Can Data Mind the Data?

In the data-rich world of higher ed, leveraging AI and automation to understand and manage data risks or recovery is par for the course. The next level: AI and ML tools that access and leverage that data for other purposes.

A **NEW REPORT FROM FORRESTER**, "**Bring-Your-Own-AI Hits the Enterprise**," reveals that more than a quarter of global IT decision-makers indicate that 51% to 75% of their employees will likely use generative AI technology by the end of 2024. Higher education security and IT leaders must consider similar potential within their own environments, where students, faculty, administrators, and other constituents stand in for "employees" in the enterprise, but the risks to data security are just as great, if not more critical given the nature of the data potentially exposed.

Another recent survey, "**Generative AI Through the Eyes of Gen Z**," notes that, of the 43% of respondents who indicated they have used generative AI tools to "help with schoolwork," 51% were college students.

While this somewhat-nascent toolset has endured a lot of hype that has yet to come to fruition, many campuses have taken a "wait-and-see" approach rather than fully embracing or preparing for AI's potential in the classroom or as part of institutional tasks. Yet when it comes to managing risk, there can be no waiting.

"Immature data governance, concerns about algorithmic bias, and ineffective data management and integration pose the greatest challenges to the implementation of AI in higher education," D. Christopher Brooks wrote in the June 2022 edition of *EDUCAUSE Review*.

As higher education's both known and dark data stores continue to grow exponentially larger and faster — mountains of student data, and data culled from research, for starters — all of that data increasingly finds itself up for grabs by untested, unsanctioned, or just plain unknown AI tools. As a result, data management and backup and recovery solutions and security are more important than ever before.

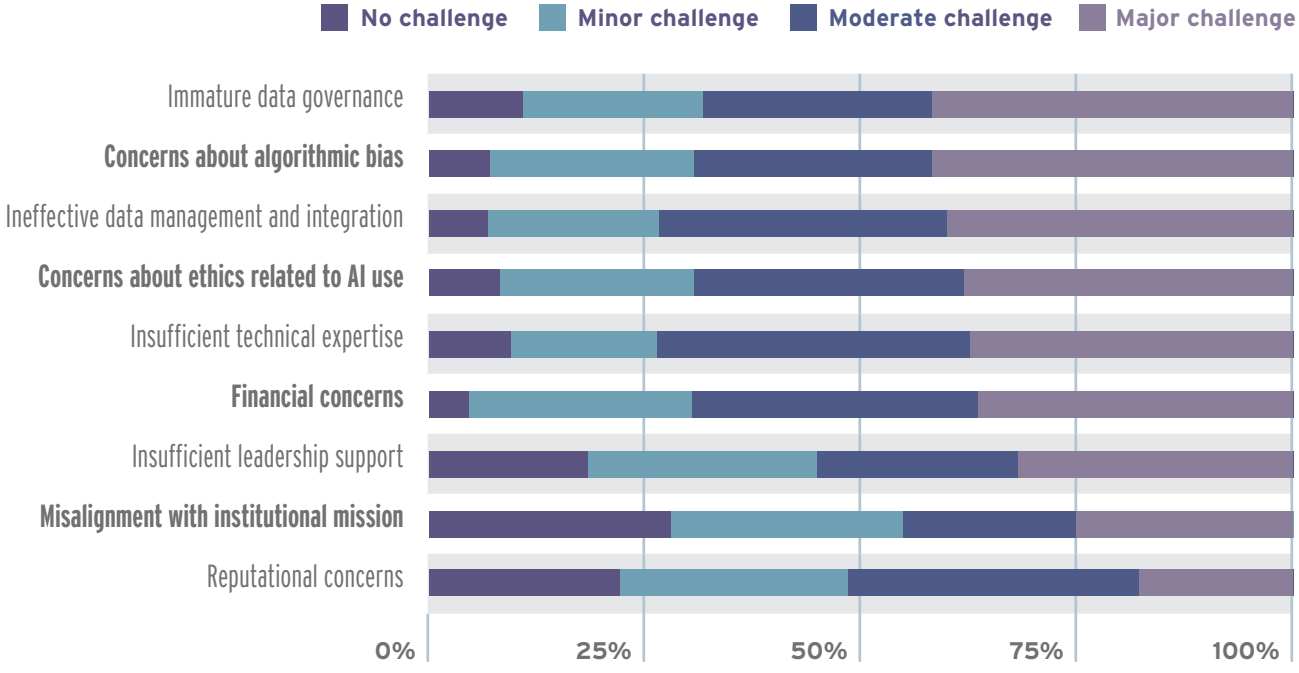
Data Management and Security Challenges Exposed Anew

