

# Can AI Keep Accelerators in Line?

REBECCA McDONALD

## Los Alamos scientists are using AI to help reduce beam loss in linear accelerators to preserve precious experiment time.

Several years ago, Los Alamos physicist Alexander Scheinker was performing some experiments at SLAC National Accelerator Lab when the plasma accelerator stopped working. The accelerator operators tried everything they could think of to fix the problem, but to no avail. The scientists were agitated, losing precious time and money that had been allotted to their experiment. Eventually, the crew called a former operator who was retired and living in Hawaii, waking him at 4:00 a.m. to ask for help. In his sleepy stupor, the retiree directed them to a specific accelerator module and told them to look behind an equipment rack in sector 32, unplug a specific ID number box that had likely overheated, wait one minute, and plug it back in. They followed his instructions, reset the box, and voila! It was fixed.

When accelerator beams are running, technicians work 24 hours a day, 7 days a week problem-solving issues—often through trial and error. Not every problem is as major as the beam not working, but a team of operators monitors thousands of parameters to maintain safe beam function, and their work is critical to keeping science on track. Through years of experience, they learn the ins and outs of these massive instruments, which is why a retiree in Hawaii remembered exactly what to do. But no matter how experienced they are, operators frequently face challenging beam-adjustment decisions in real time when experiments are in process and the stakes are high.



Use existing images to train AI model →



Give model raw data to generate into new image →

Generative diffusion-based AI models are the state-of-the-art method for creating highly accurate representations of complex objects such as photos. Diffusion models are trained by gradually adding noise to distort and blur a data sample or image to the point that it becomes a random distribution; then, the model learns to predict a new image by gradually removing noise.

Scientists at Los Alamos want to make these adjustment decisions easier for operators and they're turning to artificial intelligence (AI) for help. By training AI models to predict what the particle beam looks like at a given time and to suggest how to keep the beam optimized, the Los Alamos team is developing tools that will improve data collection while saving time, money, and even sleep.

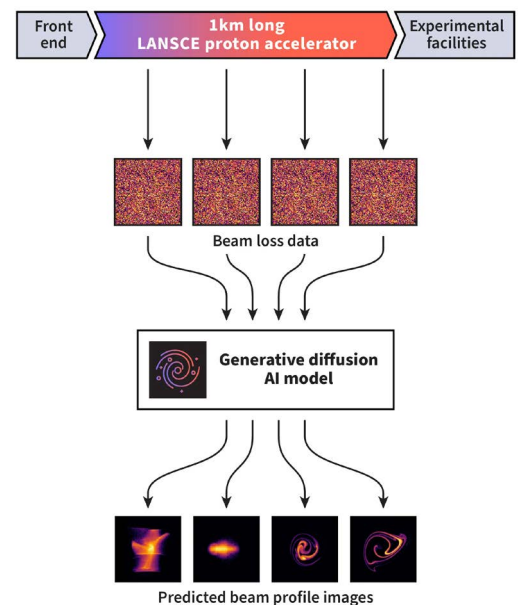
## **“What’s going on inside the beam chamber at LANSCE is an extraordinary milieu of forces operating in six dimensions: the standard three (x, y, z) plus motion for each of them”**

### **6D diagnosis**

The particle accelerator at the [Los Alamos Neutron Science Center \(LANSCE\)](#) produces a 1-millimeter-wide beam of protons traveling at about 563 million miles per hour; delivering 800 kilowatts of power. Metallic chambers called resonant cavities are fed by megawatts of radiofrequency power to create extremely powerful electric fields that accelerate the beam, while magnets force the beam to remain focused. The protons must be pointed in the correct direction when they travel through the electric field for this to work. In addition, the beam’s own internal electric fields are all the time trying to rip the beam apart, because all of the particles involved have the same charge. Altogether, what’s going on inside the beam chamber at LANSCE is an extraordinary milieu of forces operating in six dimensions: the standard three (x, y, z) plus motion for each of them.

With such powerful forces at play, the beam is never quite perfectly focused, and there are always some protons that stray away and hit the sides of the accelerator chamber. This phenomenon is called beam loss. The beam is also affected by environmental perturbations. The vibrations of the machinery can wiggle the beam out of focus, and when the sun comes out and warms the ground outside the chamber, even slight temperature changes to accelerator components can cause further beam loss. Although a small amount of beam loss is expected, three or more operators work in tandem whenever the beam is running, adjusting hundreds to thousands of parameters to minimize beam loss as best they can.

This constant need for attention is because too much beam loss can cause problems. Scientists want the beam to be focused to deliver a powerful, concentrated stream of protons to a precisely positioned experiment. In addition, when stray protons hit the chamber walls, they slowly damage it and can even make the walls radioactive. Too much beam loss means that components will wear out and have to be replaced more frequently, and if they become radioactive, any maintenance or replacement will require long delays to safely address the radiation. Finally, another issue is that with 800 kilowatts of power, too much loss could literally burn a hole through the chamber wall. The operators’ goal is clear, but their work is challenging because beam loss data are limited; it’s difficult to know exactly what is going on inside the accelerator chamber to determine which adjustments are necessary.



