

ORION FORUM

An AI Revolution for Small Businesses: RLMs in Action

MARCH 2, 2026

The potential of artificial intelligence for [small and medium-sized enterprises](#) (SMEs) remains paradoxical. While AI promises to level a playing field long dominated by large corporations, its adoption is often stalled before any technical work begins. According to [the OECD's 2025 survey](#) on SME digitalization, cost-related barriers, including maintenance expenses, hardware investment, and training costs, are the most frequently cited obstacles to AI adoption in business. Contrary to the common assumption that cost is the primary obstacle, recent industry analyses and field experience suggest the biggest barriers are managerial and organizational. A recent study also tackles the fact that for 77% of SMEs who have not yet adopted AI, the main hurdle is [a lack of knowledge or uncertainty about its use cases](#). Once this strategic barrier is cleared, businesses face the second major obstacle: technical complexity. It is this second barrier, the engineering challenge of making AI work with vast, real-world data. A recent publication from [MIT's Computer Science and Artificial Intelligence Laboratory](#) fundamentally challenges these common barriers. Accordingly, a Recursive Language Model (RLM) becomes an inference framework in which a language model programmatically examines, decomposes, and recursively calls itself over segments of input data stored externally, enabling it to process inputs up to 100 times beyond its native context window.

The RLMs in AI performance

In their research, Zhang, Kraska, and Khattab introduce a framework termed Recursive Language Models (RLMs). It demonstrates that future advances in AI performance may result not only from constructing larger neural networks, but also from enabling existing models to

interact with information more intelligently. These findings have significant implications for small businesses. Besides a long-term technological possibility, these implications also represent a near-term shift in the economics of AI accessibility and engineering complexity.

The framework apparently addresses the fact that most real-world business problems involve data volumes exceeding the processing capacity of any single AI model. Considering a regional manufacturer's supply chain documentation or [a healthcare provider's patient records](#) may encompass millions of pages. Today, the most advanced language models cannot even retain this quantity of information in their working memory. This limitation arises not only from a lack of reasoning ability, but also from the restricted 'context window' through which these models process information. Further, this constraint also compels businesses to make difficult compromises. In fact, they must either summarize and compress their data prior to model input. It results in the loss of critical details, or forgo AI entirely for their most complex, information-rich tasks. In practice, only large corporations possess the computing resources and engineering expertise necessary to manage this [trade-off effectively at scale](#).

Therefore, the MIT researchers (Zhang, Kraska, and Khattab) propose an alternative methodology. Instead of attempting to fit an entire corpus into a model's limited context window, their model is provided with a programmatic interface that enables it to generate instructions, selectively read document sections, process them, and recursively synthesize findings from the entire corpus. This iterative process, accordingly, allows the AI to revisit previous results, refine its understanding, and construct a comprehensive analysis incrementally.

The impact of RLMs as an intelligent architectural scaffold

The empirical results of RLMs are notable. On a benchmark requiring multi-hop reasoning across one thousand documents (approximately eight million tokens), [an RLM built on a frontier model achieved over 90 percent accuracy](#). According to Zhang et al., models without the recursive scaffold struggle not just because of limited reasoning abilities but also because they cannot effectively process complex inputs, which is why incorporating such a structure significantly improves performance and transforms the model's effectiveness. The strategic

significance for small businesses lies less in the direct financial savings and more in the dramatic reduction of engineering complexity. Most SMEs prefer to buy, not build, and that complexity, not raw cost, is the bigger constraint. An intelligent architectural scaffold like an RLM replaces the need for complex, in-house data processing pipelines and the specialized talent required to build them. Simplifying the how makes advanced AI accessible to businesses that lack dedicated engineering teams, thereby lowering the effective cost of adoption.

This technological shift also carries important lessons for policymakers. [Much of the current regulatory discourse around AI governance](#) is anchored in model size, parameter counts, or training computing power as proxies for risk and capability. A well-designed architecture can, therefore, multiply the effective capability of a smaller, less-regulated model by an order of magnitude. Governance frameworks that rely solely on model size as a threshold for oversight may therefore miss the mark entirely. As policymakers develop new rules for responsible AI, they must look beyond the raw specifications of individual models and consider the emergent capabilities of entire system architectures in practice. Fostering an ecosystem where smaller, more efficient models can thrive through innovative design will be essential to ensuring that the economic benefits of AI are broadly distributed rather than concentrated among a handful of technology giants.

The research trajectory in the AI field

The researchers acknowledge the limitations of the current approach. Latency increases with greater recursive depth, and cost variability can rise for particularly complex tasks. Models lacking robust coding abilities also face challenges within this framework. These engineering constraints will require time and further research to address. Nevertheless, the research trajectory is clear: the most significant future advances in AI may arise not from larger training runs, but from more sophisticated scaffolding that enables existing models to decompose problems and recursively process complex information relevant to real-world business contexts.

The pursuit of larger models will likely persist. However, for small business owners, policymakers, and analysts, the smarter investment may be in better architecture. While

architectural innovation cannot solve the primary organizational barriers of knowledge and strategy, it can dramatically lower the subsequent hurdles of complexity and cost. The most transformative AI revolution will not be broadcast from a massive data center. Instead, it will be quietly implemented in the recursive processes that finally enable machines to engage with the world as it truly exists.

Orion Policy Institute (OPI) is an independent, non-profit, tax-exempt think tank focusing on a broad range of issues at the local, national, and global levels. OPI does not take institutional policy positions. Accordingly, all views, positions, and conclusions represented herein should be understood to be solely those of the author(s) and do not necessarily reflect the views of OPI.