Data management and governance, security, and integrity continue to rankle higher education IT and cybersecurity teams — and have done so at least since the era of Big Data, if not longer. The open nature of higher education systems, in many cases, is the cause of delayed or otherwise incomplete data governance planning; however, new risks and potential for solutions and tools that leverage AI and ML capabilities have renewed urgency for effective and complete data management.

"It comes down to a risk-benefit analysis conversation, really," said Christian Westervelt, senior manager, technical sales engineering for healthcare, public sector, and education at Veritas. "Does inviting AI to accomplish a task, whether it's pattern recognition, code development — does it make sense or does it open up a potential risk vector?

"As with any new technology, in many cases it's about getting back to the basics. Even with all of the capabilities that are available and accessible to our customers, I'm still surprised by the core conversation around recovery, recovery point, and recovery time. Those fundamentals have not changed in many, many years. What's changed is the 'how.' How can we leverage that technology?

Common Challenges to the Implementation of AI source: **EDUCAUSE Quick Polls**, June 2021



Make that determination but focus on the basics. Get the plan. Make sure it's documented. Make sure it's tested. That's the rock on which we can move forward," Westervelt advised.

Can AI Improve Resiliency?

An institution's ability to protect, recover and maintain its data no matter the event or disruption is the keystone of so-called data resiliency. A resilient environment prevents data losses, minimizes down time, and helps teams or institutions wholly recover from any such attack, event, or disruption. Today, "AI, data governance and the cloud are inextricably linked to data resilience," advised author Guy Pearce in the ***ISACA Journal***. "Recognizing that data are at the heart of each of these concepts means that the converged technologies need to be the focus of an organization's data resilience strategy."

Deep insight into what is happening in an environment at any given time is required for proper recovery, Westervelt said. "There's the dark data assessment process – half of the battle is that they don't know everything about their environment, about their users, about the activity that's taking place. We need to lift those blinders off, let them see or gain access to information about the data they already have, what they're storing or protecting, so that ultimately, should the need arise, they can recover from it."

When the rapid growth of new threat vectors converges with another trend – fewer resources to handle the IT, data management and security workloads – "and with a lot of people working from home, you have a lot less control of the data," said Vishal Kadakia, solution systems engineer at Veritas. "You don't always know where it resides."

Tools or solutions that leverage AI and automation can help fill the gap, and "that's one of the reasons why it's so prevalent now," Kadakia added.

> It comes down to a risk-benefit analysis conversation, really. Does inviting AI to accomplish a task, whether it's pattern recognition, code development – does it make sense or does it open up a potential risk vector?
>
> **– CHRISTIAN WESTERVELT, VERITAS**

"Separating the data and the infrastructure – because they're two different platforms – we can look at the data, the unstructured data, and understand where it resides, who has access to it, the last time it was looked at, whether it has any sensitive information. From an infrastructure standpoint, we need to know what it's doing: Has it been compromised? Are you protecting everything that you need to?" Kadakia said. "When you say, 'AI,' what does that mean for you? Is that an application? How are you leveraging it? It really invites further conversation, which is always a good thing."

Back to Basics: Data Protection and Recovery in the Era of AI

The fundamentals of data protection remain constant in the face of emerging AI-powered threats.

CAMPUS TECHNOLOGY RECENTLY spoke with **Christian Westervelt**, senior manager, technical sales engineering for healthcare, public sector, and education, and **Vishal Kadakia**, solution systems engineer, both at Veritas, about the critical importance of data protection in the current era of rapid AI evolution.

We talk a lot about the potential harm and risk of AI, but in many cases this technology can also help in security and data protection. What are we seeing as far as advances in AI and cybersecurity practices, especially in data management and protection?

