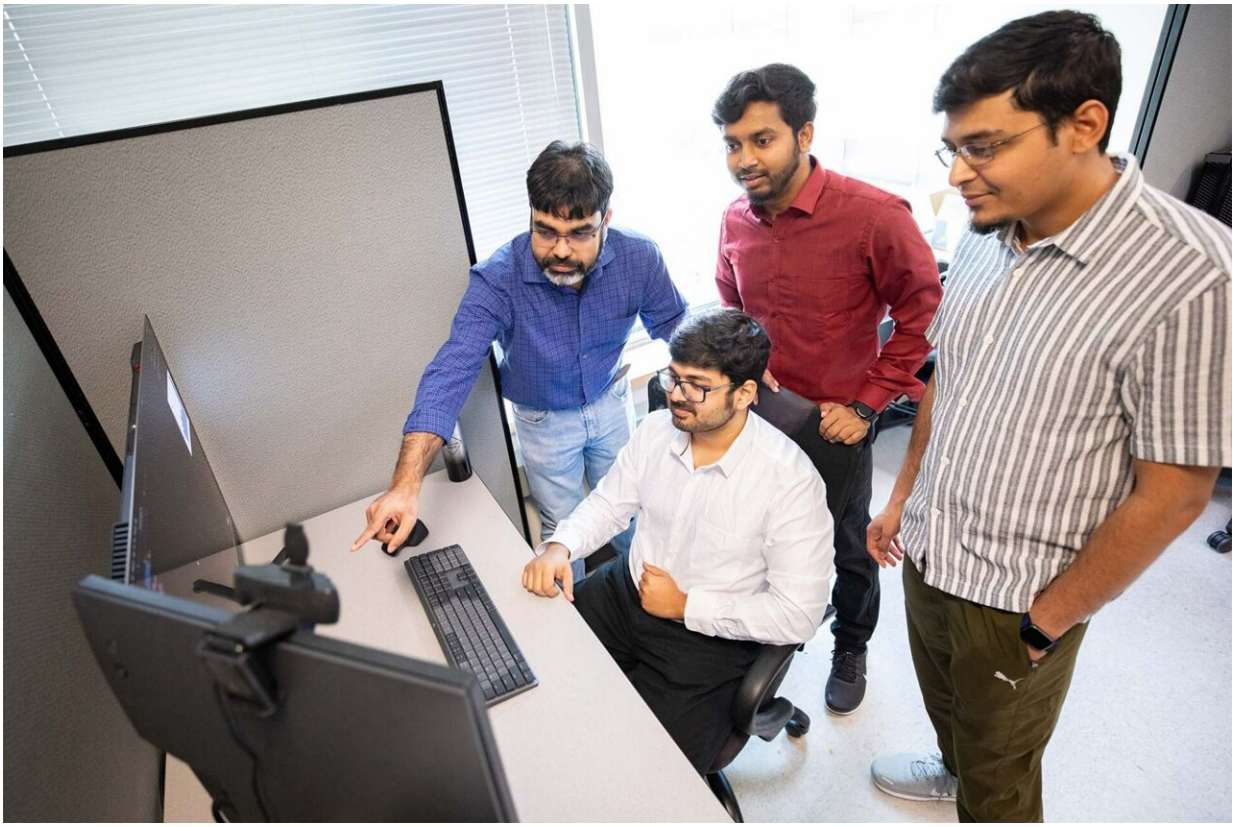
Researchers' approach may protect quantum computers from attacks

March 7 2024, by Kim Horner



Dr. Kanad Basu (left) and his colleagues developed a way to counteract the impact of attacks designed to disrupt artificial intelligence's ability to make decisions or solve tasks in quantum computers. His team includes computer engineering doctoral students Sanjay Das, Navnil Choudhury (sitting) and Shamik Kundu (right). Credit: University of Texas at Dallas

Quantum computers, which can solve several complex problems exponentially faster than classical computers, are expected to improve artificial intelligence (AI) applications deployed in devices like autonomous vehicles; however, just like their predecessors, quantum computers are vulnerable to adversarial attacks.

A team of University of Texas at Dallas researchers and an industry collaborator have developed an approach to give quantum computers an extra layer of protection against such attacks. Their solution, Quantum Noise Injection for Adversarial Defense (QNAD), counteracts the impact of attacks designed to disrupt inference—AI's ability to make decisions or solve tasks.

The team will present research that demonstrates the method at the IEEE International Symposium on Hardware Oriented Security and Trust held May 6–9 in Washington, D.C.

"Adversarial attacks designed to disrupt AI inference have the potential for serious consequences," said Dr. Kanad Basu, assistant professor of electrical and computer engineering in the Erik Jonsson School of Engineering and Computer Science. "An attack can be likened to someone putting a sticker over a stop sign: An autonomous vehicle may not recognize the stop sign properly, interpreting it as a reduced speed sign and hence, fail to stop. Our goal with this approach is to make quantum computer applications more secure."

Quantum computing is a rapidly emerging technology that uses quantum mechanics—the study of how particles behave at the subatomic level—to solve complex computational problems.

Like bits in traditional computers, qubits represent the fundamental unit of information in quantum computers. Bits in classical computers represent 1 or 0. Qubits, however, take advantage of the principle of

superposition, which means they can simultaneously be in a state of 0 and 1; therefore, qubits can represent two states, resulting in dramatic speedup capabilities compared to traditional computers. As an example, due to their computing power, quantum computers have the potential to break highly secure encryption systems.

One of the challenges of quantum computers is their susceptibility to "noise," or interference, due to factors including temperature fluctuations, magnetic fields or imperfections in hardware components. Quantum computers also are prone to "crosstalk," or unintended interactions between qubits. Noise and crosstalk can result in computing errors.

The researchers' approach leverages intrinsic quantum noise and crosstalk to counteract adversarial attacks. The method introduces crosstalk into the quantum neural network (QNN), a form of machine learning in which large datasets train computers to perform tasks, including detecting objects such as stop signs or other computer vision responsibilities.

"The noisy behavior of quantum computers actually reduces the impact of attacks," said Basu, who is senior author of the study. "We believe this is a first-of-its-kind approach that can supplement other defenses against adversarial attacks."

The researchers demonstrated that, during an attack, an AI application was 268% more accurate with QNAD than without it.

Shamik Kundu, a computer engineering doctoral student and a first co-author, said the approach is designed to supplement other techniques to protect quantum computer security. Kundu likened the framework's benefit to that of seat belts in cars.

"In case of a crash, if we do not wear the seat belt, the impact of the accident is much greater," Kundu said. "On the other hand, if we wear the seat belt, even if there is an accident, the impact of the crash is lessened. The QNAD framework operates akin to a seat belt, diminishing the impact of adversarial attacks, which symbolize the accident, for a QNN model."

Provided by University of Texas at Dallas