UC SANTA BARBARA
THE Current

October 27, 2026
Debra Herrick

Institute for Energy Efficiency invests in energy-efficient AI, data centers and quantum technologies

As artificial intelligence and advanced computing grow more powerful, so does the energy required to run them. UC Santa Barbara researchers are tackling this challenge from several directions, such as making AI models more efficient, rethinking how data centers are cooled and how quantum technologies are built.

Five projects have received seed funding from the university's Institute for Energy Efficiency (IEE), supporting early-stage research across these areas.

"The problems we're trying to solve are becoming more urgent, from the energy demands of AI and data centers to the next generation of quantum technologies," said Steven DenBaars, director of IEE. "These awards give our researchers the chance to take on those problems early, test ambitious ideas and move the most promising technologies closer to real-world use."

The projects were selected through IEE's 2026–27 Research Seed, Software Impact, Gauthier Quantum and Photonics Seed grant programs, which support early-stage work that can open new research directions and lay the groundwork for larger-scale investment.

"We want to give researchers the freedom to pursue ideas that are high-risk, high-impact and capable of opening entirely new directions," said DenBaars, a

Distinguished Professor of Materials and Electrical and Computer Engineering in UCSB's Robert Mehrabian College of Engineering and holder of the Mitsubishi Chemical Chair in Solid State Lighting and Displays.

Making AI agents work with fewer tokens

As AI agents take on longer tasks — conducting research, navigating software or completing digital workflows — they can generate and reprocess vast numbers of tokens, the units they use to reason and respond. The computational cost grows with every step. Xin Eric Wang, an assistant professor of computer science, is working to help AI agents complete those tasks using fewer tokens and less computation.

"If there are multiple ways to solve the same problem, the model should be able to choose the most efficient one," Wang said. "If we can predict which path costs less, we can choose the least expensive one with minimal cost."

Wang's team is developing an open-source system that tackles the problem in two ways: training models to reach successful answers through shorter reasoning paths and reducing the amount of information long-running agents must repeatedly process. The team is targeting a fraction of token usage while maintaining performance.

The project builds on earlier work from Wang's group that enables a model to estimate in advance how costly different routes to a solution are likely to be.

Matching AI energy use to the task

Today, a simple request can send a large AI model through much the same computational pipeline as a difficult reasoning problem. Electrical and computer engineering professors Mahnoosh Alizadeh and Ramtin Pedarsani are developing AI systems that could change that, using lighter configurations for easier requests and reserving the full model for more demanding ones.

"The focus of this seed project is on developing energy-proportional AI: foundation models that allocate computation according to the difficulty of each request," said Alizadeh and Pedarsani, whose goal is to create AI systems that use only as much computing power as each task requires.

An online learning system would choose the lowest-energy option that can still meet a required level of quality. Rather than judging efficiency only by model size or the number of computations performed, the researchers plan to measure the energy required to produce a correct response.

The work builds on earlier research from the team showing that large vision-language models contain substantial redundancy. In one project, they removed 25% of a model's layers while retaining about 89% of its average benchmark performance.

Bringing quantum light sources closer to room temperature

Many quantum technologies require extreme operating conditions, including cryogenic temperatures and ultra-high vacuum. Sriram Krishnamoorthy and Galan Moody are exploring whether semiconductor-based quantum systems could instead operate at or near room temperature.

"Color centers and point defects in wide-bandgap semiconductors offer a compelling path toward energy-efficient quantum technologies because they can host optically addressable quantum states that operate at or near room temperature," said Krishnamoorthy, a professor in the Materials Department, and Moody, a professor of electrical and computer engineering.

Their project focuses on color centers, atomic-scale defects in crystalline materials that can interact with light and, in some cases, emit individual photons capable of carrying quantum information. The researchers will investigate these defects in gallium oxide, an ultra-wide-bandgap semiconductor that could support stable single-photon emission at room temperature.

The work brings together Krishnamoorthy's expertise in high-quality gallium oxide materials and Moody's expertise in quantum photonics. Their goal is to demonstrate single-photon emission from gallium oxide for the first time, opening a potential path toward compact quantum light sources that require far less energy-intensive infrastructure.

Cooling data centers where they need it

Thermal management can account for 30% to 50% of a data center's total energy use. Yangying Zhu and Igor Mezić, professors of mechanical engineering, aim to

reduce that burden by directing cooling to the processors that need it most as computing loads change.

“By directing cooling resources where they are most needed and reducing overcooling of low-load devices, this approach enables significant reductions in overall cooling energy consumption,” said Zhu and Mezić, who are developing a more adaptive way to manage heat in energy-intensive data centers.

Their system combines advanced chip-level cooling with an AI-enabled control system that can redistribute coolant across computational processing units and graphics processing units in real time. The researchers estimate that a successful demonstration could reduce cooling energy use in AI data centers by 20%.

At the chip level, tiny silicon channels would carry coolant close to heat-generating electronics, including systems in which liquid absorbs heat as it turns to vapor. Across the larger cooling network, predictive models based on Koopman theory would anticipate changing thermal loads and adjust coolant flow accordingly.

The goal is a system that responds as computing demands shift, rather than spending energy cooling every component equally.

Designing hardware for probabilistic computing

Some computational problems depend heavily on generating random numbers and repeated mathematical operations known as dot products. Used across fields from artificial intelligence and physics to statistics and cryptography, those operations can become major computational bottlenecks.

Dmitri Strukov, a professor of electrical and computer engineering, is exploring whether specialized hardware could make these probabilistic computing tasks faster and more energy efficient than conventional processors.

“Many modern workloads are fundamentally ill-suited to general-purpose processors and even highly data-parallel architectures, such as graphics processing units,” Strukov said. “Highly efficient, purpose-built computational circuits continue to emerge, offering dramatic gains in performance and energy efficiency.”

The project reflects a broader shift toward designing specialized hardware for computationally demanding tasks that conventional processors handle inefficiently. Strukov and his collaborators will use the work to generate preliminary data toward

a scalable framework for probabilistic computing and a larger external funding effort.

Tags
Energy Efficiency

Media Contact
Shelly Leachman
Editorial Director
(805) 893-2191
sleachman@ucsb.edu

About UC Santa Barbara

The University of California, Santa Barbara is a leading research institution that also provides a comprehensive liberal arts learning experience. Our academic community of faculty, students, and staff is characterized by a culture of interdisciplinary collaboration that is responsive to the needs of our multicultural and global society. All of this takes place within a living and learning environment like no other, as we draw inspiration from the beauty and resources of our extraordinary location at the edge of the Pacific Ocean.