PowerSchool

PowerSchool: A Leader in Responsible AI for Education

Contents

Introduction

Artificial intelligence (AI) is transforming education, enabling personalized teaching and learning, enhancing decision-making, and improving efficiency. However, AI also poses challenges and risks, especially when it comes to data security and privacy. How can educators and administrators trust that the AI applications they use are safe, reliable, and ethical?

PowerSchool is committed to responsible AI, which means that our AI applications are designed and deployed with the highest standards of data protection, governance, and ethics.

PowerSchool follows six pillars of responsible AI:

1. 
   Security
2. 
   Privacy
3. 
   Transparency
4. 
   Fairness
5. 
   Accountability
6. 
   Human-centricity

These pillars guide PowerSchool's approach to developing and delivering AI solutions that empower educators and learners, while respecting their data rights and preferences.

In this report, we explore how PowerSchool ensures the security and privacy of its AI-enabled applications, focusing on two types of testing: guardrail testing and defense testing. We explain what these testing methods are, how they are conducted, and why they are important for responsible AI. Next, we delve into details of recent cybersecurity trends and attacks, and we share details about PowerSchool's commitment to providing the best and most secure web applications for your education needs.

Types of Testing: Guardrail Testing vs. Adversarial Testing

Testing is a crucial part of ensuring the security and privacy of AI applications. Testing helps to identify and prevent potential vulnerabilities, errors, and malicious attacks that could compromise the functionality, performance, and integrity of the AI systems. Testing also helps to ensure that the AI applications comply with the data protection regulations and ethical principles that govern their use.

PowerSchool conducts two types of testing for its AI-enabled applications: guardrail testing and adversarial testing. These testing methods have different objectives and scopes, but they both aim to ensure the security and privacy of the AI applications and the data they process.

Guardrail testing ensures that the information returned to a user is accurate, appropriate, and authorized. It tests the guardrails that are in place to make sure that an AI application is behaving as expected and does not violate the principles of responsible AI. Guardrail testing assumes no bad intent—i.e., that the system is working normally, and that the user is not trying to exploit or manipulate it.

Adversarial testing investigates any defenses in place to protect the AI applications from external threats and attacks. It tests the resilience and robustness of the AI systems against hackers who may try to compromise, disrupt, or steal the data or the functionality of the AI applications. Defense testing assumes bad intent—i.e., that the system is under attack and the attacker is trying to exploit or manipulate it.

To summarize, guardrails protect the user from bad information coming from the AI while adversarial testing protects a user's data from hackers.

Description of Testing Types

Guardrail testing and defense testing are conducted using different methods and tools, depending on the nature and complexity of the AI applications and the data they handle. In this section, we will describe how each type of testing is performed and what standards and best practices are followed.

Guardrail testing focuses on testing for content safety, personally identifiable information (PII) detection/protection, bias, relevance, and other elements of responsible AI use. For example, guardrail testing can determine if AI applications are returning accurate and relevant information to the user, if they are filtering out inappropriate or harmful content, if they are detecting PII, and if they are avoiding or mitigating bias and discrimination.

A simple real-world example of this would be a student attempting to get the system to generate content that is not appropriate. The student may be curious, not realize what they are asking for, or simply trying to make their friends laugh. What they are not doing is attempting to compromise the system.

Guardrail testing is aligned with the six pillars of responsible AI that PowerSchool follows: security, privacy, transparency, accountability, fairness, and human-centricity. Guardrail testing ensures that AI applications respect users' data rights and preferences, provide clear and explainable information, are accountable for their actions and outcomes, are fair and inclusive, and are human-centric and respectful.

Adversarial testing focuses on testing for web application security, network security, data encryption, authentication, and other elements of data protection. For example, defense testing can check if the AI applications are secure from vulnerabilities, such as prompt injection, broken authentication, or insecure output handling, as defined by the OWASP Top 10 for LLM. Adversarial testing may also check if AI applications are secure from other attacks, such as denial-of-service, man-in-the-middle, or spoofing, and if they use strong encryption, authentication, and authorization mechanisms to protect any data in transit or at rest.

Defense testing aligns with the security and privacy pillars of responsible AI that PowerSchool follows. Defense testing ensures that AI applications are secure from unauthorized access, modification, or disclosure, and that they comply with the data protection regulations. Defense testing also involves simulating an actual hacker who may try to exploit an AI application using various techniques and tools.

What is the OWASP Top Ten for LLM?

The Open Web Application Security Project's (OWASP) Top Ten for LLM is a list of the most common security risks for machine learning models. The list is developed by a group of AI security experts and is rapidly becoming an industry standard. LLM stands for Large Language Models, which are machine learning models that can comprehend and generate human language text.

PowerSchool is incorporating LLM features into its products to provide better services for students and teachers. By relying on OWASP, PowerSchool is leveraging a list developed by a highly qualified group of subject matter experts. As such, PowerSchool avoids having to reinvent the wheel.

Transparency

We have discussed how PowerSchool ensures the security and privacy of our AI-enabled applications for K-12 education. We have explained the two types of testing that PowerSchool conducts: guardrail testing and defense testing. We have also described how these testing methods are performed and what standards and best practices are followed.

