



AI in Higher Education and the Future of Token Economics

The new costs of scaling AI



Hernán Londoño Chief Technology and Innovation Strategist, Lenovo

Higher education has spent the last few years exploring how AI can support student success, research, administration, workforce preparation, and faculty and student use. Institutions already see its potential to improve services, increase research productivity, reduce administrative burden, strengthen cybersecurity, and prepare students for an AI-enabled economy. But as AI shifts from experimentation to institutional operations, the economic model changes.

The next phase of AI in higher education will be driven more by AI agents than chatbots. Unlike chatbots, agents pursue goals by searching knowledge, retrieving information, using tools, validating results, and iterating until a task is complete. That creates tremendous value—but also significantly more AI consumption. Understanding token economics is therefore becoming essential to scaling AI effectively across higher education.



What Token Economics Means

Tokens are the basic units of AI consumption. Every prompt, response, document, context retrieval, tool call, and reasoning step uses them. In a simple interaction, a user asks a question, the model responds, and the exchange ends. Agentic AI is different. To complete a task, an AI agent may reread context, use multiple tools, generate intermediate outputs, evaluate its work, and revise its approach. What appears to be a single request may actually involve dozens of model calls behind the scenes.

This introduces a new operating unit for AI: consumption. Unlike traditional software, where costs are tied to licenses or subscriptions, AI costs are driven by workflow design. An agent that performs well in a pilot may become costly at scale if it relies on large context windows, repeated model calls, or inefficient reasoning loops. Higher education leaders need to understand that token economics is now a strategic consideration—not just a billing issue.

Why This Matters in Higher Education

Token economics is especially important in higher education, where institutions must improve outcomes while operating under tight budgets. Agentic AI can expand capacity and reduce administrative burden, but only if its costs are managed. Without a clear strategy, AI risks widening the digital divide, leaving under-resourced institutions behind.

The answer is not less AI, but better AI architecture. By designing efficient workflows, choosing the right mix of cloud and local infrastructure, and managing token consumption, institutions can keep AI affordable, scalable, and aligned with their mission.

Why Agentic AI Changes the Cost Model

Agentic AI has the potential to transform higher education by helping students navigate administrative processes, supporting faculty and advisors, accelerating research, automating operations, and improving access to institutional knowledge. But agents are more than chatbots—they are workflow systems that use tools, reason through multiple steps, access context, and adapt as tasks evolve. As a result, an agent's cost reflects not just the final answer, but the process required to produce it.

That distinction matters because many of higher education's most valuable use cases—advising, enrollment, financial aid, research administration, compliance, IT, and student support—are complex workflows. The challenge isn't that agentic AI is too expensive, but that costs can escalate if institutions don't manage AI consumption. Just as cloud computing and SaaS required new approaches to governance and budgeting, agentic AI will require similar discipline around token economics to keep deployments efficient, predictable, and sustainable.

Token Economics: An Architecture Issue

The strategic question is no longer which AI tool to buy, but where AI should run. While frontier cloud models remain essential for advanced reasoning and complex agentic workflows, many tasks can run more efficiently on smaller models, institutional AI platforms, private clouds, on-premises infrastructure, AI PCs, or edge devices. The future of higher education AI is hybrid—not cloud-only.

Token economics helps institutions make smarter architectural decisions, including:

- Which tasks truly require frontier models?
- Which can run on smaller or on-device models?
- Which workloads should stay close to institutional data?
- Which require human review?
- Which agents need limits on tools, context, or iteration?
- Which AI workloads deliver enough value to justify their cost?

These are not just technical decisions — they are governance, budgeting, and institutional strategy decisions.



AI Phones and the Personal Edge

AI is becoming more ambient, extending beyond individual applications to work across PCs, phones, tablets, wearables, and other personal devices. For higher education, this matters because the smartphone is often a student's primary computing device and an essential tool for faculty and staff. While sensitive institutional workflows won't move entirely to phones, AI experiences will increasingly follow users across devices.