Los Alamos scientists are using AI to help maintain the health and effectiveness of the LANSCE proton accelerator beam by detecting beam loss (protons that have gone astray). First, they trained a generative diffusion model using real images of proton beams. Then, the team members can take beam loss data from various points along the 1-kilometer-long LANSCE accelerator and feed the raw data into the AI model to generate new, predicted images of what the beam looks like at each point. These images help the operators determine how to adjust the beam parameters to minimize beam loss.

Scheinker equates the challenge of diagnosing beam loss to watching a fever in a sick child: “You can’t necessarily see what’s going on inside, you just know you’re getting data that something is wrong.”

Throughout the length of the kilometer-long accelerator chamber, dozens of detectors collect data (separate from experimental data) that are used to estimate beam loss. The beam loss data provide a very sparse one-dimensional signal, which is basically the count of how many protons are hitting the walls of the accelerator in various locations. Unfortunately, these data do not accurately depict the complexity of what is really happening in the 6D space of the particle beam.

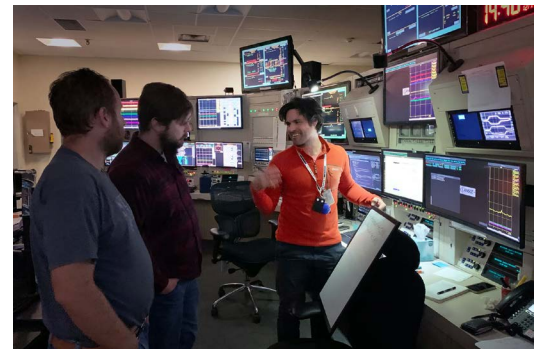
Based on the beam loss values, operators make small adjustments to the strengths of electric and magnetic fields throughout the accelerator. Many adjustments are manual, made based on experience rather than an automated algorithm, and most parts of the beam are not physically reachable, so it’s impossible to see, with eyes or camera-like images, what’s really going on.

## **“You can’t necessarily see what’s going on inside, you just know you’re getting data that something is wrong”**

“When the beam is running, the feedback data we get is minimal. This low-dimensional signal doesn’t tell us enough. We never know what the beam looks like, we just know that there is loss,” says Scheinker.

Some accelerators, like the European x-ray free electron laser facility (European XFEL) are outfitted with screens that intercept the beam to evaluate it. The screens are made of a scintillating material that sends light to a digital camera, capturing images of the beam cross section. These images help the operators “see” the shape of the beam to better diagnose and correct beam loss.

At LANSCE, screens are only available at a few locations. Other measurement devices such as wire-scanners are also available, but again only at a very limited number of locations. They are also slow and involve passing a wire through the beam to measure its profile. Furthermore, to use either screens or wire scanners, the beam intensity must be reduced. So, not only are these diagnostics sparse and slow, but they also interrupt experiments. For these reasons, Scheinker and his colleagues began to explore the idea of employing AI to help them see what the beam looks like and understand how to tune it.



Operators talk with Alexander Scheinker (right) in the LANSCE control room.

### **Bring in the computers**

“Generative AI models could give a picture of what the beam looks like without interrupting the experiment,” says Scheinker.

Using AI to help diagnose and control beam loss has a lot of potential, however, off-the-shelf systems can’t address the complex, time-varying nature of particle accelerators. In other words, because the beam is always changing based on the forces involved—the vibrations, the weather, and so forth—any AI-predicted adjustments may not be valid if the beam in question doesn’t match the beam measurements that were used to train the AI. And, stopping or slowing the beam to get new measurements to constantly re-train the AI is, in itself, destructive to the beam, and therefore not tenable.

To apply AI to the beam loss problem, the Los Alamos team is developing novel, adaptive AI that can use non-destructive signals to track the time-varying accelerator and its beam. “We are using generative diffusion models to create detailed images of the beam’s phase space directly from non-invasive measurements such as beam loss or beam current or beam position monitors,” says Scheinker.

Generative diffusion-based AI models are the state-of-the-art method for creating highly accurate representations of complex objects such as photos, videos, and even protein structures. For instance, Google's AlphaFold tool maps sequences of amino acids to their corresponding 3D protein structures via a generative diffusion process. Diffusion models are trained by gradually adding noise to distort and blur a data sample or image to the point that it becomes a random distribution, then the model learns to predict a new image by gradually removing noise.

Scheinker's team is the first to develop generative diffusion-based methods for virtual charged-particle beam diagnostics. Their model maps non-invasive measurements to high-resolution representations of a beam's 6D phase space. The team began by working at the European XFEL where they had access to large amounts of training data from that facility's screen images. Using 100,000 experimental electron beam images, the team trained an AI on a diffusion-based approach capable of creating megapixel views of the electron beam's time-versus-energy projection based only on non-invasive measurements. These beam views were then verified against real images, and the diagnostic tool was used to create new, predicted images to estimate beam loss based on unknown data. And it worked!