PowerSchool is a leader in responsible AI for education, and it is committed to providing AI solutions that empower educators and learners, while safeguarding their data and rights.

The six pillars of responsible AI guide PowerSchool's approach to developing and delivering AI applications that are safe, reliable, and ethical.

PowerSchool believes that AI can responsibly transform education for the better. By bringing AI to data, and not pushing data to AI, PowerSchool ensures that our AI applications are secure, private, innovative, and impactful. PowerSchool is proud to be a trusted partner for K-12 education, and we are dedicated to creating a future where every learner can thrive.

Cybersecurity Trends in 2023

As the education sector has become more reliant on digital technologies, we face increasing cybersecurity threats from hackers who seek to exploit vulnerabilities, steal data, or disrupt operations. In this section, we discuss two of the most common and dangerous types of cyberattacks that PowerSchool customers should be aware of: phishing and ransomware. We will also provide tips on how to prevent, detect, and respond to these attacks, and how PowerSchool can help you secure your online learning environment.

Notable 2023 Attacks

It was a rough year for school cybersecurity attacks in 2023. Below are three notable examples. It's important to note, however, that this list isn't about shaming the victims of these attacks. They are reminders that cyber criminals are targeting schools. It doesn't matter if it is a huge district or a tiny one. Cybercriminals tend to take advantage of any vulnerability that presents itself.

New Haven Public School District, Connecticut

The City of New Haven lost more than $6 million after multiple cyberattacks were carried out on its public schools. While the FBI has recovered over half of that amount, a lot of money remains lost. This is a prime example of why having proper cybersecurity can save public schools a lot of money in the long run.

The cyberthefts occurred in the summer when attackers impersonated private vendors and the city's chief operating officer by email. The hackers requested electronic transfers to be paid to fraudulent accounts. Six of the recipients paid more than $5.9 million, believing they were paying the school's bus company.

Minneapolis Public Schools

The Minneapolis Public School district revealed that a data breach and ransomware attack from earlier in the year affected over 100,000 individuals. The information believed to be breached included names, addresses, social security numbers, state student numbers, and health insurance data.

In this example, MPS district was given an ultimatum: Pay $1 million or the sensitive information will be released. After the district failed to pay the ransom, the information was uploaded to the internet by the attackers.

Colorado Department of Higher Education

Colorado Department of Higher Education fell victim to a huge school data breach that leaked sensitive data of educators and current and former students. The breach was the result of a ransomware attack. The information that was breached included names, dates of birth, social security numbers, photocopies of government IDs, and even police reports pertaining to identity theft.

Tip of the Iceberg

These three breaches represent just the tip of a very large iceberg. According to Malwarebytes Labs, a security research company, there was a 70% surge in attacks on education organizations in 2023. Corvus Insurance reported a nearly identical surge in 2023 with over 1,200 victims. There is also speculation that there are probably additional breaches that never made the news or were never reported.

Two types of attacks were found to be most prevalent in this epidemic of breaches: phishing and ransomware.

Phishing: The Art of Deception

Phishing is a form of social engineering that involves sending fraudulent emails or other messages that appear to come from legitimate sources, such as trusted organizations, colleagues, or friends. The goal of phishing is to trick the recipients into clicking on malicious links or attachments, or providing sensitive information, such as passwords, credit card numbers, or bank account details. It was likely one of the things that led to the breach at the New Haven Public School District.

According to IBM Security X-Force 2024 Threat Intelligence Index, phishing was identified as the primary infection vector in 30% of cybersecurity incidents in 2023. Moreover, phishing attacks have become more sophisticated and targeted over time, using various techniques to evade detection and increase their success rate. Some of these techniques include:

- **Spear phishing:** This is a type of phishing that targets specific individuals or groups within an organization, using personalized and relevant information to make the emails more convincing. For example, a spear phishing email may impersonate a senior executive, a vendor, or a customer, and request a wire transfer, a password reset, or a document download. Spear phishing attachments were used in 62% of phishing attacks, while links were used in 33%, and as a service in 5% (IBM Security X-Force 2023).
- **Business Email Compromise (BEC):** This is a form of spear phishing that involves impersonating or compromising the email accounts of high-level executives or employees who have access to financial or confidential information. The attackers then use these accounts to request or authorize fraudulent transactions, such as invoice payments, payroll changes, or gift card purchases. BEC accounted for 6% of incidents, with spear phishing links used in half of these cases (IBM Security X-Force 2023).
- **Brand impersonation:** This is a type of phishing that mimics the look and feel of well-known brands, such as banks, online retailers, or social media platforms, and attempts to lure the recipients into visiting fake websites or downloading malicious apps. The attackers often exploit the trust and familiarity that people have with these brands, and take advantage of seasonal events, such as holidays, sales, or promotions, to increase their chances of success. In Q4 2023, malicious links became more prevalent than malicious attachments for the first time, and there was a surge of Amazon-related campaigns during the holiday shopping season (Mimecast Global Threat Intelligence Report, October-December 2023).