A student may start a task on a phone, continue it on an AI PC, and finish it in an institutional system. Faculty, researchers, and staff will similarly move between mobile devices, workstations, and campus platforms. The phone isn't replacing institutional systems — it's becoming part of a distributed AI ecosystem. Higher education AI strategies should be designed with that reality in mind.

AI PCs and the Move Toward Local Intelligence

This is where AI PCs become important. These devices are more than faster laptops — they are a shift toward running more AI closer to the user. While cloud models remain essential for complex reasoning and advanced agentic workflows, many everyday tasks — summarization, drafting, translation, meeting support, search, and accessibility — can increasingly run locally on AI PCs.

This hybrid approach gives institutions more flexibility. Tasks that require frontier models can use the cloud, while routine work stays on the device, improving responsiveness, reducing data movement, supporting privacy, and lowering AI consumption. For higher education, where thousands of users perform routine, context-specific tasks every day, moving some intelligence to the edge creates a more scalable and cost-effective operating model.

Moving some intelligence to the edge creates a more scalable and cost-effective operating model.

Intensive Science at the Edge

On-device AI is not only about productivity — it is also transforming research. In scientific environments, the edge may be a high-performance workstation located close to the researcher and the data, reducing delays caused by shared computing resources, data movement, and limited access to compute.

NC State's hurricane research program illustrates the impact. By deploying Lenovo ThinkStation workstations with NVIDIA Blackwell GPUs, researchers created a dedicated environment for AI model development, preprocessing, and analysis. Graduate students and postdocs reported cutting modeling time by more than 50%, while hurricane forecast models ran more than five times faster than on the university's previous high performance computing platform.

The lesson is that edge computing isn't just about AI PCs. High-performance workstations can bring AI closer to research, helping faculty and students iterate faster, increase productivity, and accelerate scientific discovery. Together, AI PCs and workstations represent a shift that moves intelligence closer to where work happens

Landing the Plane

Higher education no longer needs abstract conversations about AI — it needs practical ones. As institutions move from AI experimentation to AI operations, they must balance cost, value, architecture, access, and governance. Ignore token economics, and AI pilots may become too expensive to scale. Focus only on cost, and institutions risk limiting tools that improve student success, research, and operations.

The better path is a balanced, hybrid AI strategy. Use cloud AI for testing and development, and proof of concept, institutional infrastructure where governance requires it, AI PCs and mobile devices for local productivity, and workstations for research. The future of AI in higher education will be defined not by a single model or platform, but by placing intelligence where it creates the most value. The question is no longer whether institutions should adopt AI, but where that intelligence should run.

About the Author

Dr. Hernán Londoño serves as Lenovo's Chief Technology and Innovation Strategist. With over 25 years of experience as a Chief Technology Officer (CTO) and Chief Information Security Officer (CISO) in higher education and as a Chief Technology and Innovation Strategist in private industry, he is known for his expertise in innovative academic solutions, cybersecurity, and strategic technology advancement.

Dr. Londoño holds a B.S. and M.S. in Computer Science and Information Technology from Barry University, a Ph.D. in Computer Science from Nova Southeastern University, and a Post-Doctorate Certificate from Harvard Business School, focusing on AI in business competitiveness. Currently Dr. Londoño's academic research centers on the human aspect of the impact of AI in cybersecurity risk.

Toward a Token-Aware AI Strategy

Toward a Token-Aware AI Strategy A token-aware AI strategy isn't about discouraging AI use—it's about maximizing value. Institutions should evaluate AI use cases not only by what they can do, but by the value they create relative to what they consume. High consumption workflows may be worthwhile if they improve student success, accelerate research, reduce risk, or save significant staff time.

A practical framework includes five steps:

1. Classify AI workloads by purpose, sensitivity, and performance needs.
2. Match workloads to the right compute environment, from cloud models to institutional infrastructure, AI PCs, and workstations.
3. Pilot agentic workflows with consumption monitoring before scaling.
4. Expand AI governance to include token economics, model selection, workload placement, and cost visibility.
5. Prioritize equitable access, ensuring AI remains sustainable and available to students, faculty, staff, and researchers.



Explore [Lenovo Solutions For Higher Education here.](#)