenges.

On the whole, there is promise in specialized language models to expand and improve educational opportunities at scale by "bridging the gap" between the remarkable capacities of modern state-of-the-art AI and the practical limitations of spaces for learning that are in our control. We can and should provide scalable and personalized education to students everywhere by the careful delivery of lower-cost, specialized language models. Based on the evidence we have seen so far, developing and implementing SLMs with intent and consideration can provide pedagogical benefits for students learning in classrooms, assist teachers, and work toward creating a more equitable education setting. Innovation and collaboration will be necessary to ensure future uses of AI in education are equitable and benefit all students when it comes to advancing educational potential for the next generation of learners.

References

[1] Ghosh, B. (2023, April 20). The rise of small language models—Efficient & customizable. Medium. https://medium.com/@bijit211987/the-rise-of-small-language-models-efficient-customizable-cb48ddee2aad

[2] Subramanian, S., Elango, V., & Gungor, M. (2025). Small language models (SLMs) can still pack a punch: A survey. arXiv preprint arXiv:2501.05465.

[3] Wang, R. E., Ribeiro, A. T., Robinson, C. D., Loeb, S., & Demszky, D. (2025). Tutor CoPilot: A human-AI approach for scaling real-time expertise. arXiv preprint arXiv:2410.03017.

[4] Wang, S., Xu, T., Li, H., Zhang, C., Liang, J., Tang, J., & Wen, Q. (2024). Large language models for education: A survey and outlook. arXiv preprint arXiv:2403.18105.

[5] Sezgin, E., Siracusa, J., & Lin, S. L. (2022). Operationalizing and implementing pretrained large AI linguistic models in the US healthcare system: Outlook of GPT-3 as a service model. JMIR Medical Informatics, 10(2), e32820. https://doi.org/10.2196/32820

[6] OpenAI (Achiam, J., et al.). (2023). GPT-4 technical report. arXiv preprint arXiv:2303.08774.

[7] Peláez-Sánchez, C., Velarde Camagüi, D., & Glasserman-Morales, L. D. (2024). The impact of large language models on higher education: Exploring the connection between AI and education 4.0. Frontiers in Education, 9, Article 1392091. https://doi.org/10.3389/feduc.2024.1392091

[8] UNESCO Institute for Statistics. (2016, October). The world needs almost 69 million new teachers to reach the 2030 education goals (Fact Sheet No. 39). http://uis.unesco.org/sites/default/files/documents/fs39-the-world-needs-almost-69-million-new-teachers-to-reach-the-2030-education-goals-2016-en.pdf

[9] Teki, S. (2025, January 13). Large language models: Performance vs. cost. MentorCruise Blog. https://mentorcruise.com/blog/how-does-the-quality-of-large-language-models-compare-with-their-costs/

DOI: https://doi.org/10.5281/zenodo.15188356