Next, the Los Alamos team adjusted their model to include multiple modes and developed a super-resolution diffusion process in which they added hard physics constraints to more accurately represent the beam's 6D phase space distribution. In addition to ensuring that the model can be validated by physics calculations, another important aspect of the team's virtual beam diagnostic tool is that it includes adaptive feedback. This means that as the accelerator beam changes in response to its environment, the data are combined with the diffusion-based approach to adaptively and non-invasively track the time-varying beam properties so that the beam predictions account for those changes.

Most recently, Scheinker and the adaptive machine learning team brought home their successful approach and developed the first generative diffusion-based virtual beam diagnostic for the LANSCE accelerator. Ultimately, the team plans to create a robust tool with a graphical user interface for LANSCE operators to use.

"We want to be able to ask questions like: based on these settings, what does the model think the beam looks like right now?" says Scheinker.

---

## **"Generative AI models could give a picture of what the beam looks like without interrupting the experiment"**

### **Accelerator assistant**

Figuring out what the beam looks like—and therefore the likely amount of beam loss—is of course only half the problem faced by accelerator operators. The second half is determining which adjustments are needed to correct the beam. Because most of the adjustments are done by hand, operators build, through experience and over time, a robust understanding of how to care for the accelerator beam. In addition, every setting, measurement, and activity is meticulously documented in logbooks for future problem-solving.

Scheinker's team and other members of the LANSCE Instrumentation and Controls group are collaborating with Los Alamos scientist Reeju Pokharel to help expedite the problem-solving process using AI. The team is looking to leverage decades of experience and records (LANSCE opened in 1972) into a virtual AI expert capable of giving advice on tuning the proton accelerator.

For this project, the team is using large language models (LLMs) and a specific method known as retrieval-augmented-generation (RAG). LLMs are powerful tools that can learn



from a wide range of data by looking for distinct relationships in data. They are often used to generate text by predicting each sentence word-by-word, based on the languages used for their training. However, even the most powerful commercial LLMs, like ChatGPT, cannot maintain a large file—like a PDF of a 200-page book—in their working memory, which makes it difficult to train them on new, specialized technical information such as proton accelerator physics. Using RAG, huge collections of documents (think hundreds of papers or books) are broken down into smaller pieces that can be quickly searched and incorporated as specific context for scientific applications, leading scientists to get more meaningful answers out of LLMs.

Pokharel recently used her machine learning expertise to develop PLUTO, a RAG data intelligence system for historical documents related to materials science research including plutonium. The generic framework of the RAG-based PLUTO tool makes it easy to apply to other specialized applications. Now, Pokharel is leading the development of a RAG-based, LANSCE-specific virtual expert by vectorizing a large collection of accelerator physics books, journal papers, and more than 30 years’ worth of LANSCE operations logs. When complete, the LANSCE LLM virtual expert will be able to help real-life operators navigate accelerator adjustment.

How will the virtual expert help? The LANSCE LLM virtual expert is being designed to take, as input, a handful of beam diagnostic data (including diffusion-generated images) together with natural language-based situational descriptions from operators. Then, the tool will provide suggestions based on historical data and references. For example, a possible scenario starts with a particular LANSCE component tripping off. The team that is called in to restore it would give the LANSCE LLM an overview of the day’s events, any available diagnostic data collected directly before the trip, and ask the LANSCE LLM something like “how was this problem solved in the past?” The LLM can then suggest several things for the operators to consider checking.

The LANSCE-LLM virtual expert has the potential to be invaluable for knowledge retention across generations of workers as they learn from years of logs written by experienced professionals, many of whom retired a long time ago. It can also help operators and beam physicists diagnose and solve problems more efficiently so they can get back to doing their science.

“This project is important for me personally, as a materials scientist, because if I’m doing a time-sensitive experiment at one of the neutron beamlines, then I would like to have the neutron beam restored as soon as possible after a problem, so we can conduct our experiments and record high-quality data,” says Pokharel.

Several custom RAG-based systems are now in development across the Los Alamos community, as well as other AI-driven projects—all leading to more efficient troubleshooting and ultimately, more effective science. Many of these tools are being developed by an interdisciplinary effort called the AI STRIKE team, of which Pokharel and Scheinker are both members. The STRIKE team was formed in early 2025 and is charged with providing advice and guidance to help groups from around the Laboratory implement AI for their various needs. Pokharel says that the most common questions that required support from the STRIKE team so far have been about setting up RAG systems to use AI to learn from large sets of historic documents, specialized texts, literature, policies, and safety documents.



Piles of LANSCE operations logbooks. Los Alamos scientists are vectorizing more than 30 years’ worth of logbooks and other materials to create a LANSCE virtual expert using a large language model and a technique called Retrieval-Augmented-Generation.