Phishing is a serious threat that can compromise your security and privacy, as well as expose you to legal and reputational risks. Therefore, it is essential that you and your staff are aware of the signs of phishing and know how to avoid falling victim to it. Here are some best practices to follow:

- **Do not open or respond to unsolicited or suspicious emails or messages**, especially if they ask for personal or financial information, or urge you to take immediate action
- **Do not click on links or attachments you are not expecting or that look suspicious.** Hover over the link to see the actual URL and check for spelling or grammatical errors or unusual domain names.
- **Verify the sender's identity and authenticity by checking their email address, signature, or contact details.** If in doubt, contact the sender directly using a different channel, such as a phone call or text message, or a trusted email address.
- **Use strong and unique passwords for your online accounts and change them regularly.** Do not reuse the same password for multiple accounts or share your passwords with anyone.
- **Enable multi-factor authentication (MFA) for your online accounts, especially those that contain sensitive or confidential information.** MFA adds an extra layer of security by requiring you to enter a code, answer a question, or use a device, in addition to your password, to access your account. In 80% of the organizations where a BEC attack occurred, no MFA solution was in place before their incident.
- **Keep your devices and software updated with the latest security patches and antivirus software.** This can help you prevent malware infections and protect your data from unauthorized access.
- **Report any suspicious or phishing emails or messages to your IT department, or to the relevant authorities,** such as the Federal Trade Commission (FTC) or the Anti-Phishing Working Group (APWG).

Ransomware: The Cost of Extortion

It was a banner year for ransomware in education in 2023. Reported attacks were up 68% over 2022. Ransomware played a role in two of our three example breaches from 2023, Colorado Department of Higher Education and Minneapolis Public Schools.

Ransomware is a type of malware that encrypts your files or locks your device and demands a ransom for their decryption or release. Ransomware can affect any device, such as computers, smartphones, tablets, or servers, and any type of data, such as documents, photos, videos, or databases. Ransomware can cause significant damage and disruption to your operations, as well as financial losses and legal liabilities.

According to Corvus Insurance and BlackFog Security, ransomware attacks increased by 69% in 2023, after an overall decrease in 2022. Moreover, ransomware attacks have become more sophisticated and aggressive over time, using various techniques to evade detection and increase their impact. Some of these techniques include:

- **Double extortion:** This is a type of ransomware that not only encrypts your data, but also exfiltrates it to the attackers' servers, and threatens to publish or sell it if you do not pay the ransom. This adds more pressure and leverage to the attackers and increases the risk of data breaches and privacy violations.
- **Ransomware as a service (RaaS):** This is a type of ransomware that is offered as a service by cybercriminals who provide the malware, the infrastructure, and the support to other attackers who pay a fee or a share of the ransom in return. This lowers the barrier of entry and increases the scale and diversity of ransomware attacks.
- **Targeted attacks:** This is a type of ransomware that targets specific organizations or sectors, such as healthcare, education, or government, that are more likely to pay the ransom, or that have a higher impact on society. The attackers often conduct reconnaissance and compromise the network before deploying the ransomware, and demand higher-than-average ransoms.

Ransomware is a serious threat that can compromise your availability and integrity, as well as expose you to legal and reputational risks. Therefore, it is essential that you and your staff are prepared for ransomware attacks, and know how to prevent, detect, and respond to them. Here are some best practices to follow:

- **Back up your data regularly and securely, and test your backups frequently.** This can help you restore your data in case of a ransomware attack and reduce your dependence on the attackers.
- **Do not pay the ransom or negotiate with the attackers.** Paying the ransom does not guarantee that you will get your data back, or that it will not be leaked or sold. It also encourages the attackers to continue their malicious activities and makes you a more attractive target for future attacks.
- **Isolate and disconnect any infected or compromised devices from the network, and contact your IT department, or a trusted security provider, for assistance.** This can help you contain the spread of the ransomware and prevent further damage or data loss
- **Report any ransomware attacks to the relevant authorities, such as the Federal Bureau of Investigation (FBI) or the Cybersecurity and Infrastructure Security Agency (CISA).** This can help you get support and guidance and contribute to the investigation and prosecution of the attackers.

Phishing and ransomware are two of the most common and dangerous types of cyberattacks that PowerSchool customers should be aware of. They can cause significant harm and disruption to your learning environment, as well as financial losses and legal liabilities. Therefore, it is important that you and your staff are educated and vigilant about these threats and follow the best practices to prevent, detect, and respond to them.

Conclusion

At PowerSchool, we are committed to providing you with the best and most secure web applications for your education needs. We use the latest technologies and standards to protect your data and systems from cyberattacks, and we offer you various tools and resources to help you enhance your cybersecurity posture. We also monitor and respond to any emerging threats or incidents, and we work closely with our partners and customers to ensure your safety and satisfaction. If you have any questions or concerns about cybersecurity, or if you need any assistance or support, please do not hesitate to contact us. We are here to help you.

PowerSchool

Personalized Education for Every Journey

www.PowerSchool.com

BR_PSBR-1275921_061024