Kadakia: We're trying to understand how we can use AI from a cybersecurity perspective against ransomware, and some of the things we can do in early detection. Can we leverage that to do any type of alerting? Can we provide a system that can give an early warning to say, "Hey, you may want to take a look at this. Your data may be compromised." We're seeing that, across our platform, – and it doesn't matter which portfolio we're talking about – leveraging AI and machine learning to help with those types of things, help teams where their data is at risk, where they might not even know data is at risk, with dark data assessment, with data management from start to end. Pattern recognition is key: leveraging our solution suite to detect anomalies, things that either shouldn't be occurring or stand out in the normal day-to-day data management.

There are things that we've always been doing. We have been capturing metadata. Can we use machine learning or AI to look at that metadata to give that early warning? Can we also use machine learning on user behavior patterns to discover if there are any anomalous user behaviors occurring outside of the norm? Reporting and intelligence on the metadata have been part of our core solution offerings for a long time. The newest aspect is machine learning, leveraging the benefit of pattern recognition. We've also branched out into looking at the user behaviors, the access they have – that's a different solution offering, but that's been done for many, many years.

How can teams move from a reactive state to a more proactive stance against new threats?

Westervelt: The biggest piece is a plan, and this is true not just for higher ed but for all customers. You have to have a plan. Very importantly, it has to be a documented plan and it has to be a tested plan. And that last piece is the critical element, where you may have a documented plan or an organization may have a plan, but they haven't tested it. The time of a crisis that is not the time to try to validate whether we've covered all of the bases. Make sure you've accounted for – as much as possible – inevitable scenarios: no access to network, to data, whatever it may be.

Kadakia: The other aspect of that is knowing where your data is. Understanding not just where but also what's in there. Is there any PII information? Is there any classified or very personal information that needs to be super protected? That's where we're seeing teams wanting to have the backups, wanting to have the immutable storage where no one can modify those data. I think a lot of higher education is understanding that we need to invest in our infrastructure and these plans.

What are the technologies or infrastructure components that these teams are relying on now for that protection?

Westervelt: It all comes down to the cloud, which leads to all sorts of different in-roads and conversations. How are they leveraging cloud infrastructure? How can they best leverage it, whether that's hosting mail, access, or other areas for data storage? And also importantly when it comes to recovery, what kind of value-add can be achieved through the cloud?

> Pattern recognition is key: leveraging our solution suite to detect anomalies, things that either shouldn't be occurring or stand out in the normal day-to-day data management.
>
> – VISHAL KADAKIA, VERITAS

Kadakia: A lot of teams look at isolated recovery environments, where they have a location that's kind of like a vault — it's locked down, but they know that what they're putting in there is clean information, therefore they know that it can't be compromised in any way. Indelible WORM storage, where it's write once, read many, that cannot be modified. Ironically, if you look in the protection realm, tape was a great medium because it was offline and it could not be modified. In many cases where organizations are leveraging disk-based protection, we now make offerings and options for indelible storage so that users can prove that content has not been altered or modified. That's a huge step forward and a huge part of the recovery process in the event of ransomware. Three or four years ago that was a nice-to-have, but today it's a given that we need immutable. Automated recovery and orchestration is also critical because teams don't necessarily have the resources to do all of that on a manual basis. Whether that's from an on-prem to cloud or bringing it back down, most teams need an automated method to do that. Also, in the past it would be, teams needed a file recovered or an application in the dataset recovered. Now they're requesting entire data centers and all of their apps at scale recovered.

What are the essential questions higher ed teams should be asking right now around these technologies and capabilities?

Westervelt: I really think it gets down to a risk benefit analysis conversation. Does inviting AI to accomplish a task — whether it's pattern recognition or code development — does it make sense or does it open up a potential risk vector?

It gets back to having that conversation. Our technical teams like to be seen as the trusted advisors for our customers. If there's a doubt or you're unsure, have that conversation and determine whether it makes sense for the organization. Is the risk too great by introducing an AI or third-party entity? Where's that being run? There are so many different potential upsides, but there are also some challenges there as well. Even with all of the new capabilities that are available and accessible to our customers, I'm still surprised by the core conversations we have around recovery, the recovery point and the recovery time. Those fundamentals have not changed in many, many years. It's just the "how" that's changing.