Student Data Privacy

Why Not To Share

Before we dive into some specific examples of data we shouldn't share, let's start by exploring why this is the case to begin with. Obviously the biggest reason is simply to **comply with the law** (such as **FERPA** and **GDPR**) and avoid legal actions, loss of funding, fines, and more.

However, another reason is because the **data** you enter into an AI chatbot is often **kept by the company** running that tool (although not for all AI tools as we will look at further down). For those that do store data, there is always the potential for the information getting out. One way this can happen is as **training data for the next model**.

The way companies build AI chatbots is by feeding them as much data as they can, such as every book ever written, every website on the internet, images, videos, and more. One key to building the next, more powerful model is to feed it more and more data. Some of that data can come from users.

As millions of people use ChatGPT and Gemini and Claude and others, we are entering billions of bits of information. In most cases, that **information** can be used by the company to **help train the next version** of the tool. Although AI companies work hard to anonymize such data, there is always the possibility of this information being accessed in some form through prompts in the future.

However, even if that never happens, there is always the chance of a **data breach** which would once again expose the information we have submitted. In general it is best to think of AI chatbots as potentially **public** data warehouses, rather than personal **private** tools.

🚫 What Not To Share

Below is a list of many types of student data that should not be shared with an AI chat tool, along with some more specific examples for each category.

- **Personally Identifiable Information (PII)** - Names, addresses, phone numbers, email addresses, social security numbers, student identification numbers, birth dates
- **Educational Records** - Grades, transcripts, class schedules, disciplinary records, disabilities, and Individual Education Plans (IEPs).
- **Health Information** - Medical records, health conditions, allergies, medication information, therapy records.
- **Financial Information** - Family income, financial aid information, bank account details.
- **Behavioral or Disciplinary Records** - Disciplinary actions, behavior reports, counseling records.
- **Photos or Videos** - Images or recordings of students without explicit consent.

- **Communication Logs** - Personal messages, emails, and communication with parents or guardians.

💡 How You Can Share

With those **restrictions** in mind, there are some **options** for using AI tools while still supporting student needs. See below for some ideas.

Anonymizing Data

One approach to protecting student data privacy when using an AI chatbot is to **anonymize and generalize the data**. Some ideas include:

- **Remove All Personally Identifiable Information (PII)** - Ensure that names, addresses, birthdates, social security numbers, and any other direct identifiers are excluded from prompts.
- **Generalize the Information** - For example, descriptions of behaviors, educational challenges, and needs that are framed in a general and non-identifiable manner are safer to share.
- **Use Hypothetical Scenarios** - Creating hypothetical scenarios that mirror the student's needs without revealing their identity or any specific, identifiable details are safer to share.

Note: Even with all of these steps it may still be possible to **re-identify a student** from multiple pieces of data. As such, a good rule of thumb is to share the **minimal necessary information**. Only include details essential for understanding the context and providing the needed support. Less is more when it comes to protecting privacy.

🏢 Enterprise-level AI Tools

Although **personal-use AI tools** such as ChatGPT and Gemini may store and use information provided by users, that is not the case with all AI tools. For example, businesses would not want to use tools that could potentially store and share confidential information about their operations. For such cases businesses can use **Enterprise-level AI tools** that do not store or share their data. Thankfully **schools** can use such tools as well.

Here are two examples:

Microsoft Copilot Enterprise

Microsoft offers a version of their **Copilot AI tools** that comes with **commercial data protection**. In Microsoft's own words:

"We're happy to share that we are expanding eligibility for commercial data protection to all faculty users and to higher education students ages 18 and above. Copilot provides AI chat for the web with access to models like GPT-4 and DALL-E 3 at no additional cost. Commercial data protection will be enabled when eligible users are signed in with their school account starting in early 2024. This means user and organizational data are protected, chat prompts and responses in Copilot are not saved, Microsoft has no eyes-on access to them, and they aren't used to train the underlying large language models."

You can learn more here: Resource link

Gemini for Google Workspace

Similarly **Google** has developed an Enterprise version of their **Gemini AI tools** that supports **data privacy**. In Google's own words:

"With a Gemini Business and Gemini Enterprise plan, your conversations are not used for advertising purposes, reviewed by human reviewers, or otherwise used to train generative machine-learning technologies. In fact, all of our commitments to data privacy, confidentiality, and security apply to Gemini for Workspace."

According to Google a version for **education** is coming soon:

"We're also working to bring Gemini for Workspace to education customers, and we look forward to sharing more about this in the coming weeks."

You can learn more here: Resource link

🏠 Local AI Tools

Another option for using AI tools while still protecting student data is to use a **local AI chatbot**. These are AI systems that you **install and run directly on your own computer**, rather than relying on cloud-based services. In such a case any data you enter never actually leaves your computer.

There are several **benefits** to using a local AI chatbot:

- **Data Control and Privacy** - Since the AI operates locally, the data processed by the chatbot does not leave your device or network. This means sensitive information, such as student data, remains within the confines of your control, offering a higher degree of privacy and security.

- **Compliance with Data Protection Regulations** - Using a local AI tool can help in adhering to data protection laws like FERPA and GDPR. By not transmitting data externally, you reduce the risk of unintentional data breaches or non-compliance with legal requirements.
- **Customization and Specialization** - Local AI models can be tailored to specific educational needs or requirements without the constraints of cloud-based services. This could mean a more targeted use of AI in educational settings.

However there are some **challenges** that come with local AI chatbots as well:

- **Resource Intensive** - Running an AI model locally requires a powerful computer with hard drive space to store the model, lots of memory to handle the information, and a fast processor to run the queries. Typical school computers may not have the specifications needed to run such a tool efficiently.
- **Maintenance and Updates** - Unlike cloud-based services that are automatically updated by the provider, local AI systems require manual updates and maintenance. This can be a challenge for educators or IT departments to ensure the AI tool remains effective.
- **Technical Expertise** - Implementing and managing a local AI chatbot requires a higher level of technical expertise than using a cloud-based service. This may be beyond the comfort level of many educators and require more support from the school IT staff.

Of course all of this may change very quickly as AI tools continue to evolve and computers become more powerful. As of this writing, local AI chatbots are still a very new technology.

Some popular **options for local AI chatbots** include:

- GPT4All - gpt4all.io
- Llama - llama.meta.com
- Vicuna - lmsys.org/blog/2023-03-30-vicuna
- NVIDIA Chat with RTX - nvidia.com/en-us/ai-on-rtx/chat-with-rtx-generative-ai

🏁 Conclusion

In the end, data privacy isn't just about compliance. It's about protecting our students and fostering trust in the digital age. Even with all of the considerations mentioned above, it is important for educators to adhere to your school or district's policies on student data privacy. Always aim to share the minimum necessary information and consult with legal or data privacy experts within your organization if in doubt.